RECENT PROGRESS IN

HORMONE RESEARCH

Edited by

ANTHONY R. MEANS

VOLUME 58

The Human Genome and Endocrinology

“Recent Progress in Hormone Research” is an annual publication of The Endocrine Society that is published under the editorial auspices of Endocrine Reviews.

The Endocrine Society
8401 Connecticut Avenue, Suite 900
Chevy Chase, Maryland 20815
CONTENTS

Senior Author Correspondence Information v

1. A Functional Proteomics Approach to Signal Transduction
   Paul R. Graves and Timothy A.J. Haystead 1

2. The Use of DNA Microarrays to Assess Clinical Samples: The Transition from
   Bedside to Bench to Bedside
   John A. Copland, Peter J. Davies, Gregory L. Shipley, Christopher G.
   Wood, Bruce A. Luxon, and Randall J. Urban 25

   Erich Huang, Mike West, and Joseph R. Nevins 55

4. Statistical Approach to DNA Chip Analysis
   N.M. Svrakic, O. Nescic, M.R.K. Dasu, D. Herndon, and J.R. Perez-Polo 75

5. Identification of a Nuclear Factor Kappa B-dependent Gene Network
   Bing Tian and Allan R. Brasier 95

6. Role of Defective Apoptosis in Type 1 Diabetes and Other Autoimmune Diseases
   Takuma Hayashi and Denise L. Faustman 131

7. Genomic Analysis of Glucocorticoid-regulated Promoters in Murine T-lymphoma
   Cells
   Lu Chen, Celeste Finnerty, William C. Gustafson, Craig R. Bush, Ping Chi,
   Huiping Guo, Bruce Luxon, Alan P. Fields, and E. Aubrey Thompson 155

8. Regulation of a Distinctive Set of Genes in Glucocorticoid-evoked Apoptosis in
   CEM Human Lymphoid Cells
   E. Brad Thompson and Betty H. Johnson 175

9. Overlapping but Distinct Profiles of Gene Expression Elicited by Glucocorticoids
   and Progestins
   Yihong Wan and Steven K. Nordeen 199

10. Analysis of Gene Expression in the Normal and Malignant Cerebellum
    Robert J. Wechsler-Reya 227

11. Paired-like Repression/Activation in Pituitary Development
    Lorin E. Olson, Jeremy S. Dasen, Bong Gun Ju, Jessica Tollkuhn, and
    Michael G. Rosenfeld 249
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Dynamic Changes in Gene Expression During Human Trophoblast...</td>
<td>263</td>
</tr>
<tr>
<td></td>
<td>Stuart Handwerger and Bruce Aronow</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Regulation of Hematopoietic Stem Cell Self-Renewal</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>Tannishtha Reya</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Investigation of the Transcriptional Changes Underlying...</td>
<td>297</td>
</tr>
<tr>
<td></td>
<td>Functional Defects in the Mammary Glands of Prolactin Receptor...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Christopher J. Ormandy, Matthew Naylor, Jessica Harris, Fiona...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Robertson, Nelson D. Horseman, Geoffrey J. Lindeman, Jane...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Visvader, and Paul A. Kelly</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Microarray Analysis and Identification of Novel Molecules...</td>
<td>325</td>
</tr>
<tr>
<td></td>
<td>Involved in Insulin-like</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Growth Factor-1 Receptor Signaling and Gene Expression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Joelle Dupont, Sandra E. Dunn, J. Carl Barrett, and Derek...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LeRoith</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Consequences of Elevated Luteinizing Hormone on Diverse Physiological Systems:</td>
<td>343</td>
</tr>
<tr>
<td></td>
<td>Use of the LH8CTP Transgenic Mouse as a Model of...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ovarian Hyperstimulation-induced Pathophysiology</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rachel J. Mann, Ruth A. Keri, and John H. Nilson</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td></td>
<td>377</td>
</tr>
<tr>
<td>Subject Index</td>
<td></td>
<td>379</td>
</tr>
</tbody>
</table>
Senior Author
Correspondence Information

Allan R. Brasier
Department of Internal Medicine
University of Texas Medical Branch
Room MRB 8.138, Route 1060
301 University Boulevard
Galveston, TX 77555-1060
Phone: 409-772-2824
Fax: 409-772-8709
E-mail: arbrasie@utmb.edu

Denise L. Faustman
Immunobiology Laboratory
Massachusetts General Hospital/Harvard Medical School
MGH-East, Building 149
13th Street
Charlestown, MA 02129
Phone: 617-726-4084
Fax: 617-726-4095
E-mail: faustman@helix.mgh.harvard.edu

Stuart Handwerger
Division of Endocrinology
Children's Hospital Medical Center
3333 Burnet Avenue
Cincinnati, OH 45229-3039
Phone: 513-636-4209
Fax: 513-636-7486
E-mail: stuart.handwerger@chncc.org

Timothy A.J. Haystead
Department of Pharmacology and Cancer Biology
Duke University Medical Center
Levine Science Research Center, Room C118
Durham, NC 27710-3686
Phone: 919-613-8606
Fax: 919-668-0977
E-mail: Hays001@nc.duke.edu

Derek LeRoith
Clinical Endocrinology Branch
National Cancer Institute
National Institutes of Health
Room 8D12, Building 10
MSC 1758
Bethesda, MD 20892-1758
Phone: 301-496-8090
Fax: 301-480-4386
E-mail: derek@helix.nih.gov

Joseph R. Nevins
Department of Molecular Genetics and Microbiology
Howard Hughes Medical Institute
Duke University Medical Center
366 Clinical and Research Laboratories
Durham, NC 27710
Phone: 919-684-2746
Fax: 919-681-8973
E-mail: j.nevins@duke.edu
A Functional Proteomics Approach to Signal Transduction

PAUL R. GRAVES* AND TIMOTHY A.J. HAYSTEAD†

*Department of Pharmacology and Cancer Biology, Duke University, Durham, North Carolina 27710-3686; †Serenex Inc., 323 Foster Street, Durham, North Carolina 27710

ABSTRACT

The purpose of this review is to highlight how proteomics techniques can be used to answer specific questions related to signal transduction in a wide variety of systems. In our laboratory, we utilize proteomic technologies to elucidate signal transduction pathways involved in smooth muscle contraction and relaxation, cell growth and tumorigenesis, and the pathogenesis of malaria. We see the real application of this technology as a tool to enhance the power of existing approaches such as classical yeast and mouse genetics, tissue culture, protein expression systems, and site-directed mutagenesis. Our basic approach is to examine only those proteins that differ by some variable from the control sample. In this way, the number of proteins to be processed by electrophoresis, Edman degradation, or mass spectrometry is greatly reduced. In addition, since only those proteins that change in response to a given biological treatment are analyzed, the experimental outcome provides information about specific signaling pathways. Examples of typical experiments in our laboratory are measurement of changes in protein phosphorylation in response to treatment of cells with growth factors or specific drugs, characterization of proteins associated with a bait protein in a “pull-down” experiment, or measurement of changes in protein expression. Frequently, in these experiments, it is necessary to define complex protein mixtures. To achieve this goal, we utilize a variety of techniques to isolate specific types of proteins or “subproteomes” for further analysis. In this review, we discuss strategies used in our laboratory for studying signaling pathways, including subproteome isolation, proteome mining, and analysis of the phosphoproteome.

I. Introduction

Functional proteomics can be defined as a method to identify specific proteins in a cell, tissue, or organism that undergo changes in abundance, localization, or modification in response to a specific biological condition. In functional proteomics, the aim is not to identify or characterize every protein in the cell but rather to provide information about a small number of proteins that are directly relevant to the biological question being studied. To achieve this goal, functional proteomics often is combined with complementary techniques such as protein biochemistry, molecular biology, and cell physiology. In our laboratory, we use functional proteomics to answer questions about signal transduction in a variety of systems. In this review, we describe how proteomics
techniques can be applied to study changes in global protein expression or modification in response to different conditions. In particular, we discuss how the phosphoproteome can be analyzed and provide examples of our own work to illustrate the utility of these methods.

II. The Need for Protein Studies in Signal Transduction

The completion of genome-sequencing projects will facilitate the analysis of signaling pathways in various organisms. As a first step towards this goal, changes in the expression level of mRNA have been measured using techniques such as serial analysis of gene expression (SAGE) (Velculescu et al., 1995) or DNA chip microarrays (Schena et al., 1995; Shalon et al., 1996). These methods will provide valuable information that can be used in conjunction with proteomic studies. However, these methods alone cannot provide accurate information about protein levels in the cell because mRNA levels do not necessarily reflect protein abundance. Indeed, several studies have shown a poor correlation between mRNA levels and protein expression (Anderson and Seilhamer, 1997; Gygi et al., 1999b; Ideker et al., 2001).

Another reason protein analysis is important is because proteins can undergo post-translational modifications and these modifications cannot be predicted from genomic data. Proteins can be modified by the attachment of groups such as phosphates, sulfates, carbohydrates, and lipids. These modifications can change the function of proteins in cells. This is especially important in signal transduction, where signals frequently are transmitted by protein modifications. One type of protein modification that is particularly abundant is protein phosphorylation. It is estimated that 30% of all cellular proteins contain covalently bound phosphate (Hubbard and Cohen, 1993). Thus, to obtain a full understanding of signal transduction pathways, it is essential to study protein modifications at the molecular level.

Finally, protein studies are necessary in signal transduction because signaling pathways are composed of proteins. Many proteins exist in protein complexes whose function may be elucidated only through a study of the intact complex. Currently, the composition and regulation of many protein complexes in signaling remain unknown. In a recent study of protein interactions in *Saccharomyces cerevisiae*, hundreds of multiprotein complexes were detected, of which more than 90% contained at least one previously uncharacterized protein (Gavin et al., 2002; Ho et al., 2002). To fully understand signaling pathways, the individual proteins in a protein complex will have to be identified and characterized during signaling events. These types of experiments will require a proteomics approach.
III. A Proteomics Approach to Signal Transduction

In signal transduction, experiments often are conducted to activate or inhibit specific pathways in order to elucidate the components and regulation of the pathway. For example, signaling pathways can be modulated by treatment of cells with growth factors, hormones, or small molecules. Interference of specific pathways sometimes can be achieved with drugs, gene knockouts, or mRNA interference. Whatever the condition analyzed, in a proteomics approach, the entire cell, tissue, or animal is homogenized and proteins are extracted for study. Proteins that differ by some variable then are selected for further study. The initial results can raise a host of questions. For example, what is the protein that changes in response to the condition? Is the protein modified? What form of modification is it? What are the enzymes responsible for the modification? Finally, what is the biological significance of the protein change observed?

Because all these questions must be answered to understand signal transduction, proteomics represents only the first step in this endeavor. It is no longer sufficient to produce long lists of proteins that undergo changes in a given system. It is also necessary to validate the changes by other methods. Validation of a change in a specific protein may involve cloning and expression of the protein in a heterologous system, protein biochemistry to confirm modifications, and cell biology to determine biological significance. Validation of a single protein or its modification often requires 6–12 months of work. Therefore, a proteomics experiment should be designed carefully so that only proteins relevant to the biological problem are studied. Otherwise, significant time can be spent investigating irrelevant proteins.

One of the major hurdles in proteomics occurs at the beginning of the experiment. That is, how can the complex mixture of proteins produced by homogenization of whole cells, tissues, or organisms be analyzed simultaneously? One of the difficulties inherent in a proteomics approach is that a cell extract contains thousands of distinct proteins. These proteins have to be resolved from each other so that changes in individual proteins can be observed. Another difficulty is the detection of low-copy proteins in a cell extract. Finally, because a large number of proteins are involved in proteomic analysis, strategies must be devised to focus the investigation on certain types of proteins.

A. ANALYSIS OF PROTEIN MIXTURES BY TWO-DIMENSIONAL GEL ELECTROPHORESIS

An example of how two-dimensional gel electrophoresis (2-DGE) can be used in proteomics is shown in Figure 1A. In this approach, a cell extract is prepared from two different samples and resolved by 2-DGE for comparison of protein expression or changes in protein modification. This method can provide valuable information and has been used to investigate specific signaling path-
ways (Lewis et al., 2000). However, this approach also suffers from many limitations, including poor protein representation and an inability to identify low-copy proteins (Gygi et al., 1999b; Graves and Haystead, 2002). In addition,
it is difficult to screen large numbers of samples for changes because 2-DGE is very labor intensive and time consuming. Therefore, alternative strategies are needed.

B. ALTERNATIVES TO ELECTROPHORESIS

The limitations of 2-DGE have inspired the development of methods to circumvent protein gels entirely. In one such method, the entire protein mixture isolated from a cell is converted to peptides, then the peptides are resolved by liquid chromatography. Following separation, the peptides are injected directly into a mass spectrometer in an “on-line” configuration for mass analysis and protein identification (Figure 1B). A variety of techniques have emerged using this general strategy, including multidimensional liquid chromatography (Link et al., 1999), cation-exchange and reverse-phase chromatography (Opiteck et al., 1997), and liquid chromatography combined with ion mobility spectrometry and tandem mass spectrometry (LC-IMS-TOF-MS/MS) (Kindy et al., 2002). To detect changes in protein expression levels, a method known as isotope coded affinity tags (ICAT) was introduced (Gygi et al., 1999a). In this approach, peptides from two different samples are labeled covalently with a reagent that allows for their subsequent purification and discrimination in the mass spectrometer (Gygi et al., 1999a).

The advantage of these methods is that, theoretically, more proteins can be identified, especially low-copy proteins. In addition, because protein gels are avoided, time is saved from loading and extracting proteins from gels, a process that is not easily automated. However, protein gels perform an important function. They allow for the visual selection of specific proteins from a complex mixture. In the absence of such visual selection, every protein (or peptide) in a cell extract has to be analyzed to find those that undergo changes in response to the biological condition. Frequently, it is important to activate specific signaling pathways and determine early events to understand how signals are transmitted. In these experiments, activation of a specific signaling pathway will not result in changes in the majority of cellular proteins. Therefore, if all proteins in a cell are analyzed, most of the analysis time will be spent on irrelevant proteins. In addition, the analysis of large numbers of proteins requires specialized equipment and computer systems to handle massive amounts of data.

C. TARGETED ANALYSIS OF SUBPROTEOMES

In our laboratory, we utilize an approach that enriches for low-copy proteins and allows for their visual selection. In this method, affinity chromatography is used to isolate subproteomes or specific types of proteins related to the biological question. These proteins are then visualized using one-dimensional gel electrophoresis (1-DGE) or 2-DGE (Figure 1C). The isolation of subproteomes of the
cell serves two important functions. First, it reduces the complexity of the analysis because fewer proteins are analyzed. This saves data analysis time and focuses attention on only those proteins relevant to the biological question. Second, it enriches for low-abundance proteins that otherwise would go undetected in cell extracts. Because of the reduction in protein complexity achieved with this approach, the protein samples can be analyzed for differences using conventional 1-DGE. As a result, many more proteins are represented in the gels and large numbers of samples can be screened quickly.

An example of subproteome enrichment is demonstrated by the use of the specific affinity resin, $\gamma$-phosphate-linked adenosine triphosphate (ATP)-Sepharose, originally developed in our laboratory for the purification of protein kinases (Haystead et al., 1993). Unlike previous nucleotide resins, we linked ATP in an orientation favorable for binding protein kinases based upon the crystal structure of cyclic AMP (cAMP)-dependent protein kinase (Knighton et al., 1991). From this structure, it was shown that the $\gamma$-phosphate group of ATP extended to the solvent. Therefore, this part of ATP was used for linkage to the matrix. We have shown that, in addition to binding protein kinases, $\gamma$-ATP-Sepharose has the ability to bind a large variety of purine-utilizing enzymes, including dehydrogenases, heat shock proteins, nonprotein kinases, DNA ligases, and miscellaneous purine-utilizing enzymes (Graves et al., 2001). This subproteome can be termed the "purine-binding proteome."

Because many components of signal transduction pathways (e.g., protein kinases) are present in the purine-binding proteome, $\gamma$-ATP-Sepharose can be used as a tool to dissect protein-signaling pathways. In this approach, the entire purine-binding proteome is captured from control cells or cells that have undergone some form of treatment. Proteins then are eluted from the matrix with free ATP and either resolved by 2-DGE or further fractionated by ion-exchange chromatography (Figure 2). This approach can provide information about changes in protein expression, in the activation state of protein kinases (as reflected in their ability to bind $\gamma$-ATP-Sepharose), or in the post-translational modification of proteins.

IV. Protein Identification

Protein identification is an essential part of almost every proteomics experiment. Here, we outline how mass spectrometry (MS) can be used to identify proteins and describe our approach to protein identification. (For a more comprehensive review of MS, see Burlingame et al., 1998; Yates, 1998.)
FIG. 2. Analysis of changes in the purine-binding proteome of cells. In this approach, the entire purine-binding proteome from control or treated cells is captured on γ-ATP-Sepharose and the matrix is washed to remove nonspecific proteins. Proteins are eluted with free ATP and analyzed by 2-DGE. As an alternative, the protein eluate can be further resolved by ion-exchange chromatography and 1-DGE. Proteins that differ by some variable are identified by MS. This method can allow for changes in protein expression or modification to be detected in the purine-binding proteome.
A. MASS SPECTROMETRY

MS is now the dominant technology for protein identification (Pandey and Mann, 2000). Using MS, most proteins can be identified at the femtomolar level. However, the success of protein identification can depend upon the sample preparation and on the type of mass spectrometer used. In functional proteomics, considerable effort frequently is expended to obtain a relatively small number of proteins for analysis. In this case, the most important feature of a mass spectrometer is high sensitivity and ability to identify a protein with the highest level of confidence. To this end, we have found that electrospray ionization and tandem mass spectrometry (MS/MS) are the most useful. For MS analysis, samples must be converted to gas-phase ions. Two of the most common methods to accomplish this are matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI).

1. MALDI-MS

In MALDI, a laser is used to promote the ionization of analyte molecules embedded in a crystalline matrix (Karas and Hillenkamp, 1988). The single biggest advantage to MALDI analysis is speed. In MALDI, the entire process of sample preparation and analysis in the mass spectrometer can be automated. This is not yet the case for ESI. As a result, MALDI is the preferred method for high-throughput proteomics operations. However, for many functional proteomics applications, MALDI may not be the best choice. This is because the type of data obtained may not be specific enough for protein identification. In MALDI, the masses of individual peptides derived from a protein are measured in what is known as a peptide mass “fingerprint.” In general, peptide mass fingerprint analysis is not as reliable for protein identification as peptide amino acid sequence obtained by MS/MS. Because of this deficiency, a new breed of MALDI mass spectrometers have emerged that can, in addition to peptide mass fingerprinting, also obtain peptide amino acid sequence. Examples of these machines include MALDI-Q-TOF (Loboda et al., 2000; Shevchenko et al., 2000) and MALDI-TOF-TOF (Medzihradszky et al., 2000). It remains to be seen which machine will be most successful at protein identification.

2. ESI-MS

In ESI, a sample is directly introduced into the mass spectrometer as a fine mist of charged droplets (Fenn et al., 1989; Hunt et al., 1992). In an adaptation of electrospray known as nanospray, samples are introduced through microcapillary tubes at very low flow rates (Wilm and Mann, 1996; Wilm et al., 1996). ESI has a number of advantages over MALDI in functional proteomics applications and, as a result, is the method of choice for sample analysis in our
laboratory. First, nano-electrospray MS can be used to obtain peptide amino acid sequence in a procedure known as MS/MS (Wilm et al., 1996). This is the most specific method for protein identification and greater confidence can be achieved in protein assignment. Second, from our experience, nanospray provides a higher sensitivity than MALDI, allowing for the identification of low-copy proteins. Finally, and perhaps most importantly, MS/MS is less affected by protein mixtures than peptide mass fingerprinting. This is critically important because, frequently, protein “bands” isolated from polyacrylamide gels contain more than one protein.

B. OUR APPROACH TO PROTEIN IDENTIFICATION

The general approach for protein identification in our laboratory is outlined in Figure 3. In this method, a protein band is excised from a gel, digested to peptides by trypsin, and the peptides purified by the method of Wilm and colleagues (Wilm et al., 1996). The resultant peptides are ionized by nano-
electrospray and mass analyzed with a quadrupole-time-of-flight (TOF) mass spectrometer (Q-STAR, PE-SCIEX) to obtain a spectrum of the peptides. From this spectrum, a parent ion is selected for fragmentation and amino acid sequencing. We are capable of identifying faint, silver-stained bands that equal approximately 5–10 ng of protein. However, because MS is not quantitative, frequently several different proteins can be identified from a single band isolated from a polyacrylamide gel. Thus, additional experiments often are required to confirm that the protein identified in the mass spectrometer correlates with a given protein observed on a gel.

C. DATABASE SEARCHING

Using data produced by mass spectrometers, proteins can be identified by searching DNA and protein-sequence databases. The success of protein identification depends upon the type of data utilized, the type of search conducted, and the databases searched. In cases where databases are incomplete, different search strategies are necessary.

1. Peptide Mass Fingerprint Database Searching

In this approach, peptide masses obtained from MALDI-MS are compared against theoretical spectra obtained from primary-sequence databases (Eng et al., 1994). Although this type of search method is rapid, it is not error tolerant and requires a high degree of agreement for success. Factors that alter peptide masses from those in the database (e.g., post-translational modifications) can invalidate this search method. Since many proteins in signaling pathways are modified, this search method may not result in correct protein identification. Another factor that can contribute to inaccurate protein identification is mass degeneracy. For example, a peptide with the sequence VSLNP will have the same mass as the peptide LSNVP. In addition, the masses of certain amino acids, when combined, can equal the mass of other amino acids. For example, the mass of glycine plus glutamate or aspartate plus alanine equals the mass of tryptophane. Because of these limitations, peptide mass fingerprinting is not well suited for protein identification from organisms with large genomes (e.g., human) or towards genomes that are not completely annotated (Qin et al., 1997).

2. Peptide Mass Tag Database Searching

Peptide mass tag database searching can be conducted using peptide sequence obtained from MS/MS. In this approach, a partial amino acid sequence, known as the sequence tag, is combined with the mass of the peptide to search relevant databases (Eng et al., 1994; Mann and Wilm, 1994). The advantage to this method is that it is fast and, if enough amino acid sequence is obtained, can
be a very specific method for protein identification. However, this search method also depends upon a complete and well-annotated sequence database.

3. FASTS Database Searching

The FASTS program uses amino acid sequence obtained from MS/MS to search databases from organisms whose genomes have not been sequenced completely or whose databases are not fully annotated (Mackey et al., 2002). In this approach, *de novo* peptide sequence is obtained from MS, entered into the FASTS program, and all peptide sequences are searched against a database simultaneously. Because the program utilizes amino acid sequence, it is error tolerant (Mackey et al., 2002). This is important because sequencing errors, polymorphisms, and conservative amino acid substitutions often can prevent protein identification with other search methods. A schematic of how the FASTS program functions is shown in Figure 4. The FASTS program is available to the public at http://fasta.bioch.virginia.edu/.

V. Phosphoproteome Analysis

Much of proteomics has centered around the quantitation of differences in protein expression between two samples. However, a comparison of protein expression levels provides only one aspect of protein regulation. Proteins also undergo a large number of post-translational modifications that can affect...
function. Reversible protein phosphorylation is one of the most widespread protein modifications. Thus, methods must be devised to study the phosphoproteome, defined as all phosphorylated proteins in the cell. Here, we summarize recent advances and outline some of our own strategies for analysis of the phosphoproteome.

A. PHOSPHOPROTEOME ANALYSIS BY PROTEIN RADIOLABELING

Protein labeling with either inorganic phosphate ($^{32}$P) or $\gamma$-$^{32}$P-ATP is still one of the most practical ways to study protein phosphorylation. In proteomics, a common approach is to label cells in vivo with $^{32}$P, then resolve the $^{32}$P-labeled proteins by 2-DGE. The major drawback to this approach is that although many phosphoproteins can be visualized by autoradiography, they are not present in sufficient amounts to be identified. To overcome this problem, we utilize methods to enrich for specific subproteomes prior to analysis by electrophoresis (Figure 5).

1. Analysis of Protein Phosphorylation in Vivo

In this approach, cells are labeled in vivo with $^{32}$P-orthophosphate, treated in some way, then cell extracts are prepared. The cell extracts are applied to the affinity matrix ATP-Sepharose for capture of the purine-binding proteome. Following washing to remove nonspecific proteins, ATP-binding proteins (including protein kinases) are recovered by elution with ATP and then further resolved by ion-exchange chromatography. Protein phosphorylation is visualized by electrophoresis of fractions on 1-DGE followed by autoradiography. Proteins that undergo changes in phosphorylation in response to the treatment are sequenced and identified by MS (Figure 5A). In an alternative approach, radiolabeled lysates are fractionated by ion-exchange chromatography, then directly analyzed by 1-DGE or 2-DGE.

2. Protein Kinase Profiling in Vitro

Protein kinase profiling is a way to identify protein kinases and their substrates under different biological conditions. In this strategy, a cell extract is fractionated by ATP-Sepharose or ion-exchange chromatography and the individual fractions are incubated with kinase buffer containing $\gamma$-$^{32}$P-ATP to allow labeling of phosphorylated proteins (Figure 5B). Individual fractions are resolved by 1-DGE and proteins that undergo changes in phosphorylation are selected by autoradiography. If specific proteins undergo changes in phosphorylation, they are excised from the gel and identified by MS. If individual proteins cannot be identified in the fraction, then the entire fraction can be reapplied to another ion-exchange column under different conditions. Alternatively, the fraction can
be applied to γ-ATP-Sepharose to isolate protein kinases. The advantage to this approach is that a large number of samples can be screened easily and quickly for changes in protein phosphorylation and both the substrates and the protein

FIG. 5. Strategies for phosphoproteome analysis. (A) Analysis of in vivo protein phosphorylation. In this approach, cells are grown in the presence of 32P-inorganic phosphate to allow labeling of all phosphoproteins. Following cell treatment, cell extracts are prepared and the proteins fractionated by chromatography to enrich for specific types of proteins. Proteins are resolved further by electrophoresis and changes in phosphorylation are visualized by autoradiography. Proteins that undergo changes in phosphorylation are excised and sequenced by MS. (B) Analysis of in vitro protein phosphorylation. In this approach, a protein extract is incubated with γ-32P-ATP or, alternatively, the proteins are fractionated and the individual fractions incubated with γ-32P-ATP. This approach can allow the identification of protein kinases and their substrates under different conditions.
kinases can be identified. One caveat to this approach is that because the phosphorylation is conducted \textit{in vitro}, the results must be confirmed by \textit{in vivo} studies.

B. METHODS FOR PHOSPHOPROTEOME ISOLATION

A major obstacle in the study of phosphorylated proteins is that phosphoproteins comprise only a small fraction of the total protein in a cellular lysate. As a result, many phosphoproteins cannot be identified in a cell extract. To address this problem, methods have emerged to enrich for phosphorylated proteins from a complex mixture of proteins. These techniques may facilitate the isolation and study of the entire phosphoproteome under different biological conditions.

1. Immobilized Metal Affinity Chromatography

Immobilized metal affinity chromatography (IMAC) has been used to purify phosphopeptides (Andersson and Porath, 1986) and, more recently, has been combined with MS to identify phosphoproteins from cell lysates (Posewitz and Tempst, 1999; Cao and Stults, 2000; Stensballe \textit{et al}., 2001; Ficarro \textit{et al}., 2002). However, in addition to binding phosphopeptides, IMAC is known to bind nonphosphorylated, acidic peptides through their carboxylate groups. To overcome this problem, Ficarro and colleagues converted the carboxylate groups in peptides to their corresponding methyl esters. This step was reported to reduce the binding of nonphosphorylated peptides to the IMAC matrix (Ficarro \textit{et al}., 2002).

In the approach of Ficarro and colleagues, a protein extract is prepared from the sample and digested to peptides by the action of a protease. The resultant peptide mixture is treated with methanolic hydrochloric acid (HCl) to convert carboxylate groups to methyl esters and the mixture is applied to an IMAC column. The recovered peptides are separated further by high-performance liquid chromatography (HPLC) and analyzed online by an ion-trap mass spectrometer (Ficarro \textit{et al}., 2002). However, if one assumes that approximately 30\% of all proteins are phosphorylated (Hubbard and Cohen, 1993), then the total number of phosphoproteins in a cell lysate easily could exceed 1000. This approach requires that all phosphoproteins in a cell lysate (i.e., thousands of phosphoproteins) be analyzed to detect those that undergo changes.

2. Chemical Derivitization of Phosphorylated Amino Acids

One approach to isolate phosphorylated proteins or peptides is to take advantage of the unique chemistry of phosphoamino acids in peptides. Previously, Meyer and colleagues developed a method for the identification of phosphoserine in peptides by its conversion to S-ethylcysteine and Edman
sequencing (Meyer et al., 1991). Under alkaline conditions, phosphoserine will undergo β-elimination to dehydroalanine, which can be converted to S-ethylcysteine by Michael addition with ethanethiol. In our laboratory, we optimized conditions for the derivitization of phosphoserine with 1,2-ethanedithiol (EDT) for the purpose of fluorophore labeling of proteins (Fadden and Haystead, 1995). This same strategy can be used to attach biotinylated moieties to purify phosphoproteins or peptides. An example of this approach recently was provided by Oda and colleagues who, using the same chemistry described herein, converted phosphoserine residues to biotinylated residues in proteins. The biotinylated proteins were purified by avidin chromatography, digested to peptides, and identified by MS (Oda et al., 2001). In an alternate method, Zhou and colleagues devised a technique to isolate peptides phosphorylated on serine, threonine, and tyrosine residues (Zhou et al., 2001).

The main disadvantage of these two methods is that the current chemistries require significant amounts of protein or peptide for identification by MS to be successful. As a result, we have not found these methods to be a practical approach for isolation of phosphoproteins. In addition, the selectivity of these methods has not been confirmed. Several proteins were isolated using these methods that are not known phosphoproteins (e.g., glyceraldehyde 3-phosphate dehydrogenase). Therefore, further work needs to be done to confirm these results and to determine the specificity of this approach (Oda et al., 2001; Zhou et al., 2001).

C. IDENTIFICATION OF PROTEIN PHOSPHORYLATION SITES

In addition to identifying phosphoproteins, it is important to identify the phosphorylation sites in proteins, so that information can be obtained about the function of the phosphorylation event. The two predominant techniques for phosphorylation site analysis are Edman degradation and MS.

1. Phosphorylation Site Identification by Edman Degradation

Edman degradation is still one of the most practical methods to determine phosphorylation sites in peptides. This is because the technique is relatively simple, is very sensitive, and can be applied to a large variety of peptides (Aebersold et al., 1991; Boyle et al., 1991; Wettenhall et al., 1991). If enough radioactivity can be incorporated into the phosphoprotein of interest, sites can be determined at the sub-fmol level. In our hands, this can be as little as 1000 cpm (not ideal). In this approach, a 32P-labeled protein is digested with a protease and the resulting phosphopeptides are purified by reverse-phase HPLC or thin-layer chromatography (TLC) (Figure 6). The isolated peptides then are cross-linked via their C-termini to an inert membrane and the radioactive membrane is subjected to several rounds of Edman cycles. The radioactivity is collected after each
FIG. 6. Strategies for phosphorylation site identification. (A) Proteins that are labeled with $^{32}$P are resolved by electrophoresis and converted to peptides by the action of proteases. (B) Peptides are resolved by high-performance liquid chromatography (HPLC) and phosphorylated peptides are selected by measurement of radioactivity. (C) Phosphorylated peptides are subject to Edman degradation and the resultant fractions are monitored for release of $^{32}$P. (D) Analysis of phosphorylated peptides by MS/MS. The location of phosphoserine in a peptide can be determined by obtaining amino acid sequence and characterizing the mass difference between peptide fragments. (E) Cleaved radioactive peptide analysis is a bioinformatics program that assists the assignment of phosphorylation sites in peptides.
cleavage step and the released $^{32}$P is measured in a scintillation counter. This method positionally places the phosphoamino acid within the sequenced phosphopeptide (Figure 6). Of course, this is meaningful only if the sequence of the phosphopeptide is known. The analysis ceases to become quantitative beyond 30 Edman cycles (even with efficient, modern Edman machines) due to well-understood issues with repetitive yield associated with Edman chemistry.

Recently, our laboratory has extended the usefulness of phosphorylation site characterization by Edman chemistry through the development of the cleaved radioactive peptide (CRP) program (MacDonald et al., 2002). In CRP analysis, one requires only that the sequence of the protein be known. Purification and sequencing of individual peptides is not required. Radiolabeled proteins (isolated following immunoprecipitation from $^{32}$P-labeled cells, for example) are cleaved at predetermined residues by the action of a protease. The phosphopeptides are then separated by HPLC or TLC (or, if only one site is present, no peptide separation is required), cross-linked to the inert membrane, and carried through 25–30 Edman cycles. The sequence of the target protein is entered into the CRP program. This program predicts how many Edman cycles are required to cover 100% of the serines, threonines, and tyrosines from the site of cleavage. Generally, one round of CRP analysis narrows the number of possible sites to 5–10 for most proteins. Phosphoamino acid analysis can be used to reduce the number of possibilities still further. The CRP analysis is repeated following cleavage with a second protease (usually one cutting at R but M and F are alternatives). The second round of CRP usually unambiguously localizes the phosphoamino acid to one possible site. The technique does not work if sites are greater than 30 amino acids away from all possible cleavage sites. The finding that CRP analysis is not applicable may, in itself, confine a phosphorylation site to a segment of the protein that is likely to produce very large proteolytic fragments. The CRP program is accessible at http://fasta.bioch.virginia.edu/crp/ and was written in collaboration with Aaron Mackey and Bill Pearson of the University of Virginia (MacDonald et al., 2002).

2. Phosphorylation Site Identification by MS

Phosphorylation sites in peptides can also be analyzed by MS/MS. In this approach, the phosphopeptide is sequenced in the mass spectrometer and the site of phosphorylation is determined unambiguously (Figure 6). Phosphopeptides can be identified from a mixture of peptides by a method known as precursor ion scanning (Neubauer and Mann, 1999). Serine and threonine phosphorylated peptides undergo loss of 98 Da ($H_3PO_3$) and 80 Da ($HPO_3$) from the molecular ion. These ions can be used to identify phosphorylated peptides. Having identified a phosphopeptide, the peptide mixture is sprayed under acidic conditions and the phosphopeptide sequenced by conventional MS/MS. Upon fragmentation of
the phosphopeptide, phosphoserine can be identified by the formation of dehydroalanine (69 Da), the β-elimination product of phosphoserine. Similarly, phosphothreonine can be identified by the formation of its β-elimination product, dehydroamino-2-butyric acid at 83 Da (Neubauer and Mann, 1999).

VI. Proteomics Applications

A. YEAST GENETICS AND PROTEOMICS TO ANALYZE SIGNALING PATHWAYS

One application of proteomics that holds great potential is the analysis of signal transduction pathways in yeast. Proteomics techniques, when combined with yeast genetics, can allow dissection of signaling events at the molecular level. In our laboratory, we have combined proteomics and yeast genetics to identify physiological targets of protein phosphatases. In S. cerevisiae, protein phosphatase-1 (Glc7p) and its binding protein, Reg1p, are essential for regulation of glucose-repression pathways. However, little was known about the physiologic substrates of this phosphatase. Therefore, in a collaborative effort with Dr. Marion Carlson, we undertook a proteomics approach to identify the physiological substrates of the phosphatase (Alms et al., 1999). To accomplish this goal, we examined the effects of deletion of the REG1 gene on the yeast phosphoproteome. Analysis by two-dimensional phosphoprotein mapping identified two distinct proteins that were greatly increased in phosphate content in the reg1 deletion mutants. Peptide microsequencing identified these proteins as hexokinase II (Hxk2p) and the E1α subunit of pyruvate dehydrogenase.

We went on to validate these findings in a comprehensive biochemical study. Consistent with increased phosphorylation of Hxk2p in response to REG1 deletion, fractionation of yeast extracts by anion-exchange chromatography identified a Hxk2p phosphatase activity in wild-type strains that was selectively lost in the reg1Δ mutant. Having carried out these studies, we attempted to rescue the reg1Δ phosphoprotein phenotype by overexpressing both wild-type and mutant Reg1p in the deletion strains. Here, both the phosphorylation state of Hxk2p and Hxk2p phosphatase activity were restored to wild-type levels in the reg1Δ mutant by expression of a LexA-Reg1p fusion protein. In contrast, expression of a LexA-Reg1p protein containing mutations at phenylalanine in a putative PP-1C (the catalytic subunit) binding site motif (K/R)(X)(I/V)XF was unable to rescue Hxk2p dephosphorylation in intact yeast or restore Hxk2p phosphatase activity. These results demonstrate that Reg1p targets PP-1C to dephosphorylate Hxk2p in vivo and that the peptide motif (K/R)(X)(I/V)XF is necessary for its PP-1 targeting function. These studies, therefore, demonstrate how a proteomics approach can be used to first identify enzyme targets in cells, then direct further analysis to verify the findings. Clearly, a combined proteomics
and genetics approach greatly enhances one’s ability to directly answer key biological questions. We believe that a similar strategy could be adopted with transgenic or knockout mouse work, particularly in cases where there is no obvious phenotype.

B. SMOOTH MUSCLE CELL PHYSIOLOGY

Proteomics, when combined with cell physiology, can be a valuable tool to understand basic cellular processes. In our laboratory, we apply proteomics techniques to study the regulation of smooth muscle contraction (SMC). SMC is regulated principally by free intracellular calcium levels and the level of myosin light-chain phosphorylation (Sellers and Adelstein, 1985). The steady-state level of myosin light-chain phosphorylation is, in turn, regulated by the relative activities of myosin light-chain kinase and myosin light-chain phosphatase (SMPP-1M). SMPP-1M itself is regulated by phosphorylation and its activity can be inhibited by phosphorylation of its targeting subunit. However, until recently, the protein kinase responsible for this phosphorylation was unknown. Therefore, we utilized a proteomics strategy to identify the relevant endogenous kinase responsible for phosphorylation and inhibition of SMPP-1M.

Because γ-ATP-Sepharose is capable of binding protein kinases (Haystead et al., 1993), we applied the entire SMC complement of proteins to γ-ATP-Sepharose. Using this affinity purification, we successfully obtained amino acid sequence for the previously unidentified MYPT1-kinase (MacDonald et al., 2001a). Further experiments, including addition of recombinant enzyme to permeabilized SMC, confirmed a role for the MYPT1 kinase as a regulator of SMC contractile state (MacDonald et al., 2001a,b; Borman et al., 2002).

We currently are using a functional proteomics approach to identify novel protein kinase substrates that are phosphorylated in response to agonist-induced smooth muscle relaxation. In our approach, permeabilized smooth muscle strips, contracted by treatment with calcium, are treated with 8-bromo-cGMP to induce relaxation. Cell extracts are prepared and proteins are resolved by 2-DGE. Proteins that undergo changes in phosphorylation in response to 8Br-cAMP are sequenced and identified by MS. Using this method, several known phosphoproteins — including HSP20, HSP27, and telokin (MacDonald et al., 2000; Walker et al., 2001; Borman et al., 2002) — were identified. In addition to these known proteins, we have identified novel phosphoproteins. Their role in smooth muscle relaxation is being investigated (J.A. MacDonald and T.A. Haystead, unpublished results).

C. PROTEOME MINING

Proteome mining is an iterative process in which a targeted proteome is screened against compound libraries to identify highly selective drug-protein interactions (Graves and Haystead, 2002). The advantage of this approach is
the ability to identify a drug-like molecule from a compound library that is specific for a protein or group of proteins and simultaneously identify the protein targets of the drug. The process begins with the isolation of a specific subproteome from a cell, tissue, or animal source by application of saturating amounts of tissue extract to an affinity matrix (Figure 7). The affinity matrix is composed of a natural ligand immobilized in an orientation that favors interaction with its respective protein targets. For example, if the natural ligand is ATP, only those proteins that bind ATP or ATP-resembling molecules will be captured on the matrix. This can be validated by sequencing and identifying proteins that are specifically absorbed to the matrix. The captured proteome then is screened against large chemical libraries and, if a compound from the library is able to compete with the immobilized ligand for protein binding, the bound protein is displaced from the matrix and recovered. Since the drug has the potential to interact with all of the proteins bound to the matrix, information about drug specificity can be obtained by identifying the eluted proteins.

FIG. 7. Proteome mining is a method to screen combinatorial compound libraries for protein targets and gain information about compound specificity. In this strategy, proteins from a cell line, tissue, or organism are isolated on identical affinity column arrays composed of natural ligands, drugs, or inhibitors designed for the capture of specific types of proteins. After washing to remove nonspecific proteins, the arrays are exposed to a compound library and proteins that are eluted are resolved and identified by MS. If the protein is relevant, a sublibrary can be applied to the bound proteome to further define the lead molecule.
We recently applied proteome mining to investigate the protein targets of antimalarial drugs. Malaria is responsible for an estimated 1–3 million deaths annually. Because parasites have evolved resistance to many common antimalarial drugs, there is a great demand for development of novel therapeutics (Foley and Tilley, 1998). In our approach, we captured the purine-binding proteome of human red blood cells, *Plasmodium falciparum*, on γ-ATP-Sepharose and screened the bound proteomes against several of the quinoline antimalarial drugs (e.g., chloroquine, primaquine). Interestingly, several quinoline antimalarial drugs eluted two proteins from the human red blood cell purine-binding proteome identified as aldehyde dehydrogenase and quinone reductase 2 (Graves et al., 2001). We are investigating the role of these proteins in the pathology of malaria. Thus, proteome mining can allow the identification of drug targets and further investigation of those targets can reveal the molecular action of the drugs in vivo.

ACKNOWLEDGMENTS

We thank Elizabeth Herrick for figure making and design and Meredith Borman and Justin MacDonald for critical review of the manuscript.

REFERENCES


Alms GR, Sanz P, Carlson M, Haystead TA 1999 Reg1p targets protein phosphatase 1 to dephosphorylate hexokinase II in *Saccharomyces cerevisiae*: characterizing the effects of a phosphatase subunit on the yeast proteome. EMBO J 18:4157–4168


Anderson L, Porath J 1986 Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. Anal Biochem 154:250–254


MacDonald JA, Borman MA, Muranyi A, Somlyo AV, Hartshorne DJ, Haystead TA 2001a
Identification of the endogenous smooth muscle myosin phosphatase-associated kinase. Proc
Natl Acad Sci USA 98:2419–2424

MacDonald JA, Eto M, Borman MA, Brautigan DL, Haystead TA 2001b Dual Ser and Thr
phosphorylation of CPI-17, an inhibitor of myosin phosphatase, by MYPT-associated kinase.

MacDonald JA, Mackey AJ, Pearson WR, Haystead TAJ 2002 A strategy for the rapid
identification of phosphorylation sites in the phospho-proteome. Mol Cell Proteomics 1:314–
322

MacDonald JA, Walker LA, Nakamoto RK, Gorenne I, Somlyo AV, Somlyo AP, Haystead TA
2000 Phosphorylation of telokin by cyclic nucleotide kinases and the identification of in vivo

Mackey AJ, Haystead TAJ, Pearson WR 2002 Getting more from less: algorithms for rapid protein
identification with multiple short peptide sequences. Mol Cell Proteomics 1:139–147

Mann M, Wilm M 1994 Error-tolerant identification of peptides in sequence databases by peptide

Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame
AL 2000 The characteristics of peptide collision-induced dissociation using a high-perform-

Meyer HE, Hoffmann-Posorske E, Heilmeyer LM Jr 1991 Determination and location of
phosphoserine in proteins and peptides by conversion to S-ethylcysteine. Meth Enzymol
201:169–185

Neubauer G, Mann M 1999 Mapping of phosphorylation sites of gel-isolated proteins by nano-

Oda Y, Nagasu T, Chait BT 2001 Enrichment analysis of phosphorylated proteins as a tool for

Opitcek GJ, Lewis KC, Jorgenson JW, Anderegg RJ 1997 Comprehensive on-line LC/LC/MS of


Posewitz MC, Tempst P 1999 Immobilized gallium(III) affinity chromatography of phosphopep-

Qin J, Fenyo D, Zhao Y, Hall WW, Chao DM, Wilson CJ, Young RA, Chait BT 1997 A strategy
for rapid, high-confidence protein identification. Anal Chem 69:3995–4001

Schena M, Shalon D, Davis RW, Brown PO 1995 Quantitative monitoring of gene expression
patterns with a complementary DNA microarray. Science 270:467–470

Sellers JR, Adelstein RS 1985 The mechanism of regulation of smooth muscle myosin by

Shalon D, Smith SJ, Brown PO 1996 A DNA microarray system for analyzing complex DNA
samples using two-color fluorescent probe hybridization. Genome Res 6:639–645

Shevchenko A, Loboda A, Ens W, Standing KG 2000 MALDI quadrupole time-of-flight mass

Stensballe A, Andersen S, Jensen ON 2001 Characterization of phosphoproteins from electro-
phoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry
analysis. Proteomics 1:207–222

Science 270:484–487

Site-specific phosphorylation and point mutations of telokin modulate its Ca2+-desensitizing


The Use of DNA Microarrays to Assess Clinical Samples: The Transition from Bedside to Bench to Bedside

JOHN A. COPLAND,* PETER J. DAVIES,† GREGORY L. SHIPLEY,‡ CHRISTOPHER G. WOOD,‡ BRUCE A. LUXON,§ AND RANDALL J. URBAN*

*Endocrinology Division, Internal Medicine, University of Texas Medical Branch, Galveston, Texas 77555; †Department of Pharmacology, University of Texas-Houston Medical School, Houston, Texas 77030; ‡Urology Department, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030; §Human Biological Chemistry & Genetics Departments, University of Texas Medical Branch, Galveston, Texas 77555

ABSTRACT

The advent of gene array technology brings the ability to classify disease states to the molecular level by examining changes in all mRNAs expressed in cells or tissues. Comparing changes in gene expression patterns between normal and diseased cells and/or tissues has elucidated unique subsets of genes identifiable to a specific disease. Already, new subclassifications of specific cancers have been discovered, belying that genomic profiling can uniquely distinguish a specific disease state and tissue of origin. This technology bestows the ability to examine global changes occurring in a cell or tissue(s), thereby allowing the elucidation of alterations in dysregulated biological, biochemical, and molecular events leading to disease states such as diabetes, hypertension, infertility, obesity, osteoporosis, and atherosclerosis. Furthermore, genomic profiling will lead to new molecular targets for the development of drug therapeutics. Futuristically, one could envision personalized patient therapies based upon identification of specific aberrant signaling pathways that can be targeted for drug therapy.

I. Introduction

The introduction of gene array technology in the mid 1990s (Schena et al., 1995; DeRisi et al., 1996) has led to a phenomenal ability to develop new subclasses of common diseases, predict disease outcome, and identify novel molecular targets for potential drug therapy. DNA arrays allow for simultaneous quantitative measurement of mRNA expression by thousands of genes in a biological sample. It is expected that this “gene profiling” will play a key role in understanding drug side effects at the molecular level and, one day, provide the rationale for individualized drug therapies for each patient, providing effective treatment and decreased side effects. Thus, clinicians with valuable patient tissue and blood samples are needed to use gene array technology to help create these novel molecular profiles, especially for endocrine-linked diseases such as diabe-
tes, hypertension, multiple endocrine neoplasia (MEN) syndromes, other cancers, infertility, obesity, osteoporosis, atherosclerosis, and leiomyomas (Korach, 2002). Large biological sample numbers will be required to develop precise gene profiles.

As shown in Figure 1, clinical applications demonstrate the uses of gene array data in which multivariate analyses via gene array detection supercede the conventional gene-by-gene analysis approach that is limited by biological insight. This systemic approach allows for elucidation of molecular complexities in alterations in signal transduction pathways that alter disease processes. Thus, the overall goal should be to complete construction of the roadmap identifying each molecule in all signaling pathways in each and every cell type known to regulate cellular functions as well as to characterize between signaling pathways (for 14 elegant reviews demonstrating signal transduction pathways important to cell function, see Science 296:1632–1657). Unique alterations in the roadmap should be predictive of specific diseases and the phenomena of different endocrine-driven stages of life, such as onset of puberty, pregnancy, and aging.

The premise of gene microarray technology is that steady-state mRNA levels are altered in diseased cells/tissues. These alterations determine changes in function and phenotype. Thus, it would be expected that tens to hundreds of genes out of the total expressed genome would be altered, depending upon the disease state. Therefore, it should be expected that multiple alterations within a single signaling pathway would occur as well as among signaling pathways due to changes in the dynamics of crosstalk between paths. The premise further dictates that, for each specific disease state, a unique subset of mRNAs will be

![Diagram](image_url)

**FIG. 1.** Clinical applications of microarray technology.
altered, creating a signature or fingerprint pattern that identifies that disease as well as the cell type of origin. In the cancer arena, this has proven true: gene array profiles have been performed upon different types of cancers, each replicating a signature pattern correlating to tissue of origin (Ross et al., 2000; Ramswamy et al., 2001; Nielsen et al., 2002). Moreover, new cancer subclassifications have been discovered using gene profiling for lung adenocarcinoma, breast carcinoma, leukemia, lymphoma, melanoma, and colon carcinoma (Khan et al., 1998; Golub et al., 1999; Alizadeh et al., 2000; Ross et al., 2000; Garber et al., 2001; Notterman et al., 2001; Sorlie et al., 2001; Armstrong et al., 2002). This strategy—combined with clinical parameters—is leading to well-defined disease classifications and the ability to predict clinical outcome (Bhattacharjee et al., 2001; Lakhani and Ashworth, 2001; van’t Veer et al., 2002).

While gene array technology represents a marvelous opportunity to meld the clinical researcher with the basic scientist, three significant issues must be considered when attempting to derive meaningful array data from a clinical research protocol. These problems are 1) the quality and amount of tissue sample from which the cDNA for hybridization is derived, 2) the type of array to be used for the tissue derived-cDNA, and 3) the extent and nature of the data analysis.

II. Tissue Samples

The source of human mRNA remains a major challenge, slowing widespread use of gene array technology in clinical research. The two major problems with obtaining tissue are the amount that can be obtained (which limits the quantity of mRNA that can be isolated) and the complexity of the tissue sample itself. Therefore, clinical studies that are published using DNA gene arrays are performed in settings where tissue/blood are abundant and readily obtainable. Typically, 1–2 micrograms of mRNA or 10 micrograms of total RNA are required for an oligo or cDNA array. Nylon microarrays, which use radioactive detection, need only nanogram quantities of mRNA (Bertucci et al., 1999). Another solution is to subject the sample mRNA to linear amplification methods, which require \( \leq 50 \text{ ng} \) of mRNA for gene array analysis (Van Gelder et al., 1990; Phillips and Eberwine, 1996; Wang et al., 2000). Ambion (Austin, TX) sells a kit based upon Eberwine’s linear amplification of mRNA into antisense RNA. This is an important development, since many clinical samples are small.

Although nearly all clinical gene array analyses have been performed on cancer specimens, to profile differences between tumor and normal tissues as well as between different tumors, we have included other examples of clinical samples that have utilized gene arrays. The reader is referred to the June 2002 issue of Endocrinology, which highlights the impact of the human genome upon endocrinology. Gene expression patterns were measured in subjects with scleroderma from inflammatory cells obtained by bronchoalveolar lavage (Luzina...
et al., 2002). Circulating leukocytes and peripheral blood mononuclear cells have been used to assess kidney diseases (Alcorta et al., 2002), expression of cytokine- and chemokine-related genes in lupus patients (Rus et al., 2002), and gene expression profiles of mononuclear cells in humans after infection with human immunodeficiency virus (HIV) type 1 RF (Vahey et al., 2002). In some clinical settings, tissue may be readily available— for example, gene expression markers that have been measured in subjects with endometriosis (Eyster et al., 2002), from osteoarthritic cartilage (Aihara et al., 2002), and from brain tissue with gene expression patterns in schizophrenia (Mimmack et al., 2002).

The tissue itself is a second problem in determining the value of gene array data. Processing tissue rapidly to maintain RNA integrity is crucial. Artifactual gene array data are generated from degraded mRNA. Therefore, having access to a competent tissue bank linked to searchable databases that contain the clinical, biological, and biochemical characteristics of the sample is key to obtaining meaningful diagnostic interpretations from integration of these clinical correlates and reliable gene array data.

Most tissue samples obtained from humans are a mixture of different cell types. For example, a muscle biopsy sample taken from the vastus lateralis muscle will contain not only skeletal muscle but also blood vessels, connective tissue, nerve tissue, and stromal cells. Therefore, changes in gene expression patterns, when comparing two different muscle biopsy samples, are a reflection of all the cell types present in that sample. Many claim that this can confound the analysis and limit applicability of results. Methods such as laser capture microdissection that allow for isolation of individual cells still are limited technologically (Simone et al., 1998; Brail et al., 1999; Luo et al., 1999; Best and Emmert-Buck, 2001). However, many others argue that all the cell types influence the function of the tissue in question and the gene expression pattern as a composite of the whole is more meaningful than any one isolated cell type. Clearly, this is an area of active debate. Yet, virtually everyone agrees that altered signaling and interactions between cell types are informative and diagnostic of a specific disease. In fact, this has proven true in profiling (Eisen and Brown, 1999; Young, 2000; Ramaswamy and Golub, 2002). The significance of this discussion can be determined only through comparing array data between individual cell types and the tissue as a whole (Alizadeh et al., 2001).

III. Type of Array

Another important decision that a clinical investigator must make is choosing the type of array to use with the clinical sample (Figure 2). The technique is based upon the same principle as Northern and southern blotting, in which a labeled (radioactive or fluorescent) complementary DNA (cDNA) probe has been created from reverse transcription of mRNA and hybridized to cDNA or short
oligonucleotides representative of unique gene sequences immobilized on either a nylon membrane or glass slide (in excess of mRNA). Early on, only a few arrays were available. However, as the field has advanced, more-specific arrays that contain focused genes are being developed. A recent study identified Apo L proteins as a group that shows increased expression in the brain tissue of humans with schizophrenia (Mimmack et al., 2002). This finding was made using a custom candidate gene cDNA array comprising 300 genes that were implicated in schizophrenia (Mimmack et al., 2002). On the other end of the scale from

FIG. 2. Oligonucleotide and cDNA microarrays. Oligonucleotide microarrays utilize a direct synthesis of oligonucleotides (25–60 mer) onto solid surface with a single-color readout of gene expression from a patient sample. A cDNA microarray utilizes polymerase chain reaction (PCR)-generated products (1–2 Kb) from cDNA libraries onto a solid surface with simultaneous, two-color readout (Cy3 and Cy5) of gene expression from patient sample and a reference standard sample.
custom cDNA arrays are commercial oligonucleotide arrays such as the Human Genome U133 GeneChip® produced by Affymetrix (Santa Clara, CA) that contains 39,000 of the best-characterized human genes (Gershon, 2002). Clinical investigators must decide between a specialized array and a more-generalized one that would be more inclusive but also generate many more data for interpretation. Cost is another consideration. A number of excellent reviews describe design and benefits of various arrays (Alizadeh et al., 2001; Bertucci et al., 2001; Arcellana-Panlilio and Robbins, 2002; Grant et al., 2002).

IV. Gene Array Data Analysis

As detailed above, many methods are available to allow investigators to begin to interpret the significance of the array data. Moreover, investigators must validate key gene expression patterns using methods such as Northern blotting, quantitative reverse transcription-polymerase chain reaction (QRT-PCR), or RNAse protection assays. However, most clinical investigators must collaborate with 1) a basic scientist, to generate the array data, and 2) bioinformatics experts, to analyze the array results and indicate significant gene expression patterns. As the field continues to develop, this process will become more automated and, with this progression, clinical investigators without access to the multiple groups of investigators found at major medical institutions will begin to use array technology. An example of this continuing automation of a DNA analysis device that will both synthesize oligonucleotide probes and perform hybridization is being tested by Geniom, a German biotechnology company (Gershon, 2002).

A detailed analysis of differential expression, including clustering and profiling, requires a rich data field. Therefore, experiments should be set up to construct a data matrix or “dataframe” consisting of a row for each gene and a column for each chip. The first column contains the identification of the gene in each row. To achieve the best results when clustering, it is useful to have at least four columns of array data. Because very large volumes of data are generated by microarray experiments, it is important to select for analysis only those genes that appear to show differential expression due to the experimental conditions. If this is not done, the large number of genes whose expression was either not changing or was due to random fluctuations are likely to wash out many important experimental effects. To filter out these genes, a strategy should select only those genes fit for further analysis. Good experimental design makes this easier to do in a systematic manner (Tian et al., 2002). Typically, we set up our experiments to be suitable for either a one-way or two-way analysis of variance (ANOVA) filter (Figure 3). Factorial designs are very useful for microarray experiments and often fit the experimental situation very well. Note that a $2 \times 2$ factorial design can be satisfied with as few as four GeneChips and is ideal for a two-way ANOVA (though at least three replicates are preferred for statistical reasons) and
construction of a good dataframe. For Affymetrix arrays, we first filter out genes that the initial analysis rates as “Absent” in each of the GeneChips. These are discarded as uninteresting (i.e., unresponsive) to this set of experiments. Next, we perform the ANOVA separately for each gene, keeping only those that show a probability that the F-ratio (Pr(F)) is significant at some level of confidence such as 95% (i.e., having a Pr(F) value ≤ 0.05). This means that the differential expression for that gene is likely due to the experimental conditions rather than to random fluctuations at that level of confidence. Genes that don’t meet this criterion are discarded from further analysis. The remaining genes are most likely...
to demonstrate responses that correlate in some manner with the experimental conditions and are most useful for further discovery. This method avoids the problems of multiple t-tests and is statistically more satisfying than simply requiring $\geq 3$-fold change as a cutoff. If no other method is available or suitable due to the experimental design, however, fold-change or log ratio cutoffs can be used to effectively filter out genes that appear to be unresponsive to the experimental conditions.

A primary goal of microarray analysis is discovery of hidden patterns of differential expression within the data field. Clustering methods especially are the tools of choice (Figure 3). This is usually best done iteratively using a partitioning method (e.g., k-means clustering, partitioning around medoids (PAM)). This will break the initial large group of genes into smaller subclusters, based on the similarity of their patterns of differential expression across all the experiments (GeneChips) simultaneously. Thus, a type of pattern-recognition algorithm groups genes because their expression is similar. Each gene within a subcluster will have enhanced or depressed expression in unison, when plotted across the different experimental conditions. It will be useful to give some of the details about how this is done because the underlying methodology is common to virtually all clustering techniques and it helps to understand why good experimental design resulting in a proper dataframe is usually the best approach.

A microarray dataframe can be described as an n-by-p matrix containing n rows corresponding to the "objects" (i.e., genes, probe sets) on the microarray and p columns, each corresponding to a separate microarray (or the average of a set of replicates) corresponding to a different experimental time or condition (e.g., tissue sample, temperature, dosage). Thus, each gene/probe set is identified with a row, $n_i$, which is a vector of order p where each point, $x_{ij}$, describes the fluorescent intensity of that probe set or gene in microchip experiment $p_j$. Associated with the n-by-p matrix of experimental measurements is an n-by-n table that is a collection of proximities describing the comparisons of all possible pairs of objects (i.e., genes). For the purposes of clustering microarray data, these proximities most often describe the dissimilarities between the differential expression patterns of two genes (or its conjugate, similarity) or covariance. Most simply, dissimilarity $d_{ij}$ can be explained as the Euclidean distance (i.e., crossproduct) between two vectors in a data set, each representing the differential expression of a particular gene (e.g., any two of the $n_i$). Using the $d_{ij}$ table, usually called a dissimilarity matrix, as input, a variety of clustering methods can be used to identify those objects (genes or probe sets) that behave most alike across a given set of experiments. Two of the more-common methods for doing this are k-means clustering and hierarchical clustering.

The k-means clustering continues iteratively, partitioning the genes into a growing number of smaller and smaller subclusters, until the pattern of expression for all members of the subcluster are not significantly dissimilar. Once the
subclusters are identified, they can be clustered hierarchically to show the juxtaposition of each gene within the subcluster, based upon the similarity of their differential expression patterns. This is shown graphically as a dendrogram (similar to a family tree). Hierarchical clustering techniques can be either agglomerative or divisive, depending upon whether they start with each of the member genes as an individual, then group them together into families, or whether they start with one large family and divide it up progressively into smaller and smaller subfamilies, until each gene is a separate branch. Usually, there are small qualitative differences between the results of the two methods but occasionally larger differences show up that need to be reconciled. We typically use an agglomerative nesting technique called AGNES. An often-insightful use of hierarchical clustering is to cluster the transpose of the dataframe associated with a specific subcluster. This shows the interrelationships between the columns (GeneChips) of the dataframe rather than the rows (genes). From this, we can see how genes within the subcluster differentiate the experiments. It is becoming increasingly common to perform hierarchical clustering for both the rows and columns and to show their dendograms on the same graphic aligned along the top and side of a “heat map” (Figures 4–7). A heat map is a graphical matrix where each cell corresponds to the signal intensity of a specific gene in a specific experiment. The rows and columns of a heat map are arranged to show simultaneously the interrelationship between the different experiments and the genes within the subcluster. The color of each cell is significant and is selected from a gradient of colors (typically, red to green), where the shade of the color is proportional to the signal intensity (or log ratio) of that gene in that experiment. These values often are normalized or scaled to z-scores for best effect. Normally, we represent high values as shades of red, intermediate values as shades of gray to black, and low values as shades of green. (A few authors do it the opposite way, so be sure to check the legend when reading articles containing heat maps.) Due to color limitations in this review, we have used white to represent high values, black to represent low values, and shades of gray to represent gradations of gene expression levels between high and low.

Another, oftentimes very useful way to generate the information for a heat map is called profiling. This method is used in Example 1 (see Section VA below) to identify genes commonly upregulated and downregulated by two thiazolidinedione compounds and genes uniquely regulated by these compounds (Figures 4–6). In this method, one constructs a model profile of how one would expect some genes to behave across the columns of a dataframe to demonstrate a pattern of meaningful biological significance. Each of the rows (genes) of the dataframe is ranked according to how similar its expression profile is to the model profile, then assigned a similarity coefficient between $+1$ and $-1$. A coefficient of $+1$ represents a perfect correlation, while $-1$ represents perfect anticorrelation. Zero means that there is no correlation. In practice, one chooses
genes that are highly correlated (and usually, highly anticorrelated) to the model profile, then constructs a heat map showing their hierarchical distribution among the genes and across the experiments. The choice of a suitable cutoff for the similarity coefficient is somewhat arbitrary but, similar to an R² correlation coefficient in regression analyses, absolute values below 0.70 often will not show good correlation with the model profile. This is something to trial and error on data using a spreadsheet and charting program.

Once a subset of genes that service a working hypothesis has been selected, it is very instructive to search out their functions (using PubMed, gene ontology entries (GO), etc.) and to characterize the subcluster for the similarity or
V. Practical Examples

A. EXAMPLE 1. DISTINGUISHING MOLECULAR ACTIONS OF DRUGS OF THE SAME CLASS, THEN LINKING THESE ACTIONS TO BIOLOGICAL ACTIVITY

Thiazolidinediones (Tzds) belong to a class of compounds that are ligand agonists for peroxisome proliferator-activated receptor gamma (PPARγ), a nuclear transcriptional factor. Tzds are clinically used in the treatment of type 2 diabetes as insulin sensitizers (Horikoshi and Yoshioka, 1998; Willson et al., 1998).
However, these compounds have different profiles with regard to side effects (Fujiwara and Horikoshi, 2002). PPARγ agonists are known to control adipocyte differentiation pathways via activation and suppression of key regulatory molecules determining adipocyte phenotype (Kliewer et al., 2002; Walczak and Tontonoz, 2002). Tzds also recently have been described to have antiproliferative as well as chemopreventive activity against human tumors (Debril et al., 2001; Sporn et al., 2001). We have found that troglitazone (TRO) and rosiglitazone (ROSI) have dissimilar profiles with regard to growth inhibitory profiles and induction of apoptosis in two human pancreatic tumor cells lines, Mia Paca-1 and Panc-1 (Cowey et al., 2001). Thus, we were interested in identifying and determining whether genes that control cell proliferation and cell death were regulated by these compounds. Cells were treated in culture with 20 μM of either TRO or ROSI for 4 and 24 hours, followed by RNA isolation and analysis of gene expression by Affymetrix analysis. The heat maps shown in
Figures 4–6 highlight the ability of TRO and ROSI to identify commonly and divergently regulated genes. Of the 12,558 probe sets on the human Affymetrix gene chip, 8249 genes were marked as “Present” (e.g., expressed). A pairwise comparison of control to treatment group using a 3-fold change cutoff retained 4158 probe sets. Using idealized profiles, we performed profile comparisons utilizing Spotfire DecisionSite 7.0 to identify correlates (upregulated genes) and anticorrelates (downregulated genes) with a similarity to the idealized profile of 0.7. The three profile searches performed are reflected in Figure 4 (genes commonly regulated by TRO and ROSI and different from their appropriate control), Figure 5 (genes altered uniquely by TRO), and Figure 6 (genes uniquely altered by ROSI). Thus, we were able to identify 42 genes commonly regulated by TRO and ROSI, 58 genes altered uniquely by TRO, and 39 genes altered by ROSI treatment at 4 and 24 hours.

The next challenge is to determine which of these genes are known to be regulated by Tzds. Bioinformatics programs written to query PubMed, Medline, and other databases can rapidly make the connection by probing the gene set identified against key words such as thiazolidinedione and PPARγ. Another

![Figure 7](image-url)
challenge is to identify functions of genes in a data set and link them to biochemical pathways that can be interpreted in a meaningful manner related to the demonstrated action of the compound: in our case, inhibition of cell proliferation (TRO and ROSI) and apoptosis (TRO). We have written a program that links gene name, GenBank number, chromosome location, and functions (biochemical, biological, organismal, and molecular; www.bioinfo.utmb.edu). Table I shows genes that play a role in inhibiting cell proliferation and stimu-

<table>
<thead>
<tr>
<th>Function</th>
<th>Gene</th>
<th>Troglitazone</th>
<th>Rosiglitazone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoptosis</td>
<td>Retinoic acid receptor alpha</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>Gadd45</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Cell cycle arrest</td>
<td>Protein kinase A inhibitor</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>Gas6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Antiapoptosis</td>
<td>Transforming growth factor beta-inducible early protein</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>++</td>
<td>0</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>DEC1</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Tumor suppressor</td>
<td>AD7c-neuronal thread protein</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Activating transcription factor-3</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>Zinc nuclear finger 10</td>
<td>++</td>
<td>+</td>
</tr>
</tbody>
</table>

[Genes were identified by cluster analysis or cutoff techniques described in the text for Mia Paca-1 (M) and Panc-1 (P) cells treated with TRO, ROSI, or vehicle control for 4 (04) or 24 hours (24). A program linked to PubMed/Gene Ontology was used to identify genes in this subset that regulate apoptosis and cell proliferation. Visualization of shades of gray from the heatmaps of Figures 4–6 correlates with expression levels of these identified genes. ++ = highly expressed (+ = increased expression), compared to matched untreated control, and correlates with white on heat map. 0 = no difference from control. – = a decrease in gene expression of treated group, compared to its matched control value. M = Mia Paca-1 cells; P = Panc-1 cells; 04 = 4 hours of treatment; 24 = 24 hours of treatment.]
lating apoptosis. For the first time, we demonstrated that retinoic acid receptor alpha (RARα) is regulated by Tzds. Also not previously identified to be regulated by Tzds are genes that are transcriptionally regulated by TGFβ and p53 (Conner et al., 1999; Kimura et al., 2001; McDonald and El-Deiry, 2001; Kondo et al., 2002; Yun et al., 2002; Zawel et al., 2002). Incidentally, Panc-1 and Mia Paca-1 cells express nonfunctional mutant forms of p53. Thus, we can identify multiple genes in a specific signaling pathway. Clearly, there is overlapping yet distinct regulation of these genes by TRO and ROSI. Thus, one can test the role of each of the newly identified genes by either overexpressing the respective gene in cells or by selectively blocking expression of each gene (antisense RNA, small interfering RNA (siRNA), or dominant-negative expression constructs). Yet another challenge is to identify unknown genes or expressed sequence tags (ESTs). Multiple strategies can be taken in this respect using data search bases and programs. Using the Genbank number for the unknown gene, a DNA sequence can be copied and pasted into programs (e.g., www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html) that will perform searches for homologies to the unknown gene.

B. EXAMPLE 2. IDENTIFYING GENE SIGNATURES RELATED TO TUMORIGENESIS IN PATIENTS AND DEMONSTRATING UNIQUE GENE EXPRESSION PATTERNS IN EACH PATIENT

In Figure 7, a heat map demonstrates the ability to identify differences in gene expression between normal tissue and that of matched primary tumor and metastasis in three patients diagnosed with metastatic clear cell renal cell carcinoma (RCC). Hierarchical clustering was used to determine whether genes were differentially expressed in tumors versus normal kidney tissue. As shown in Figure 7, 253 genes were identified, clearly separating out gene expression of normal tissue of patients 1–3 (see right top of heatmap in Figure 7). Although genes were different in primary tumor and metastasis samples, compared to their respective controls, few genes delineated a difference between the two groups. This had been observed previously by others; it was suggested that the primary tumor already had been programmed to metastasize and that gene expression levels will be similar in the primary tumor and metastasis tissues (Bhattacharjee et al., 2001). However, expression signatures are being used to diagnose cancer subtypes (van’t Veer et al., 2002). Golub and colleagues have designed supervised learning algorithms to classify tumor samples into specific diagnostic categories based on their gene expression signatures. They and others have demonstrated successfully that an unknown tumor sample can be classified correctly, based upon its signature (Khan et al., 2001; Ramaswamy et al., 2001; Yeang et al., 2001). Using these strategies, the number of genes needed to diagnose a disease can be narrowed down to only those that act as clear indicators.
of that disease state. Thus, the three-patient population sample shown in Figure 7 must be expanded by hundreds of matched normal and tumor samples in order to define the gene signature to appropriately define diagnosis. These data linked to patient outcome will lead to predicting prognoses based upon gene expression patterns. Even this small sample allowed us to identify dysregulation of multiple signaling pathways and multiple genes within the same signal pathway (data not shown). This has been verified by RT-PCR and immunohistochemical analysis in a larger set of matched tissue samples.

Another strategy that can be used to identify gene expression patterns that change in similar fashions is to profile a particular gene, then search the database for genes that have the same or opposite profile (anticorrelate). Again, genes that change in similar or diametrically opposite patterns may interact in a signaling pathway or coregulate a phenotypic change in the cell or tissue (Figure 8). Using the gene expression of pattern for E-cadherin (downregulated 50- to 100-fold in all primary and metastatic tumors, compared to matched normal controls) demonstrates the ability to identify genes in the data set from Figure 7 that are downregulated similarly to E-cadherin or are upregulated (anticorrelate). These strategies should allow one to begin to identify key signaling pathways and crosstalk between pathways in regulating proliferation, differentiation, and cell death (Figure 9). Filling in the pieces to this puzzle and understanding all the functions of each gene product will provide the roadmap for developing effective treatment regimens.

C. EXAMPLE 3. IDENTIFYING MUSCLE GENES REGULATED BY ANDROGENS IN OLDER ADULTS

Figure 10 displays an analysis of muscle biopsy samples taken from older men given testosterone injections for 6 months (Ferrando et al., 2002). This study found that administration to older men to increase their testosterone concentrations to those of younger men increased muscle strength, lean body mass, and the net balance of muscle metabolism (Ferrando et al., 2002). To further investigate the cause of the increase in strength, Atlas™ cDNA Expression Arrays made by CLONTECH Laboratories, Inc. (Palo Alto, CA) were hybridized with the tissue samples. The CLONTECH DNA array chip can determine the expression of gene targets from five major subgroupings. However, for this analysis, insufficient muscle biopsy tissue was available to analyze each sample. Therefore, for the baseline, 1-month, and 6-month time points, total RNA from the seven subjects that received testosterone for 6 months was pooled for the hybridization. Of the 2300 genes screened, significant changes were detected in 230. One group of genes that responded to testosterone was mitochondrial proteins involved in the oxidative phosphorylation pathway. Two of the protein complexes (IV and V) showed a response of the nuclear-encoded mitochondrial proteins. From these
data, three representative genes were selected for verification. Western blot analysis was performed on each of the seven subjects individually, with antibodies to the genes being verified. As shown in Figure 10, an increase in protein expression occurred at 1 month that returned to pretreatment levels by 6 months. Cytochrome c oxidase Vb is one of the smaller subunits of the cytochrome c oxidase complex that is nuclear encoded and is one of the key regulatory subunits of the complex (Basu et al., 1997; Lenka et al., 1998). Adenosine triphosphate (ATP) synthase γ forms the central core unit of the ATP synthase complex around which other subunits orient themselves (Konno et al., 2000; Tsunoda et al., 2001), including ATP synthase α. These mitochondrial oxidative phos-
phorylation proteins are nuclear encoded, indicating that they are responding to
the administration of testosterone in these older men. This pattern of expression,
which follows that of the androgen receptor (Ferrando et al., 2002), indicates that
further studies are needed to assess the effects of cycling testosterone on muscle
mass and strength.

VI. Verification of Gene Array Analysis

DNA microarray technology represents a powerful tool for expression
profiling and the scanning of very large numbers of genes for potential differ-
ences in gene expression. However, data generated by microarray-based studies have inherent limitations that complicate its use as a stand-alone technology. The problems with microarray-based studies stem largely from their dependence on differential hybridizations to detect differences in transcript abundance. Many factors may affect nucleic acid hybridization reactions. The standardized hybridization conditions of array experiments have been developed to minimize, on average, the impact of these factors. However, for any individual transcript, the hybridization conditions may not be optimal and apparent differences in gene expression may, on occasion, be biased by these effects. In some instances, it may be possible to use sufficient replications, coupled with statistical tools, to

FIG. 10. Western analysis of mitochondrial proteins during testosterone administration. Three nuclear-encoded mitochondrial proteins in the oxidative phosphorylation pathway were analyzed by western blots. The top panel is a representative subject and the graph below represents the mean ± standard error from the seven subjects. Actin (not shown) was used to standardize the blots. ATP, adenosine triphosphate; Cox Vb, cytochrome c oxidase Vb. *, p ≤ 0.05, as determined by ANOVA.

FIG. 10. Western analysis of mitochondrial proteins during testosterone administration. Three nuclear-encoded mitochondrial proteins in the oxidative phosphorylation pathway were analyzed by western blots. The top panel is a representative subject and the graph below represents the mean ± standard error from the seven subjects. Actin (not shown) was used to standardize the blots. ATP, adenosine triphosphate; Cox Vb, cytochrome c oxidase Vb. *, p ≤ 0.05, as determined by ANOVA.
demonstrate confidence in the results obtained (Lee et al., 2000; Tusher et al., 2001). More commonly, however, investigators resort to an independent measurement of transcript abundance (e.g., Northern blot, quantitative RT-PCR) to confirm their results. “Real-time” quantitative RT-PCR (q-PCR) has proven to be an invaluable tool not only for confirming microarray-based results but also for extending them into a more-general context.

A. REAL-TIME QUANTITATIVE RT-PCR

q-PCR is an ingenious technique that allows for the continuous measurement of products generated during the course of a multicycle PCR reaction (Gibson et al., 1996; Heid et al., 1996). A variety of experimental strategies are available that use changes in fluorescence emission intensities to monitor the progress of PCR reactions. The Taqman procedure, one of the earliest real-time q-PCR techniques developed, is based upon measuring the increase in fluorescence that results from the template-dependent hydrolysis of a fluorescent hybridization probe. This fluorescent hybridization probe is a sequence-specific oligonucleotide that is complementary to the amplicon being assayed. The probe contains two fluorescent dyes covalently attached to specific nucleotides such that fluorescence emission of the dye being excited is quenched by intramolecular fluorescence resonance energy transfer (FRET). During each PCR cycle, Taq polymerase hydrolyzes the hybridization probe molecules that are bound to the template, releasing the fluorescent dyes and dequenching the fluorescence. The increase in fluorescence is directly proportional to the number of probe molecules hydrolyzed that, in turn, under appropriate conditions, are directly proportional to the number of template molecules in the PCR reaction tube in that particular PCR cycle. During the geometric amplification phase of the PCR reaction, the amount of template generated is geometrically related to the number of template molecules present at the start of the reaction and the number of PCR cycles completed. Thus, by continuously monitoring the level of fluorescence in the PCR reaction in real-time PCR instrumentation, it is possible to accurately quantitate the number of templates present in the unknown sample at the beginning of the PCR reaction.

B. THE VALUE OF COMBINING q-PCR WITH MICROARRAY-BASED STUDIES

The value of real-time q-PCR as a technique for confirming microarray results stems from its ability to provide an accurate measure of the abundance of specific transcripts in an RNA preparation. Under most circumstances, microarray studies provide a semiquantitative estimate of transcript abundance and differential gene expression. Once a transcript of interest has been identified, it is relatively straightforward to design a specific real-time q-PCR assay for the
transcript, then use the assay to accurately measure the level of the transcript in the two RNA samples being compared. Not only will the technique allow for confirmation of an apparent differential level of expression but it also provides an accurate measure of the degree of differential expression. This is particularly important and useful in the analysis of transcripts that show modest levels of induction in the microarray studies. Transcripts that show large (i.e., > 3-fold) and consistent changes in expression in microarray studies generally are readily confirmed by an independent technique such as q-PCR, Northern blots, or RNase protection assays. However, many of the most important changes in gene expression, particularly in in vivo and clinical studies, are represented by more-modest changes (i.e., ≥ 1.5-fold) in groups of related genes such as those comprising key metabolic pathways. These levels of change often are difficult to distinguish from noise in the microarray studies but can be readily evaluated using q-PCR. The precision and reproducibility of q-PCR assays allow investigators to apply relevant statistical tests to the data to confirm the differential expression of genes with modest inductions. The combination of microarray and q-PCR measurements allows investigators to work with confidence in the “gray” zones of 1.5- to 3-fold changes, where most of the important biology is occurring (Singh and Liu, 2001).

A particular value of real-time q-PCR assays is that they are readily adapted to the analysis of multiple samples and can be run in a high-throughput mode. Since the progress of the PCR reaction is monitored optically, a variety of instruments have been developed that allow for simultaneous quantitation of transcripts in either 96-well or 384-well format. Thus, in one experiment, even allowing for replicates and controls, it is possible to quantitate transcripts for large numbers of samples (i.e., 20–200). Given that the usual run time for a real-time q-PCR reaction is less than 2 hours, this capability means that it is possible to quantitate many transcripts in many samples in a relatively brief period of time. This high-throughput capability is particularly useful in clinical studies. Very often, it is not logistically feasible to run microarrays on large numbers of individual patient samples. The clinical investigator may be restricted to running only a few samples, due to either limited availability of RNA requiring pooling of specimens or the expense of running large numbers of individual chips. To “validate” the results of the microarray study, it often is desirable to extend the analysis to a much larger pool of samples or subjects. We have found that real-time q-PCR analysis of a much larger series of subjects than could be included in the microarray study leads to a much more accurate and useful estimation of the extent to which the changes detected on the microarray can be applied to the patient population as a whole.

In addition to its suitability for high-throughput analyses, real-time q-PCR assays are particularly useful adjuncts to microarray studies. Their extreme sensitivity permits analysis of transcripts in very small amounts of input total
RNA. Under normal conditions, it is easy to develop assays with a lower limit of detection of $10^2$ to $10^3$ transcript molecules. This level of sensitivity permits detection of even low-abundance transcripts in nanogram quantities of total RNA. Since most microarray techniques perform best utilizing micrograms of RNA, it is often convenient to carry out the initial microarray experiments on pools of patient-derived samples, then switch to the much more-sensitive PCR-based technologies to confirm the array results in the panel of individual samples that contributed to the pool. In this way, it is possible not only to “confirm” the array result but also to acquire quantitative information on the distribution of differential gene expression in the patient population of interest.

A further useful aspect of the real-time q-PCR technique is that it does not require intact RNA to provide meaningful data on transcript abundance. The quality of data recovered from most microarray-based procedures depends heavily on the quality of the RNA used to generate cDNAs. On the other hand, it is possible to develop real-time q-PCR assays based on very short amplicons (60–80 nt). These amplicons will remain intact in RNA preparations that have been subjected to extensive degradation. This is particularly useful in a clinical context, where it is often difficult to control for the handling of biological specimens at the time of collection. This feature of real-time q-PCR assays can be applied to RNA recovered from formalin-fixed tissue blocks (Uray and Connelly, 2001). RNA fragments (usually several hundred nucleotides in length) can be extracted from formalin-fixed, paraffin-embedded tissue blocks such as those routinely maintained in pathology archives. This capability can provide a powerful complement to microarray-based studies. For instance, we have used RNAs derived from a limited number of surgical specimens with sufficient material to permit microarray-based analysis. We have then used real-time q-PCR assays to measure the expression of the transcripts identified as of potential interest by the microarray studies, in a much larger series of cases for which archival formalin-fixed specimens exist. An additional benefit of this approach is that it can be combined with laser capture microdissection techniques that allow for recovery of RNA from specific cellular subsets of diseased tissues. In this way, it is possible to extend the results of the microarray studies to a detailed analysis of the pattern of gene expression in large numbers of well-documented clinical cases.

In summary, we have found real-time q-PCR to be one of the most-useful approaches to first confirm and then extend the results obtained from microarray-based analyses.

C. APPLICATION PROCEDURES

1. Real-time q-PCR assays can be designed using target gene sequences accessible from genomic databases such as GenBank. Although a number of
primer design algorithms can be used for this purpose, we routinely employ Primer Express (Applied Biosystems), since it allows for the simultaneous design of both the PCR primers and fluorescent Taqman probe. We generally design our amplicons to be < 100 nt in length (usually, 60–80 nt). The PCR primers, the fluorescent probe (usually with a fluorescein amidite (FAM) reporter dye and either tetramethyl rhodamine (TAMRA) or a “black hole” quencher dye) and a single-stranded sDNA of the amplicon (a long oligonucleotide for use as a standard) are ordered from one of several commercial vendors.

2. Assay conditions are optimized with the sDNA amplicon standards by adjusting primer and Mg\textsuperscript{2+} concentrations to generate assays with a slope of the standard curve (Ct versus log template molecules) of $-3.2$ to $-3.5$ and a lower limit of detection of $10^2$ amplicon molecules.

3. Total RNA samples are assayed after DNAse I pretreatment. The range of RNA concentrations will vary based on the amount of material available and the anticipated transcript abundance. We routinely use 10–100 ng of total RNA per determination. Each RNA sample is assayed in triplicate, with a fourth aliquot that is run without reverse transcriptase (−RT control) to control for signal generated by genomic DNA rather than RNA.

4. Samples are subjected to RT prior to PCR amplification. Although we use the reverse PCR primer for the reverse transcriptase reaction, other investigators report equivalent success with random primed RT reactions.

5. In parallel with the unknown RNA samples, we run a standard curve with known amounts of amplicon ranging from $10^3$–$10^7$ molecules. The values of template molecules in the unknown samples are determined by interpolation of the $C_t$ (PCR cycles to reach an arbitrarily set threshold) of the unknown samples on the amplicon-specific standard curve. The data analysis protocols are embedded in the software included with the commercially available real-time PCR instrumentation.

6. We use robotics to assemble both the RT and PCR reactions, since robotics enhances both the throughput of the assays and the precision of the data that are generated.

VI. Conclusions

The era is dawning in which integration of large data sets (molecular, biochemical, and histological) create a multidisciplinary approach to accurately determine diagnosis, prognosis, and effective treatment regimens for each patient. Clearly, organizing the large quantity of data generated from gene array profiling of mRNA from biological samples linked to analyses tools will be an effective approach toward these goals. It is expected that an understanding of molecular signaling pathways will result in identifying each and every disease by its unique signature. It should also be expected that, in identifying key molecules
in each signaling pathway, drugs able to regulate biological activity will be identified. Thus, one could envision diagnosis by gene array and individualizing drug treatment for each patient (Figure 11).

The clinician has a critical role in identifying patient samples and creating the link to have these unique biological samples analyzed to create gene signatures. Currently, the application of array technology to the clinical investi-
ator is limited by the tissue sample, available arrays, and analysis of data generated from the array. A clinical investigator must have access to a collaborative group of scientists that assist in utilizing the gene array technology. This is a unique time for the clinician to help create diagnoses and prognoses tools for her/his patients and to be involved in bedside to bench to bedside experience. In the near future, advancing automation in the array field may reduce this dependence and broadly expand the use of array technology in clinical investigation. One should expect clinical diagnosis to follow a pattern described in Figure 11.

VI. Website Resources

Commercial Gene Arrays
Agilent Technologies: DNA_microarray@agilent.com
Affymetrix: www.affymetrix.com
Clontech: www.clontech.com
Incyte Pharmaceuticals: www.incyte.com/reagents/catalog/support/

Facilities Performing Gene Arrays/RT-PCR/Tissue Arrays/Databases
National Institutes of Health (NIH) genomics:
Stanford University: http://cmgm.Stanford.edu/pbrown/mguide
  http://genome-www4.stanford.edu/MicroArray/SMD/
Massachusetts Institute of Technology (MIT) cancer genomics website:
  www.genome.wi.mit.edu/MPR/
University of Texas Houston-Medical School RT-PCR/arrays:
  http://girch2.med.uth.tmc.edu/
University of Texas Medical Branch (UTMB):
  Affymetrix gene array facility: www.scms.utmb.edu/genomics
  Bioinformatics Group: www.bioinfo.utmb.edu/
Nature: http://genetics.nature.com/

Analysis Software
ArrayPro (Media Cybernetics): http://www.mediacy.com/arraypro.htm
ArrayStat (Imaging Research): http://www.imagingresearch.com/
Spotfire: www.spotfire.com

Public Sources of Software
University of Texas Houston-Medical School RT-PCR/arrays:
  http://girch2.med.uth.tmc.edu/
University of Texas Medical Branch (UTMB):
  Affymetrix gene array facility: http://www.scms.utmb.edu/genomics
  Bioinformatics Group: www.bioinfo.utmb.edu/
ACKNOWLEDGMENTS

A special thank you is extended to Shauna LeGrand for the design of Figure 9 and to Joanna Burris for editorial comments. This work was supported in part by grants from the National Institutes of Health AG/AR-11000 (RJU) and M01-RR-0073 (General Clinical Research Center, University of Texas Medical Branch) as well as John Sealy Memorial Endowment Fund for Biomedical Research at UTMB (JAC) and American Cancer Society Institutional Research Grant IRG-96–152-04 (UTMB).

REFERENCES


expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA 98:13790–13795
Eyster KM, Boles AL, Brannian JD, Hansen KA 2002 DNA microarray analysis of gene expression markers on endometriosis. Fertil Steril 77:42
Kimura SH, Ikawa M, Ito A, Okabe M, Nojima H 2001 Cyclin G1 is involved in G2/M arrest in response to DNA damage and in growth control after damage recovery. Oncogene 20:3290–3300


Korach KS 2002 Bringing genomics research to endocrinology. Endocrinology 143:1982


Minnmack ML, Ryan M, Baba H 2002 Gene expression analysis in schizophrenia: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. Proc Natl Acad Sci USA 99:4680–4685


Young RA 2000 Biomedical discovery with DNA arrays. Cell 102:9–15

Yun Z, Maeker HL, Johnson RS, Giaccia AJ 2002 Inhibition of PPAR gamma 2 gene expression by the HIF-1 regulated gene DECI/Stra13: a mechanism for regulation of adipogenesis by hypoxia. Dev Cell 2:331–341

Gene Expression Profiling for Prediction of Clinical Characteristics of Breast Cancer

ERICH HUANG,* MIKE WEST,† AND JOSEPH R. NEVINS*‡

*Department of Molecular Genetics and Microbiology; †Institute of Statistics and Decision Sciences; ‡Howard Hughes Medical Institute; ‡Duke University Medical Center; Duke University, Durham, North Carolina 27710

ABSTRACT

We have applied techniques of gene expression analysis to the analysis of human breast cancer by identifying metagene models with the capacity to discriminate breast tumors based on estrogen receptor (ER) status as well as the propensity for lymph node metastasis. We assess the utility and validity of these models in predicting status of tumors in cross-validation determinations. The practical value of such approaches relies on the ability not only to assess relative probabilities of clinical outcomes for future samples but also to provide an honest assessment of the uncertainties associated with such predictive classifications, based on the selection of gene subsets for each validation analysis. This latter point is of critical importance to the ability of applying these methodologies to clinical assessment of tumor phenotype. It is also clear from ER predictions that these analyses identify genes known to be involved in ER function but also identify new candidate genes involved in ER function. We believe these gene expression phenotypes have the potential to characterize the complex genetic alterations that typify the neoplastic state in a way that truly reflects the complexity of the regulatory pathways that are affected.

I. Introduction

Breast cancer is a prime example of a disease where further molecular characterization is needed to improve diagnostic and therapeutic strategies. Numerous studies have correlated genetic alterations with clinical outcome, including a strong correlation between the amplification of the erbB-2 receptor gene (Her-2) and poor clinical outcome (Tandon et al., 1989; Ciocca et al., 1992). In addition, overexpression of erbB-2 is a strong predictor of response to adriamycin-based therapy (Muss et al., 1994). Nevertheless, such correlations are few and often do not adequately define tumor subtypes from the viewpoint of substantially impacting therapeutic decisions. The inability to define a subclass of tumor type that may be refractory to standard therapies restricts the development of new, more- efficacious therapeutic strategies.

The analysis of gene expression represents an indirect measure of the genetic alterations in tumors since, in most instances, these alterations affect gene
regulatory pathways. Given the tremendous complexity that can be scored by measuring gene expression with DNA microarrays, together with the absence of bias in assumptions as to what type of pathway might be affected in a particular tumor, the analysis of gene expression profiles offers the potential to impact clinical decision making based on more-precise determinations of tumor cell phenotypes. It is critical that such analyses characterize the inherent variability and the resulting uncertainty about the predicted clinical status of tumors with out-of-sample predictions, in order to properly assess the potential utility of such information in therapeutic decision making. This has been the focus of much of our work (West et al., 2001; West, 2002).

II. Using Genome-scale Gene Expression Analysis to Study Cellular Phenotypes

Nucleic acid arrays, and the genome-scale expression data they capture, represent a major advance in the biological sciences, not only for the efficient, high-throughput data collection they afford but also for the fresh analytic methodologies and new avenues for interpreting biology demanded by data sets of such novelty, mass, and complexity. Whatever the specific platform, microarrays depend on the same principle: that complimentary nucleic acid sequences will hybridize preferentially to sequences mounted on a substrate. As such, microarrays are no more than refinements of a technology that has existed and been employed ubiquitously since its description by Southern (1975). The power of microarrays depends on two significant advances: that the genomes of several organisms have been characterized or substantially characterized and the ability to precisely fix many thousands of sequences reproducibly on a substrate. The synthesis of these genomic and technical revolutions brings about an unprecedented capability to assess and quantify a significant portion, if not virtually all, mRNA sequences present in a tissue at one time. However, the sheer quantity of such data forces the biologist to face the challenge of interpreting hundreds of thousands of data points in a statistically robust and responsible manner that can be conveyed readily to the scientific public.

In its current nascent state, analysis of gene expression data is hampered by the fact that the number of expressed genes being assayed is generally orders of magnitude greater than the number of experimental samples. A typical experiment will involve at most a couple of hundred samples, while microarrays typically assay tens of thousands of sequences. This results in algorithms or models that are highly susceptible to “overfitting,” where random noise in a data set is mistaken for substantive biological structure. Until the point is reached where sample numbers begin to match the number of genes represented on an array, it is exceedingly important that investigators first genuinely understand the data-analysis methodologies they employ and how effectively these methods
distinguish noise from valid structure. Out-of-sample predictive evaluation of models, using cross-validation or bootstrap techniques, is fundamental to validate analyses and their interpretations. Traditional analytical methodologies sufficient for small data sets involving a handful of replicates and a few thousand data points are insufficient to critically evaluate genome-scale data. A further danger is assuming that traditional modes of thinking about biology are sufficient for explaining gene expression data. Such data already are demonstrating that assumptions about the linearity and independence of biological pathways belie the complexity and richness seen in genome-scale data.

An important innovation in analyzing and interpreting gene expression data has been the use of unsupervised learning procedures to identify structure in microarray data sets. A variety of methodologies—hierarchical clustering, K-means clustering, and self-organizing maps—identify genes that share similarity in their patterns of expression and group them taxonomically according to that similarity. Again, patterns of expression are determined at a single-gene level but now similarity metrics allows one to group genes into coregulated clusters.

Alternatively, one can approach genome-scale expression data as a complex molecular phenotype and seek methods that treat such data as a composite of both up- and downregulated genes in toto rather than as ontologically isolated genes. This approach also seeks to define structure in expression data but differs from clustering methods by directly relating structure to phenotype, rather than simply describing that structure. We term these alternative integrated data structures of coordinately up- and downregulated genes “metagenes.” We apply them to models that test the association of genes with phenotypic states and provide a predictive capability that facilitates rigorous cross-validation of conclusions generated from microarray experiments.

A decisive point in developing any analytical method for genome-scale data is drawing the distinction between using methods to describe the data and developing predictive models from the data. Data-mining techniques (e.g., unsupervised or machine learning procedures) are exceedingly powerful in identifying and describing features in large, complex data sets. Their weakness is that they only describe the data in hand and do not formally address whether features of that data are generalizable and applicable to the real world in a predictive sense. Assuming a particular data set is a reasonable sample of reality, it is quite likely that features identified by a technique such as hierarchical clustering can be thought to be representative of genuine biological processes. Still, there is no proper methodology for directly establishing or testing this link. In contrast, by interpreting a data set through developing models at the outset, the model can be used to immediately predict new samples or, more importantly, the analysis can be stressed and tested to formally ascertain whether it properly predicts outcome. While this may be a subtle point, it is exceedingly important. A data set is a sample of a real-world distribution. There always is the possibility
that the sample is biased due to chance and the wisdom extracted from the data is relevant to itself and irrelevant to the real world.

In cases where the number of samples in a data set are far fewer than the predictor variables, the only feasible means to correct for overfitting is to test a model derived from the data as rigorously as possible. By performing such corroboration, one can test the possibility that features of the data thought to be important are generated by chance. Validation procedures (e.g., testing against an independent data set, out-of-sample cross-validation) are exceedingly important. Validation represents the thin line between mistakenly accepting adventitious or confounding structure in data as being biologically relevant and truly defining structure that corresponds to biological function. Most data-mining techniques are sensitive enough to find any and all structure present. It is the responsibility of the scientist to ascertain which structure is genuinely useful.

To demonstrate the importance of this issue, we generated two data sets for analysis. The first was a “real” data set comprising replicate samples from cells with ectopically expressed E2F genes (E2F1, E2F2, and E2F3) versus a Control set. The second was a “mock” data set in which a random number generator was used to create a synthetic gene expression data set with identical but arbitrary class assignments: A, B, C, and Control. The only stipulation for the mock set was that it possess the same mean and variance as the real set. Each data set was screened for the top 100 most-correlated genes and analyzed using metagene modeling techniques (West et al., 2001; Spang et al., 2002). Simply stated, metagenes are singular factors (or principal components) that are derived using singular-value decomposition methods, a standard method of data decomposition that isolates unrelated linear combinations of genes that each measure aspects of the patterns of variation and covariation in the full set of data. A three-dimensional metagene plot is created for the resulting metagenes, one for each experiment (Figure 1). It is evident that metagenes can “find” structure in the data that unambiguously separate the E2F1-, E2F2-, and E2F3-expressing cells from each other and from the Control. Strikingly, a similar robust separation was achieved with the mock data set of randomly generated values. Clearly, the massive complexity of this data set allows one to find structure simply by chance.

In order to generate a predictive model for E2F1 versus Control, we sought out structure specific to the E2F1 phenotype, as can be seen in the fitted classification based on combined metagene scores. The technique successfully identifies structure that can distinguish E2F1 samples from Control and does so with an estimate of probability of accurate classification. Once again, a similar analysis for the mock data set generates a classification that is as clean as for the real data. For all intents and purposes, the fitted model for the mock data appears to be as valid as the fitted model for the genuine data set.

Only under the stringent conditions of out-of-sample cross-validation does it become obvious that a model generated from the randomly generated data set
fails to deliver a genuine, generalizable predictive model for distinguishing experimental from controls. To assess this, analysis was repeated a number of times, each time removing one of the samples from the data set, then predicting its class based on the remaining data. This hold-one-out cross-validation approach is the most-searching assessment of predictive value of a model and has been stressed as a cornerstone of our work to date. As evident in Figure 1, the mock classification model is no more accurate than a coin flip — even after many repetitions, the classification of samples was merely stochastic, while the model generated from the real data set robustly predicts E2F1 status versus Control. This example illustrates how all techniques for exploring and modeling data can be sensitive to the fallacies of overfitting. Representations of data — such as multidimensional scaling, principal component plots, or clustering schemes — can demonstrate misleadingly unambiguous class separation. However, unless that structure is built into a model and tested, it cannot be assessed whether that structure is genuine or generated by chance. This issue is important because data sets of the complexity associated with gene expression data inevitably will possess random structure that can be taken erroneously as biologically relevant.

III. Application of Gene Expression Analysis to Breast Cancer

Expression data have opened a realm of possibilities for understanding the process of neoplastic disease at the molecular level. Relying on the assumption that the complement of transcripts in a cell represents fundamental characteristics of phenotype, many groups have sought to create essentially a “molecular pathology” of neoplastic disease using gene expression data. This in the hope that data sets comprising many thousands of molecular features, as opposed to the few available previously, will allow for more-precise diagnosis, prognosis, and prediction of treatment response. Highly promising preliminary studies of leukemias, lymphomas, and solid neoplasms demonstrate that gene expression data can highlight differences between otherwise histologically identical diseases, point toward new prognostic methodologies, and even emphasize patterns idiosyncratic of specific individuals.

We have investigated the potential of metagene analysis using DNA microarray data for a better understanding of breast cancer, with the goal of harnessing gene expression data to identify novel prognostic or predictive methodologies to enhance clinical decision making. Often, traditional methods of phenotypic characterization are limited and are not able to discern subtle differences that may be important for developing a better understanding of the tumor and advancing therapeutic strategies for disease treatment. We have taken a two-pronged approach in creating a novel statistical method that provides robust probabilistic prediction and classification of tumors based on gene expression data and also permits formal assessment of the uncertainties inherent in
FIG. 1. An example of overfitting of data in microarray analysis. The top set of panels (A-C) represents analysis of data from a “real” gene expression data set derived from the deregulated expression of E2F proteins. Quiescent mouse embryo fibroblasts were infected with recombinant adenoviruses expressing either the E2F1, E2F2, or E2F3 protein. RNA was prepared 18 hours after infection and used to generate hybridization probes. These probes then were applied to Affymetrix Mu11K GeneChips for analysis. Panel A depicts a plot of the analysis of the data based on three...
any predictive model. Such an approach is critical in an arena where clinicians must gauge their certainty of a tumor’s phenotypic properties against the potential morbidities of specific interventions.

IV. Modeling Breast Cancer Based on Metagene Analysis

Metagene-based methodology for interpreting breast cancer, or any type of expression data, can be summarized in the following manner: each gene in a microarray experiment may be thought of as representing one dimension. Thus, an array representing 10,000 individual sequences defines a gene space of 10,000 dimensions. Each individual tumor sample that is hybridized to an array represents a point plotted in that 10,000-dimensional space. Therefore, 200 individual points plotted in 10,000 gene dimensions would represent a group of experiments encompassing data from 200 tumors. The data “cloud” of 200 points plotted in high-dimensional space possesses a certain structure that can be related to biological features of the data. The essence of metagene definition is drawing multiple regression lines through this cloud of data that successively whittle down the cloud’s structure. Each regression line represents a metagene, a composite summarizing the impact of many genes that quantitatively weights the “pull” that each gene dimension exerts on that particular regression line. The operation of deriving metagenes is equivalent to a singular value decomposition (SVD), a standard matrix factorization in linear algebra. Many genomics visualization packages use SVD for principal component plots of experiments. The difference in the approach we describe here is that metagenes are utilized as composite-weighted proxies for multiple genes and fed into a binary regression model for classifying experiments by their gene expression patterns. Thus, each metagene is treated as if it were a single gene, although each actually summarizes the inputs of many genes, thereby simplifying the process of interpreting complex, high-dimensional data by aggregating many dimensions into fewer. In many cases when applying this approach, we can render a question of 100 dimensions into five to 10 without “losing” data. A further advantage is that aggregating genes into metagenes is an additional method of grouping genes that

(Figure 1 caption, continued)

metagenes that provide discrimination of the samples expressing the individual E2F proteins from control. Panel B depicts a fitted classification using the top 100 genes from a combined metagene to classify the E2F1 and control samples, with estimated probabilities for the classification. Panel C depicts a one-at-a-time, out-of-sample cross-validation for the E2F1 classification, demonstrating the ability of the E2F1 metagene score to predict the status of samples treated as unknown. The bottom set of panels (D-F) represent an identical analysis but using a “mock” data set generated with random numbers and with a dimension (number of data points, mean, and variance) the same as the real data set. It is evident from this analysis that patterns can be identified in this mock data that separate the samples equally well as in the real data set but then fail upon cross-validation predictions.
share some biological property. In our models of estrogen receptor (ER) status, many of the most-influential genes in metagene analysis are known ER targets.

In order to generate a classification model, we first define a biological or clinical question, then agnostically seek the metagenes that best answer the question, dispensing with noninformative or adventitious metagenes. With most questions, multiple metagenes are required. The logistic regression problem becomes a relatively straightforward binary regression involving multiple metagene predictors that, in turn, each represent multiple genes. As a proof-of-principle undertaking, we elected to model ER status because it signifies a characteristic that is both prognostic and predictive. It also is independently verifiable by both immunohistochemistry (IHC) and protein immunoblotting and many aspects of ER biology function are understood. The advantage of the considerable body of knowledge related to the estrogen hormonal axis is that the roles of genes identified as being important in the model are more likely interpretable in the context of previous knowledge. To construct this classification, we focused on developing a model that could distinguish between tumors evaluated by IHC as having no ER present (ER\(^-\)) versus those with any receptor present (ER\(^+\)).

V. Modeling ER Status in Breast Cancer

For this study, we used 49 tumors collected in the Breast Cancer Program at Duke University Medical Center. Tumors were classified as ER\(^+\) or ER\(^-\) at time of diagnosis by IHC. These findings later were confirmed by protein immunoblotting for ER. Conflicts between these assays were identified in five cases. Consequently, these five cases — and an additional four randomly selected tumors — were separated out and treated as validation samples for testing the predictive model developed from the remaining training cases. Of the latter, two samples were rejected due to failed hybridization, leaving a training set of 18 ER\(^+\) and 20 ER\(^-\) cases, as determined by both IHC and immunoblotting. The five tumors with contradictory tests for ER raise concerns about heterogeneity in the tumor sample; hence, these were treated as of equivocal status and scrutinized on the basis of expression-based predictions of status employing our metagene model.

For the 38 training arrays of unequivocal ER status (as confirmed by both IHC and immunoblot), we first sought to reduce noise contributed by genes irrelevant to an ER predictive model by computing sample correlation coefficients between individual genes and ER\(^+\)/ER\(^-\) binary outcomes. After repeated empirical experimentation, we found that selecting about 100 genes with the largest absolute correlation coefficient values minimized noise, while providing enough relevant information for metagene models. In the absence of screening, results were comparable to what we found with a filtered subset but included all variation in the entire data set, which subsequently was integrated into the
singular-factor metagenes. This contributed adventitious influences to the metagenes and clouded the discriminatory ability of metagene models. In the ER analysis using all genes, results are broadly similar to those reported here. However, due to the much higher level of noise influencing the analysis, all predictive probabilities have much higher associated uncertainties and one or two tumors are much less well classified. Screening to a smaller, relevant, discriminatory subset of genes is guaranteed to reduce such unwanted noise, with cleaner and more-accurate results.

With a gene expression space defined by 100 genes culled from genes possessing the highest absolute correlation coefficient for ER status, we derived metagenes, of which the first (Figure 2A) provides strong discrimination between ER+ and ER− cases. By inferences on the gene regression vector in the binary regression, we found that this particular metagene displayed many significant values (not illustrated here). Though it appears that a single metagene can serve in a discriminatory model, in higher-dimensional space, many metagenes make subtle contributions. Thus, our classification model integrates several metagenes that are coalesced into a metagene score and binary regression model developed.

**FIG. 2.** Factor analysis for ER+/ER− comparison. (A) Pairwise factor analysis for the discrimination of breast tumors based on estrogen receptor (ER) status. Individual tumor samples are depicted in a scatter plot on two dominant factors underlying 100 genes selected in pure discrimination of the training cases. Each tumor is indicated as either ER+ (P) or ER− (N). Only the tumors in the training set are plotted. Factor 1 is clearly discriminatory. (Factor 4 is chosen purely for display purposes.) (B) Fitted classification probabilities for training cases from the factor regression analysis. The values on the horizontal axis are estimates of the overall factor score in the regression. The corresponding values on the vertical axis are fitted/estimated classification probabilities, with corresponding 90% probability intervals marked as dashed lines to indicate uncertainty about these estimated values. Coding is as described in panel A. [Reprinted with permission from West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR, Nevins JR 2001 Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci USA 98:11462–11467. Copyright 2001 National Academy of Sciences, U.S.A.]
from these inputs. Figure 2B is a graphical depiction of estimated classification probabilities for each of the training cases, with vertical bars indicating 90% probability intervals reflecting degree of uncertainty. This illustration of our fitted model, looking at each of the tumors in the developmental subset, actually represents in-sample discrimination between ER+ and ER− rather than prediction. Only by performing out-of-sample procedures do we genuinely test the efficacy of the model in classifying tumors by ER status.

Genes can be ordered by the absolute values of the estimated regression vector to provide an assessment of their relevance in the discrimination. Figure 3 depicts expression levels of the genes, with each row representing an individual gene, ordered from top to bottom according to the absolute values of the estimated regression coefficients. The group of genes includes some that function in the ER pathway, including the ER gene itself as well as a number of known ER targets (Table I). Several others contribute to the discrimination inversely with ER+ status (negative coefficients); some of these encode proteins that are known to have inverse relationships with ER function (e.g., maspin, glutathione-S-transferase (GST)-Pi). Also included are genes that are not regulated by ER but are known to function in concert with ER, such as hepatocyte nuclear factor 3 alpha (HNF3α) and androgen receptor. Although the model is not designed to discover regulatory mechanisms, these factor models may generate clues about relationships between genes that do relate to underlying functional pathways.

VI. Cross-validation Analysis of ER Status and Honest Prediction

A major practical interest and potential clinical value of such statistical analyses lies in the ability to predict clinical conditions based on gene expression profiles of the primary tumor. In the pilot study of ER status, the prediction of status is based on gene expression patterns. The goal is to develop a rational, theoretically well-founded estimate of the probability of ER status for any new case, accompanied by a realistic assessment of uncertainty. Because such uncertainties may be high, due to limited information and population heterogeneity, it is critical that this uncertainty be reported and communicated to clinical researchers and clinicians along with point estimates of outcome probabilities.

Using the set of 100 genes selected from the full training sample study, the regression model was refitted repeatedly to the training data, each time removing the ER status of one of the tumors and then estimating the classification probability for that tumor. This is a standard, “one-at-a-time” cross-validation analysis; the status of each tumor in the training sample is predicted based on the remaining cases. For a true predictive assessment, gene screening and selection must be performed separately in each “hold-one-out” analysis, mirroring the real-life circumstances that will be faced in using such models and methods to predict future outcomes. In each of the 38 analyses, this leads to a different subset
of 100 screened genes. These subsets are highly overlapping but reveal additional genes on a case-by-case basis, reflecting sample variability and inherent heterogeneity in expression profiles. Figure 4 illustrates the results: uncertainty intervals tend to be fairly wide for tumors whose predicted probabilities are in the central region (i.e., nearer 0.5 than 0 or 1). This reflects the ambiguity discovered in the expression profiles of these cases relative to the 100 genes found to be most discriminatory among the other 37 cases. These “uncertain” cases are of obvious special interest for further study. Case 16 clearly has an expression profile more in accord with those of the ER+ cases than with those sharing its designated ER− status. This case has a low level of ER gene expression, consistent with its
<table>
<thead>
<tr>
<th>Rank</th>
<th>Weight</th>
<th>Unigene cluster</th>
<th>Estrogen relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.08</td>
<td>Trefoil factor 1 (pS2)</td>
<td>Estrogen induced</td>
</tr>
<tr>
<td>2</td>
<td>0.079</td>
<td>Estrogen receptor (ER) 1</td>
<td>ER</td>
</tr>
<tr>
<td>3</td>
<td>0.067</td>
<td>Cytochrome P450, subfamily IIB</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.064</td>
<td>Trefoil factor 3</td>
<td>Estrogen induced</td>
</tr>
<tr>
<td>5</td>
<td>0.061</td>
<td>(Insulin-like growth factor)</td>
<td>Estrogen induced</td>
</tr>
<tr>
<td>6</td>
<td>0.057</td>
<td>Human clone 23948 mRNA sequence</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.056</td>
<td>Microtubule-associated protein tau</td>
<td>Estrogen induced</td>
</tr>
<tr>
<td>8</td>
<td>0.055</td>
<td>Hepsin</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.048</td>
<td>GATA-binding protein 3</td>
<td>Coexpressed with ER</td>
</tr>
<tr>
<td>10</td>
<td>−0.043</td>
<td>Serine protease inhibitor, clade B, member 5 (Maspin)</td>
<td>Induced by tamoxifen; inverse with ER</td>
</tr>
<tr>
<td>11</td>
<td>0.041</td>
<td>N-acetyltransferase 1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>−0.041</td>
<td>S100 calcium binding protein A9</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>−0.041</td>
<td>Retinoic acid receptor responder 1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>−0.039</td>
<td>Small inducible cytokine subfamily D, member 1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.039</td>
<td>Hepatocyte nuclear factor 3 Alpha</td>
<td>Synergistic with ER</td>
</tr>
<tr>
<td>16</td>
<td>0.038</td>
<td>37 kDa Leucine-rich repeat protein</td>
<td>Physical interaction with ER</td>
</tr>
<tr>
<td>17</td>
<td>0.038</td>
<td>(Androgen receptor)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>−0.038</td>
<td>Cathepsin C</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.037</td>
<td>Inositol polyphosphate-4-phosphatase, type II, 105 kD</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.036</td>
<td>Purinergic receptor P2X, ligand-gated ion channel, 4</td>
<td>Estrogen biosynthesis</td>
</tr>
<tr>
<td>21</td>
<td>−0.036</td>
<td>KIAA0125 gene product</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.036</td>
<td>(Neuropeptide Y receptor Y1)</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.035</td>
<td>Meis (mouse) homolog 3</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.035</td>
<td>LIV-1 protein</td>
<td>Estrogen induced</td>
</tr>
</tbody>
</table>

[Genes are listed according to the discriminatory ranking, with gene 1 having the greatest weight in the discrimination. Negative values indicate an inverse correlation with ER+ status (and thus a positive correlation with ER− status). Reprinted with permission from West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR, Nevins JR 2001 Proc Natl Acad Sci 98:11462. Copyright 2001 National Academy of Sciences, U.S.A.]
ER- determination, but with relatively elevated levels of other genes in the top group, such as a marginally elevated level of ps2. Cases 40 and 43 share similar expression characteristics to tumor 16, exhibiting elevated levels of several known estrogen-regulated genes. In some cases, the discrepancy in clinical classification versus molecular classification is evident from the expression data. The ER- cases (16, 40, and 43) that are most borderline exhibit patterns that lie somewhere between the ER+ and ER-, as does the ER+ case of tumor 11. Tumor 31, whose laboratory ER status determinations were conflicting, strongly exhibits a pattern consistent with an ER+ state.
With these exceptions, the predictive accuracy of the analysis is very high. In particular, 34 of the 38 tumor samples are predicted accurately with a high degree of confidence. Thus, not only do these expression patterns derived from regression analysis have the capacity to classify on the basis of ER status, they also have an ability to honestly predict ER status of unknown samples, demonstrating the validity of the link between expression and clinical phenotype. Note, again, the clear differences between this display and that of Figure 2B and the extent to which the true prediction in Figure 4 highlights the increased uncertainties about cases 16, 40, and 43 in the middle ground.

The validation procedures we have performed are essential for testing whether the metagene structure genuinely reflects, in a predictive manner, biological characteristics related to ER status. In the absence of such testing, the identified distinctions between ER+ and ER− tumor samples are, at best, highly promising but, at worst, the product of chance.

VII. Models That Predict Lymph Node Metastasis in Breast Cancer

The analysis of ER status demonstrates the power to predict status of samples with associated assessments of predictive uncertainties. A second analysis concerns the clinically important issue of metastatic spread of the tumor. Determination of the extent of lymph node involvement in primary breast cancer is the single most important risk factor in disease outcome (Shek and Godolphin, 1988). Here, the analysis compares primary cancers that have not spread beyond the breast to ones that have metastasized to the axillary lymph nodes at the time of diagnosis. The potential power in making this determination from the primary cancer is significant in those instances where a positive lymph node might be missed or where a tumor is poised to metastasize to the lymph node but has not yet done so.

We identified tumors as “reported negative” when no positive lymph nodes were discovered and “reported positive” for tumors with at least three identifiably positive nodes, resulting in 12 reported positives [1] and 22 reported negatives [0]. Following screening to select the top 100 most-correlated genes, the metagene analysis was performed as described for ER discrimination. As in the ER study, this first analysis used an overall screened subset of 100 genes and again provided a good classification based on lymph node status, quite comparable to that for ER discrimination. Figure 5 illustrates the practically relevant cross-validation analysis that adopts a screen to select potentially different genes for each held-out case. The screened subsets of 100 most-discriminatory genes vary more widely than that seen in the ER analysis as we move across tumors, reflecting higher levels of natural variation in gene expression patterns with respect to nodal status. All of the reportedly positive cases appropriately have estimated probabilities above 0.5, though some are close to that boundary with
FIG. 5. Analysis for nodal comparisons. One-at-a-time cross-validation predictions in the analysis of lymph node metastasis. Each case is predicted based only on the nodal status of the remaining training tumors, with the subset of 100 genes reselected in each case. As such, the analysis exhibits the resulting uncertainties about the extent of true predictive accuracy in a practical setting, reflecting inherent variability due to heterogeneity of expression profiles. Node-positive tumors are indicated as P and node-negative tumors as N. [Reprinted with permission from West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zazan H, Olson JA Jr, Marks JR, Nevins JR 2001 Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci USA 98:11462–11467. Copyright 2001 National Academy of Sciences, U.S.A.]

moderate uncertainty. Perhaps most interesting are the few reportedly negative cases whose predicted probabilities slightly exceed 0.5. Cases like this are of paramount interest, since identifying genomic predictors of the progression from node negative to positive is a major goal from the viewpoint of potential therapeutic implications. These cases could, in principle, represent tumors that have metastasized but were missed in the nodal determination or they could be cases that have not yet metastasized but are poised to do so. Such analysis of nodal status clearly illustrates the importance of honest cross-validation studies of predictions in gauging the validity of the classification. The honest cross-validation predictions reveal realistic levels of uncertainty, likely due to heterogeneity in the profiles and the clinical phenotypes, and stress the importance of

GENE EXPRESSION PROFILING & BREAST CANCER
the validation studies to verify the significance of the classification. Nevertheless, the analysis does identify gene expression patterns that have predictive capability. Clearly, it is the analysis of those tumors in the uncertain region that must be the focus of further studies.

VIII. Future Considerations

Numerous studies of leukemia, lymphoma, and breast cancer indicate that analysis of gene expression reveals patterns that can serve to classify tumors and define tumor subtypes. The method we describe here first explicitly poses a clinical question — in this case, whether ER+ and ER− tumors can be distinguished by the mRNA of primary tumor samples and whether axillary lymph node status can be classified by similar methods. Methodologically, this differs significantly from the dominant analytical method for functional genomics, hierarchical clustering. The most important difference is that it is, at the outset, a model-building approach that seeks an outcome, finds the data that best fit the outcome, and critically evaluates whether the outcome and data are robustly linked. In contrast, no question is posed with clustering. Clustering organizes a data set without human intervention solely by a similarity criterion, thus generating taxonomic trees of genes and experiments that appear related. This technique belongs to a class of data-mining tools called unsupervised learning algorithms. Clustering has demonstrated remarkable power for delineating the superorganization of data inherent in gene expression data sets. From the point of view of delineating structure, clustering and metagene-based approaches are quite comparable. It is in actively evaluating the link between a clinical or phenotypic question that the metagene approach has the advantage of providing quantitative relations of genes to outcome rather than to themselves. In purely utilitarian terms, because the process of deriving metagenes directly rotates and recomputes the values in gene expression hyperspace for each gene in the context of overall structure in the data, the rotated expression value can be fed directly into the models linking that data to an outcome. With clustering, no rotation takes place: relations between genes are drawn by similarity metrics. Inevitably, because of the complexity and sheer quantity of gene expression data, much structure is due to chance. Testing whether the composition of a cluster or membership of a particular gene in a cluster are the consequences of random processes becomes very difficult to assess. Because metagenes are rigorously assessed by their ability to successfully answer a clinical question, the relation of structure and phenotype is more rigorous and transparent.

At the same time, a model-based methodology forces the investigator not only to pose a question but also to pose the right question. Seeking a classification that cannot be corroborated in gene expression will result in weak models. Even if qualitative visual assessments by two-dimensional metagene plots,
multidimensional scaling, or clustering show a dramatic separation between experiments and controls, this separation is only quantifiable and testable if one proceeds to the next step and builds a predictive model based on that ostensible separation. In our comparison of a genuine and a synthetic data set, data-mining techniques found unambiguous and seemingly irrefutable gene expression patterns that properly separated two different classes. Even fitted models appeared to probabilistically confirm the validity of structure found in the synthetic set that drew the distinction between classes. Out-of-sample cross-validation, the gold standard stress test for a predictive model, represents the last word on whether structure is real or merely stochastic. It was only through this method that real data could be distinguished from randomly generated data. Similarly, with any gene expression experiment, if a biological question cannot be answered by the data, the weakness of the predictive model built on that assumption will be exposed — and very likely only be exposed — by validation testing.

With these considerations in mind, the analyses presented here demonstrate that clinically relevant phenotypes can be determined for primary breast tumor samples through the analysis of gene expression. We also develop predictive analyses that bring gene expression analysis to real-world clinical applicability, facilitating the use of complex gene expression patterns as discrete prognostic or predictive factors. Similar studies have utilized gene expression profiles in out-of-sample cross-validation studies. Most approaches use some form of initial gene screening to select discriminatory subsets. We have stressed and illustrated the practical importance of repeating such gene-screening exercises within each cross-validation, in order to adequately assess realistic uncertainties about predictions and avoid misleading confidence in predictive accuracy and validity.

Classifications of leukemias and lymphomas that have been achieved in recent analyses of gene expression patterns represent a significant step in the development of methodologies to phenotype tumors (Golub et al., 1999; Alizadeh et al., 2000). The analysis of breast cancer phenotypes likely represents a context of considerably more biological heterogeneity, reflecting subtle aspects of tumor phenotype. As such, the fact that the cross-validation predictions reveal tumors with an uncertain classification, particularly for the lymph node analysis, is not unexpected. Indeed, it would be surprising to find that such an analysis would yield two cleanly separated groups. In this context, it is critical to develop methods that not only validate classifications with out-of-sample cross-validation methods but also provide appropriate and adequate assessments of the inherent uncertainties found with such predictions. The predictive or prognostic capacity demonstrated here is particularly relevant because clinical decision making depends on a rational, theoretically well-founded model for assessing data from new patients. Because such prognostic and predictive factors are couched in probabilistic language, clinicians can make judgments based on unbiased assessments of the uncertainties in a classification.
The assay of ER status by IHC is far from perfect and can produce erroneous results, as highlighted by our study. In addition, such assays would not score alterations that disable the ER pathway as opposed to the receptor itself. Thus, if the clinically significant determination is the functional status of the pathway, not the status of ER itself, then measurements of gene expression profiles that reflect activity of the pathway could provide an important advance in understanding the behavior of breast cancers. The finding that the group of genes that contribute most weight to the discrimination include not only ER and ER pathway genes but also genes that encode proteins that synergize with ER (e.g., HNF3α, androgen receptor) points to the potential power of the analysis in identifying functionally significant relationships.

An additional important benefit of these analyses is the potential for identifying gene pathways underlying an observed phenotype. A key point is the capacity to identify not just highly expressed genes but also those whose expression pattern highly correlates with the phenotype, regardless of level of expression. Perhaps most important is the fact that these analyses identify not only genes expected to be involved in the phenotype (i.e., ER-regulated genes), thus validating the process, but also genes for which a connection is not immediately clear. It is the identification of this latter group of genes that represents a major focus of these studies: the use of expression analysis to identify genes that highly correlate with the observed phenotype, thus providing additional insight into the underlying biological pathways.

Finally, we note that the presence of metastatic breast cancer in axillary lymph nodes is the most-significant factor in overall survival (Shek and Godolphin, 1988). Although the determination of lymph node status is relatively routine, the surgical procedure is highly invasive. Selectivity in the process of identifying nodes for examination induces biases that suggest some reported negatives may, indeed, be truly positive (Kjaergaard et al., 1985; Hill et al., 1999). Furthermore, the ability to accurately predict axillary lymph node status based on an expression profile of the primary tumor may obviate the need for axillary lymph node dissection and the significant morbidity associated with this procedure. Perhaps of more significance is the patient with truly negative lymph nodes but a primary tumor that is poised to metastasize. Many more data are needed to determine the precision of the predictive capability for lymph node status but it is clearly possible that a gene expression profile could predict metastatic potential, even in the absence of reportedly positive nodes.
REFERENCES


West M 2003 Bayesian factor regression models in the “large p, small n” paradigm. Bayesian Statistics, in press.
Statistical Approach to DNA Chip Analysis

N.M. Svrakic,* O. Nesic,† M.R.K. Dasu,‡ D. Herndon,‡ and J.R. Perez-Polo†

*Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri 63110; †Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, Texas 77550; ‡Shriners Hospital for Children, Department of Surgery, University of Texas Medical Branch, Galveston, Texas 77550

ABSTRACT

Statistical methods for analyzing data from DNA microarray experiments are reviewed. Specifically, we discuss common experimental setups, methods for data reduction and clustering, and analysis of time-course experiments. While early microarray studies focused mainly on the basic methodological and technical aspects of DNA arrays, emphasis has shifted to biological, medical, and clinical applications. We mention several of these and present results from our recent research as illustrative examples. New developments in this ever-growing field are outlined.

I. Introduction and Outline

The recent development of DNA array technology, or DNA arrays — also called microarrays, gene chips, DNA chips, and biochips — has enabled researchers to monitor simultaneously levels of thousands of genes as they are expressed in tissues, cell lines, patient samples, etc., at particular times and under a variety of different conditions. Developed in 1996 (Lockhart et al., 1996), this new technology has experienced a remarkable growth in popularity and utility, as reflected in the number of papers published on the subject. A casual search of the literature shows that the number of publications dealing with DNA array technology has increased nearly exponentially over the past several years. In 1997, about eight papers were published on the subject, followed by 23 in 1998, 94 in 1999, 296 in 2000, and 1156 in 2001. From incomplete data for the first half of 2002, one finds about 800 papers citing some aspect of microarray technology.

While early microarray studies mostly focused on the basic methodological and technical aspects of DNA arrays (e.g., data normalization, error correction, replication), emphasis has shifted to biological, medical, and clinical applications. DNA chips are being used in pharmacogenomics and pharmacogenetics, toxicogenomics, human disease studies, disease screening, profiling and classification, diagnosis and clinical applications, and basic biological science studies. In each case, the experimental design has to be planned to fit the questions being
addressed. Following is a brief list describing the most-frequently used experimental designs.

In a typical experimental situation with microarrays, one may want to:

1. Compare gene expressions obtained from two or more different tissues — for example, healthy versus diseased tissue — in order to compare or classify them. This type of experimental design has been used, for instance, to get clues regarding mechanisms and causes of disease processes or to classify specific clinical varieties of cancers/tumors according to their expression profiles, in order to better predict prognosis (Golub et al., 1999; Dirix and van Oosterom, 2002).

2. Compare gene expression data obtained from the same tissue or cell line at different time points, in order to follow the time course of expression. This design can be used to monitor temporal gene expression patterns during the cell cycle (Spellman et al., 1998; Tamayo et al., 1999) or during development (Wen et al., 1998), temporal progression of a disease (Agrawal et al., 2002; Pomeroy et al., 2002; Spies et al., 2002), or response to a treatment (Nesic et al., 2002; Sotiriou et al., 2002).

3. Compare gene expression data obtained from different parts of the same tissue, in order to reconstruct spatial distribution patterns of gene expressions. An expanded version of this design, called voxelation, was used recently by Brown et al. (2002), who correlated microarray data with the site of gene expression in tissues by creating signatures of expression patterns in coronal hemissections at the level of the hippocampus of the human brain. By combining the data for the entire surface of a volume of brain section, a three-dimensional spatial pattern of gene expression was generated. This important study (Peterson, 2002) combines DNA array technology and brain-imaging technique, like functional magnetic resonance imaging (fMRI), to represent the expression patterns of the whole organ.

Irrespective of the type of microarray employed (e.g., cDNA, oligonucleotide, spotted), such experiments generate tens of thousands of data points per each measurement. In addition, depending on the experimental design, or the number of samples, or the number of time points, the complete data set to be analyzed often contains hundreds of thousands of gene expression levels. These data are most commonly presented in tabular form (Quakenbush, 2001), called an expression matrix (see Table I).

In Table I, “Gene 1” is the name of the first gene, “Gene 2” the second gene, and so on. The column labeled “Experiment 1” lists data obtained from the first microarray (or under one condition, or measured at one time point), the column labeled “Experiment 2” lists the data obtained from the second microarray (or under a second condition, or measured at a second time point), etc. Finally, the entry (number) “E_{i,k}” is the measured expression level of gene $i$ in the experiment $k$, so that the entry “E_{1,1}” is the expression level of “Gene 1” in “Experiment 1,” the entry “E_{1,2}” is the expression level of “Gene 1” in “Experiment 2,” etc.
Typically, a single microarray contains many thousands or tens of thousands of different probes ("Genes") and the complete experimental design may require measurements from tens or even hundreds of such microarrays ("Experiments"). Complete data are collected in a matrix similar to Table I of a size that may be $10,000 \times 50$, which translates to half a million entries to be analyzed.

The main difficulty in statistical analysis of such data sets stems primarily from the fact that one must deal with a small number of samples or "Experiments" (i.e., cell lines, patients, time points), relative to the large number of probes ("Genes"). Moreover, the unnormalized, raw expression levels of different genes in the same experiment (or under one condition, or at one time) (i.e., the numbers $E_{1,1}, E_{2,1}, E_{3,1}, \ldots$) may have values that range over several orders of magnitude — from values close to unity to values on the order of $10^5$. The ultimate goals are to establish how the expression level of some gene changes from experiment to experiment and to identify groups of genes that exhibit similar coexpression patterns. Statistical methods designed to deal with these issues continue to be adapted and developed, since they are crucial for providing useful data and for extracting reliable biological information from DNA array experiments. This chapter reviews some of these methods, starting from the most basic and working towards more complex ones. Some of our results, obtained by using Affymetrix GeneChips, are described briefly in the form of illustrative examples.

### II. Fold Changes in Expression Levels

The critical issue is statistical handling of expression data, as one typically wants to identify genes of potential interest and search for those that are systematically up- or downregulated across experiments. For this limited purpose, it suffices to perform simple statistical analysis of gene expression levels. Early papers reported such analyses by presenting a list of genes that show $\approx 2$-fold change in expression level. But even with this simple analysis, care must be taken because of the aforementioned "large number of probes vs. small
number of experiments” problem. This is especially important if one wants to attach statistical significance to the observed changes. For instance, to determine the expression level of a single gene in one experiment (or under one condition), one needs to make several replicate measurements — the more the better. Performing many replicate experiments, however, often is not feasible, due to the high cost of DNA chips or the limited amount of RNA or DNA material available. Nevertheless, it has been our experience, in agreement with more-formal reliability studies (Lee et al., 2000), that at least three replicates per experiment must be made to have reasonable statistical confidence in the expression values obtained. Once the expression value of a gene has been established through replicate experiments under one condition, one wants to compare that with the expression value of that same gene under some other condition. A usual way to make the comparison is through a 2-sided t-test, assuming normal distribution of replicated expressions, or by some other non-parametric method. With three or so replicates per experiment, the statistical significance of the difference between the two experiments typically is not very impressive and only those genes that exhibit large up- or downregulation between the two experiments can be identified with some confidence. Thus, due to the various sources of errors or chance variations between two measurements, DNA arrays cannot be used with great confidence to detect small (i.e., less than 1.5-fold changes) in expression levels across experiments. Even with this constraint, one is left with, for example, 1000 genes that are identified as significantly changed. To assign some confidence level to this finding, one can perform t-tests, one for each gene, requiring 1000 total tests. Then, to correct for the repeated testing, one can impose the usual Bonferroni correction to the individual significance levels (i.e., require that the p-value for each gene be 1000 times smaller than, say, 0.05). Unfortunately, under these settings, the Bonferroni condition turns out to be extremely restrictive and almost no genes with significantly changed expression levels are detected with required statistical confidence.

One way out of this impasse was suggested by Tusher and coworkers (2001), who proposed a new method, significance analysis of microarrays (SAM). The SAM procedure assigns “observed score” to each gene, depending on that gene’s expression level scaled by the standard deviation of replicated measurements. Next, a number of “balanced” permutations of expression values are performed and a similar score in each case is assigned, which is then finally averaged over all permutations to compute the “expected score.” The scatter plot of observed vs. expected scores is then used to identify significant changes in gene expression. With the additional adjustable threshold parameter, parameter Δ in the original article (Tusher et al., 2001), one can control the overall false discovery rate (FDR), the percentage of genes discovered to be potentially significant by chance alone. With an FDR of \( \approx 5–10\% \), which is deemed acceptable, one is still
left with dozens of genes that show statistically significant changes in expression levels. SAM analysis has become a standard statistical technique for detecting groups of genes with potentially significant change in their expression levels. (SAM software is available at http://www-stat.stanford.edu/~tibs/SAM/index.html.) Following such analysis, one can compile a table of significantly over- or underexpressed genes, under different circumstances, with the expectation that these genes most actively participate in the phenomenon under study.

III. Data Classification and Clustering

To go a step beyond simple recording of fold changes of gene expression levels, various methods of data reduction and classification have been devised to identify groups of genes that show similar expression patterns. To present the results of such classification, it is useful to have an intuitive visual representation (Eisen et al., 1998). This often is achieved by drawing dendrograms and/or color-coded representations of similarly expressed genes. The most-common approach to perform classification or grouping of data is by one of the many clustering methods. Even though clustering methods are deterministic and reproducible, they still are subjective, since they may yield different results depending on the selected algorithm, normalization, distance metrics, etc. The challenge is to select the most-suitable one for the purpose of the experiment, so that the clustering produces results appropriate to the question being asked or the hypothesis being tested.

To illustrate the issues involved, consider Table I. With each “Gene” (e.g., “Gene 2”) from this table, one can associate an “expression vector” with entries “$E_{2,1}$, $E_{2,2}$, $E_{2,3}$...” that are simply read off from the row corresponding to that gene. In other words, the expression vector of a gene contains expression levels of that gene in different experiments. The number of components (dimension) of the expression vector equals the number of experiments ($N_E$) and the number of expression vectors equals the number of genes ($N_G$). Geometrically, one can think of the expression vector as a point (tip of the expression vector) in the $N_E$-dimensional “expression space,” so that each gene is uniquely assigned a single point. The dimensionality of the expression space is equal to the number of experiments (typically, between 10 and 100), while the number of points in this space is equal to the number of genes (typically, several thousand). In order to group the genes (or points in expression space) into clusters, one needs to define some measure of distance between them. The most-straightforward and most commonly used one is the geometric, Euclidean distance between the two points (expression vectors) $i$ and $j$, the square of which is defined as

$$D_{i,j}^2 = \sum_k (E_{i,k} - E_{j,k})^2$$
where the sum runs over all experiments $k$ and $Es$ are the appropriate expression levels from Table I. The most-similar points are the ones with the shortest Euclidean distance between them. Another possibility is to use some nongeometric measure of similarity, such as the Pearson correlation, which basically measures how similar are the directions in which the two expression vectors point. Thus, one attributes greatest similarity to the points with the highest correlation score. This method is widely used but has the drawback that it may sometimes falsely attribute high correlation score to expression vectors that are dissimilar. This may happen when there is an outlier in the data, such that overall expression levels of two genes are unrelated but for a single experiment, in which there is a common large peak, thus producing artificially high correlation. This can be remedied by employing a different correlation measure, a jackknife correlation, which is robust to single outliers, as proposed by Heyer et al. (1999). Many other distance measures can be used but discussing them is beyond the scope of this chapter.

As a somewhat “orthogonal” procedure to gene clustering, it often is useful to perform experiment clustering. To achieve this, one can represent each experiment by an “experiment vector,” with its entries read off from the corresponding column of the expression matrix (see Table I). Thus, “Experiment 2” would be represented by an experiment vector with entries “$E_{1,2}$, $E_{2,2}$, $E_{3,2}$, etc.” The number of experiment vectors equals the number of different experiments, while the length (dimension) of this vector equals the number of genes. Each point in this “experiment space” corresponds to one experiment. By introducing appropriate distance measure between two experiment vectors, one then can cluster experiments according to their similarity.

Clustering experiments is particularly useful as a preliminary step to discover, for instance, eventual gross discrepancies between microarrays that may occur with faulty arrays or because of other systematic errors. As an illustration, we recently reported on a microarray experiment involving burn injury in rats (Spies et al., 2002), where gene expressions in the skin tissues from burned rats and normal rats were compared at four time points (2 hours, 6 hours, 24 hours, and 240 hours after the injury). We used three replicate experiments for each group: thus, $3 \times 2 \times 4 = 24$ experiments (arrays). After clustering of experiments (arrays), it was discovered that one of the 24 arrays differed markedly from the rest. In this particular array, only about 800 genes were expressed, while in all others, the number of expressed genes averaged around 4000. This difference was immediately visible in the clustering of experiments. The faulty array was discarded from further analysis and a proper one was substituted.

Experiment clustering also can be used to determine the overall effect of treatment, or healing, on global expression profiles. For instance, in a recent study of spinal cord injury (SCI) in rats (Nesic et al., 2002), we compared
expression levels from spinal cord tissues of 1) rats with injured cord, 2) rats with injured cord that were treated with N-methyl-D-aspartate (NMDA) receptor antagonist MK-801, and 3) control (sham) animals. We used three replicates per group and performed hierarchical clustering of nine experiments. The resulting dendrogram, shown in Figure 1, correctly demonstrated that, overall, the injured and MK-801-treated groups are more similar to each other than to the sham group.

In a similar manner, one can use experiment clustering to follow the overall healing process. To again cite an example from our study of burn injury in rats (Spies et al., 2002), after all 24 experiments (arrays) were clustered, it was evident that samples from burned skin 10 days after the injury were more similar to control (unburned) samples than to samples from other burned groups. This gave us the molecular imprint of the onset of healing process that already had started 10 days post-injury.

With the defined distance measure, whether in the expression space or in the experiment space, the next step is to select the appropriate clustering algorithm.

A. HIERARCHICAL CLUSTERING ALGORITHMS

Most gene-clustering algorithms are hierarchical. These methods are derived (Eisen et al., 1998) from algorithms used to construct phylogenetic trees; the most-similar genes are clustered first, while those with more-diverse profiles are subsequently included in a stepwise hierarchy of increasing diversity. This means that, in the first clustering step, the single most-similar expression profiles are linked to form nodes, the most similar of which are linked further in the second clustering step, and so forth, until all nodes finally are linked and the complete hierarchical tree of proximities (dendrogram) is obtained. Starting from the second clustering step and higher, each node may consist of two or more objects. The distances between nodes must be recomputed at each step. This can be done,

![Dendrogram](image)

FIG. 1. Dendrogram obtained by hierarchical clustering using average linkage between normalized values of expression levels for 1322 probe pairs in nine DNA chips: three sham samples (1–3), three injured spinal cord samples (4–6), and three injured spinal cord samples treated with MK-801 (7–9).
for example, by computing the distance between the nodes as the average distance between its objects, as in the average linkage procedure, or as the distance between two of its closest objects, as in the nearest-neighbor linkage procedure. Other options include distances computed between the centers of mass of clusters or their modifications. In most cases, however, average linkage procedure is considered acceptable.

These different linkage choices are made to compensate for potential problems with hierarchical clustering. Namely, as clusters grow in size, at higher levels of hierarchy, the expression vector that represents the cluster may no longer be representative of any of the genes in the cluster. Thus, actual expression patterns of the genes themselves become less relevant on higher levels of hierarchy. If a gene is assigned to the “wrong” cluster, this error cannot be corrected later under hierarchical clustering.

B. NONHIERARCHICAL CLUSTERING ALGORITHMS

1. K-means Clustering

Sometimes, when a priori knowledge exists about the number of clusters that should be obtained, one can use nonhierarchical K-means clustering to partition the data. In this procedure, one first specifies the number of clusters (K), then randomly assigns expression vectors to them. Distances between clusters are recomputed, expression vectors are reassigned to the nearest cluster, and the procedure is iterated until the point is reached when no new assignments are made. The K-means clustering procedure simply partitions expression data into K groups and does not produce a dendrogram, although one can be constructed later by a hierarchical procedure.

2. Self-organizing Maps

Another frequently used nonhierarchical procedure is self-organizing maps (SOMs), a neural network-based procedure for clustering. In this algorithm, one also specifies in advance the number of clusters, chosen usually as the nodes of a grid. The nodes are mapped into K-dimensional space, initially at random, and then iteratively adjusted. During each iteration, a data point is randomly selected and the node is moved towards that point by the amount proportional to its proximity, so that more distant nodes are moved the least amount. In this way, neighboring points in the initial geometry are mapped to nearby points in the data space. This process usually is iterated tens of thousands of times. SOMs are particularly useful for exploratory data analysis, in order to expose the global patterns in the data.
C. PRINCIPAL COMPONENT ANALYSIS

A somewhat more-familiar method for data reduction is the singular value decomposition (SVD), or principal component analysis (PCA) as it is known in statistics (Alter et al., 2000; Yeung and Ruzzo, 2001). In this procedure, expression data from the "genes × experiments" expression space are transformed to diagonalized "eigengenes × eigenexperiments" space, where eigengenes (or eigenexperiments) are unique orthonormal superposition of genes (or experiments). PCA is essentially a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, with each succeeding component accounting for as much of the remaining variability as possible. This transformation represents the data in the new reduced coordinate space, in which individual genes or experiments appear to be classified into groups of similar functions or similar cellular state or phenotype. A simple illustration of PCA is given in Figure 2, in which the first principal component of a two-dimensional data set is shown by a straight line.

So far, the discussion has focused on the most common and most frequently used clustering algorithms. One should note, however, that other approaches exist, such as Bayesian and neural network algorithms, together with their numerous variants and modifications, that have been implemented in many DNA array studies.

FIG. 2. Principal component analysis (PCA) of a two-dimensional data cloud. The straight line shown is the direction of the first principal component, which gives an optimal (in the mean-square sense) linear reduction of dimension from two to one.
D. BIOMEDICAL APPLICATIONS OF CLUSTERING

Use of clustering/classification procedures in microarray experiments has been particularly fruitful in cancer research because cancers are complex, multigenic diseases with a natural control group for the analysis — noncancerous tissue (Alon et al., 1999; Lin et al., 2002). This was studied in prostate cancer behavior (Singh et al., 2002), where a set of gene expression differences between healthy and diseased tissues was detectable at the time of diagnosis. Alternatively, clustering procedure can be used to compare cancerous tissues of the same type and to distinguish between clinical subtypes, as was done in two types of breast cancer (Hedenfalk et al., 2001). The procedure also was very efficient in finding genes that distinguish between small blue cell tumors and leukemias (Tibshirani et al., 2002) as well as in the discovery of a new subset of melanomas (Bittner et al., 2000). The general conclusion drawn from these and other studies is that different cancers can be classified by the characteristic expression patterns of not more than dozens of genes. With more than 200 types of cancer, DNA microarray experiments are becoming an important tool to distinguish between their types and subtypes on the molecular level.

E. CLUSTERING OF TIME-COURSE EXPERIMENTS

An important class of DNA array experiments, in which data classification and clustering have been used successfully, are time-course experiments. In this setup, genome-wide expressions are measured at different time points in order to discover the temporal pattern in the course of development, or during a response to a treatment, or during a healing process. In this context, we mention the important pioneering work by Tamayo et al. (1999), where the temporal patterns of gene expression during the yeast cell cycle were classified by the SOM algorithm. The expression measurements were taken at 16 equally spaced, 10-minute intervals over two cell cycles (160 minutes), yielding a total of 30 different patterns. The classification was able to successfully extract yeast cell-cycle periodicity as the most prominent feature in the data and to select the appropriate group of genes that participate in the cycling process.

Following this work, numerous articles have reported results of temporal gene expression patterns under a variety of conditions. These include the temporal gene expression mapping of central nervous system development in rat’s cervical spinal cord (Wen et al., 1998); response of human bronchial cells to smoke and hydrogen peroxide (Yoneda et al., 2001); differentially expressed genes in human myometrium during pregnancy and labor (Aguan et al., 2000); and a range of experiments in toxicogenomics that measure response following exposure to toxicants, to identify drugs that provoke adverse reaction (Castle et al., 2002). A large-scale study of development and metabolic pathways in mice, with approximately 1.8 million measurements of gene expressions based on 294
microarray analyses of 49 adult and embryonic tissues (Miki et al., 2001), is perhaps the best illustration of the versatility of time-course DNA array experiments.

The time points where expression levels are measured in time-course experiments need not be equally spaced, since biologically important events often occur over different time scales. To give an example (Spies et al., 2002), we analyzed the time course of healing and recovery of burn wounds in rats, with measurements made at the following four time points: 2 hours, 6 hours, 24 hours, and 240 hours after the burn injury.

The goal of our study was to identify local responses and initial cellular responses to skin thermal injury by comparing expression profiles in burned and unburned rat skin tissue. The associated genomic events include differential expression of genes involved in cell survival and death, growth regulation, metabolism, inflammation, and immune response. The dynamics of these events is most clearly seen when genes with similar temporal expression patterns are clustered together.

With only four data collection time points, the temporal change of the ratio of burned vs. unburned expressions can be analyzed in considerable detail. Note that, in this case, there are 27 possible dynamical patterns of temporal development. At each time point, the value of the expression ratio (burned vs. unburned) can be 1) increased with respect to the previous time, 2) decreased, or 3) remain the same as the value at previous time. Thus, starting at some base value at time 1, the expression ratio can change/not change at later time points 2, 3, and 4, giving \(3^3 = 27\) possibilities of development (Figure 2).

More generally, with \(t\) time points, there are \(3^{(t-1)}\) dynamical patterns. This number can become quite large quickly with increasing numbers of time points. For instance, with 17 time points equally spread over a 160-minute interval during two yeast cell cycles, as used in the already-mentioned work of Tamayo et al. (1999), there are over 43 million \(3^{16}\) mathematically possible different patterns. Of course, in this and similar cases, it is quite unrealistic and, indeed, completely unnecessary to consider all possible patterns in detail. What one needs is to classify existing data into a small number of characteristic patterns (clusters) with some global features like “clusters with peak expressions at 25–45 minutes and 85–105 minutes” (Tamayo et al., 1999) that correspond to some meaningful biological events. This is precisely what a clustering algorithm such as SOM performs: it searches through the large “pattern space” for the small set of characteristic patterns that reflects global features of the entire set. Alternatively, in dynamical system parlance, one can think of the final characteristic clusters as attractors and of sets of patterns assigned to them as points that fall into their domains of attraction.
Returning to the example at hand, the complete set of 27 patterns that can be obtained with four time points is shown in Figure 3. The total number of genes in our data is 781.

Note that Figure 3 shows only the generic shapes of possible patterns, not the actual data, and that the four time points (1–4) are drawn as equidistant to simplify the graphics. By simple visual inspection, one can see that the “population” of patterns, specified by the number “N” in the figures, differs widely between the patterns, from N = 1 in pattern 4 to N = 133 in pattern 17. This indicates that some types of expression patterns are much more frequent than others. In particular, of 781 genes in this example, 88% (685) of them show significant over- or underexpression between time points 1 and 2 (i.e., in the period between 2 and 6 hours following the injury). The remaining 12% (96) do not change their expression levels appreciably during the same period. Moreover, just eight genes (patterns 2 and 3) exhibit increased (decreased) late-stage activity only, during the 24- to 240-hour period, without significant change in their activity prior to this time. The dynamics that emerges suggests that in the early phase (i.e., an hour or so after the injury), most genes (88%) involved in the entire 10-day process change their activity. This is consistent with the dynamics of the wound-healing process, which can be divided into an early phase of abrupt energy depletion and necrosis, followed by a two-stage inflammatory phase, delayed cell death, formation of granulation tissue, and matrix formation and remodeling (Spies et al., 2002).

The particular set of genes participating in these processes can be analyzed in detail by examining each cluster separately. Consider, for illustration, cluster 6 (i.e., pattern 6) that includes 24 genes, as shown in Figure 4. The numbers 1–24 in this figure label the particular genes that belong to the cluster (their names are listed in the separate table, not included). In this and other figures, the time points are drawn as equidistant to simplify the drawing.

With images like this, one can see that all genes in this cluster show a peak of activity at time 3 (24 hours post-injury), as specified by generic pattern 6 from Figure 3. A better view of the patterns of change is obtained in the three-dimensional rendering of this image, where the third dimension is the value of the expression ratio (see Figure 5).

Yet another view of cluster 6, this time from the direction of time axis, shows more precisely the amount of over- or underexpression of the genes involved (Figure 6). From this view, one can simply read off the amount by which the genes in this cluster from the burned tissue are over- or underexpressed with respect to the unburned one. With this information from all clusters, and knowledge of the particular genes involved (especially their metabolic functions, protein products, etc.), one can reconstruct patterns of biological activity during the wound-healing process. A detailed presentation of this and other analyses for all clusters will be published separately (Nesic et al., 2002).
FIG. 3. Generic shapes of 27 possible dynamical patterns of gene expressions from four time-point measurements. Number "N" inside each box denotes the number of genes in our data that exhibit that pattern. The labeling (1–27) of patterns is arbitrary.
Clustering of gene expression patterns can be improved by including additional, more-complex relationships beyond simple coexpression that are implicit in time-course patterns. For example, a gene may activate or control another gene downstream in the pathway, thus introducing a time-delayed response. Another possibility is that two genes have opposing influences on each other, so that when the activity of one increases, the activity of the other decreases, producing inverted correlation. A study in this direction has been reported by Qian et al. (2001) where, instead of simple direct correlation, four different correlation measures between gene expression patterns have been taken into account: 1) simultaneous correlation, 2) time-delayed correlation, 3) inverted correlation, and 4) inverted and time-delayed correlation. The method was applied to the yeast cell-cycle data set of Tamayo et al. (1999) and new interactions were identified, implying new biological relationships between genes. Still, with this and other improvements such as time warping (Aach and Church, 2001) and dynamical modeling (Holter et al., 2001; Ramoni et al., 2002), much research remains to be done on the systematic classification and clustering of gene expression patterns from time-course DNA array experiments.

IV. Beyond Simple Clustering: Genetic Regulatory Networks

Clustering gene expressions into similar patterns usually is performed with the expectation that “coexpressed genes are coregulated,” a plausible assumption.
(Spellman et al., 1998) that is, however, not universally true. Simultaneous detection of overexpression of two different genes does not necessarily imply that they are regulated by the same pathway, even if they appear together in the same cluster after a stimulus is applied to the cells. Many stimuli are known to initiate several different genomic-scale processes simultaneously, so that observed synchronicity in gene expression at certain times may be purely coincidental.

In order to move beyond simple coexpression, one has to establish which genes in some cluster also share common regulatory elements that control their expression levels (a group of genes regulated by a common element has been dubbed “regulons” in recent literature). The final goal of such analysis is to construct genetic regulatory networks and to identify the function of many thousands of novel genes (Tavazoie et al., 1999). This approach has been successful in yeast (Lyons et al., 2000), for which the complete sequence of promoter regions is known. Unfortunately, this is not the case with mammalian or other systems, where untranslated first exons, followed by introns greater than 10 kb in size, can make promoter identification...
extremely difficult. In many organisms, the promoter regions have not been fully sequenced. To construct the network for the phenomenon in question, one must use statistical algorithms for clustering and motif discovery in combination with genomic data, cis-regulatory analysis, and known molecular biology of the process studied. In spite of all the difficulties, several genetic regulatory networks, or parts of them, have been constructed. Davidson et al. (2002) have mapped a gene regulatory network in sea urchin embryo that controls the specification of endoderm and mesoderm. Such studies reveal that, in addition to comprehensive gene expression maps (Kim et al., 2001) obtained by DNA array measurements, one needs as much other genome-wide information as can be mustered to unravel the intricate patterns of genetic interactions in biological processes.

Simultaneously with this development, additional knowledge is accumulating regarding the statistical nature of naturally occurring networks. Many biological networks (e.g., genetic, metabolic) exhibit “small world”-scale free behavior (Watts and Strogatz, 1998). This means that although the network may possess thousands of nodes, the path leading from one node to another is remarkably short. Such architecture may serve to minimize transition times between metabolic states or provide robustness against mutations (Fell and Wagner, 2000; Jeong et al., 2000; Wagner, 2000). These new insights, combined with the knowledge of biological processes, may lead us for the first time towards understanding biology at the systems level (Kitano, 2002).

FIG. 6. Side view of the gene expression patterns shown in Figure 5, exhibiting the range of expression change over all time points.
REFERENCES


Identification of a Nuclear Factor Kappa B-dependent Gene Network

BING TIAN AND ALLAN R. BRASIER

Department of Medicine and the Sealy Center for Molecular Sciences, The University of Texas Medical Branch, Galveston, Texas 77555-1060

ABSTRACT

Nuclear factor-kappa B (NF-κB) is a highly inducible transcription factor that plays an important role in the hepatic acute-phase response, innate/adaptive immunity, and cellular survival through the induction of genetic networks. The major transcriptional-activating species Rel A-NF-κB is a cytoplasmic complex whose nuclear translocation is controlled by its association with a family of inhibitory proteins, termed IκBs. Activation of NF-κB results in the targeted proteolysis of IκB, releasing NF-κB to enter the nucleus and bind to specific sequences in target promoters. Because the genomic actions of NF-κB are influenced by the stimulus applied and the promoter context/chromatin structure in which it binds, the spectrum of NF-κB-regulated genes has not been elucidated. We have begun to address this question, exploiting a tightly regulated cellular system expressing a nondegradable IκBα mutant that completely inhibits NF-κB action. High-density oligonucleotide microarrays were used to identify genetic responses in response to complex biological stimuli (viral replication) in the presence and absence of NF-κB. Using statistical and informatics tools, we identified two groups of NF-κB-dependent genes with distinct expression profiles: 1) a group with high constitutive expression whose expression levels fall in response to viral exposure and constitutive mRNA expression increases from NF-κB blockade, and 2) a group where constitutive expression was very low (or undetectable) and, after stimulation, expression levels strongly increased. In this group, NF-κB blockade inhibited the viral induction of genes. This latter cluster includes chemokines, transcriptional regulators, intracellular proteins regulating translation and proteolysis, and secreted proteins (e.g., complement components, growth factor regulators). These data reveal complexity in the genetic response to NF-κB and serve as a foundation for further informatics analysis to identify genetic features common to up- and downregulated NF-κB-dependent promoters.

I. Introduction

The mechanisms controlling RNA polymerase II-dependent gene expression in eukaryotes have been investigated intensively over the past several decades. From this work, it is widely accepted that transcriptional initiation is the primary mechanism controlling gene expression in response to tissue specific-, developmental-, and hormonal-induced cellular signaling (McKnight and Yamamoto, 1992). Transcriptional activation is a multistep process initiated by sequence-specific proteins binding to upstream regulatory sites of DNA flanking
transcribed regions of a gene. In highly inducible gene promoters, enhanceosomes — nucleoprotein complexes assembled on inducible enhancer sequences in the regulatory regions — allow for multiple, intracellular signaling cascades to modulate target gene expression (Thanos and Maniatis, 1995). Formation of the enhanceosome is initiated by the binding of sequence-specific transcription factors and results in the cooperative binding of accessory chromatin-remodeling proteins, nuclear coactivators, kinases, and/or histone acetylases into the nucleoprotein complex. The presence of chromatin-remodeling factors induces formation of an activated preinitiation complex that, in turn, controls DNA-dependent RNA polymerase II activity and subsequent gene expression. Although the basic biochemistry of preinitiation complex formation and the identification of how signaling molecules control transcription factor have resulted from the study of promoters on an individual level, the spectrum of genes controlled by specific signaling pathways within the context of a complex biological stimulus has not been explored systematically.

The development and application of high-density microarrays is an enabling technology that allows global understanding of genome-wide changes in cellular mRNA abundance in response to specific stimuli (Tavazoie et al., 1999; Bussemaker et al., 2001; Berman et al., 2002). However, as it is commonly used, mRNA profiling is a correlational tool that records patterns of gene expression in response to specific cellular perturbations. Although these studies provide new insight into cellular responses, allowing the identification of genes controlled by a stimulus not previously considered, little mechanistic information arises on the interaction of downstream signaling pathways in controlling specific subsets of genes. Herein, we describe our approach to experimentally identify genetic networks under inducible transcription factor control using mRNA profiling. We have chosen the highly inducible transcription factor, NF-κB, and sought to identify its role in response to a complex biological stimulus (viral replication).

A. NF-κB AS AN INDUCIBLE TRANSCRIPTIONAL REGULATOR

NF-κB is a ubiquitously expressed, highly inducible transcription factor that plays an important role in the hepatic acute-phase response, innate and adaptive immunity, and cellular survival through the induction of genetic networks (Barnes and Karin, 1997; Karin, 1999). Composed of five distinct gene products related by a common NH₂-terminal Rel homology domain, the NF-κB family is divided into two functionally distinct groups that freely heterodimerize. The first group consists of members translated as mature proteins that bind DNA weakly and contain potent COOH-terminal transcriptional activation domains (Rel A and c-Rel). The second group has members encoded by large precursor proteins that are proteolytically processed into small (~50 kDa) DNA-binding subunits with strong DNA-binding activity and weak transcriptional activation potential (p105/
NF-κB1 and p110/NF-κB2) (reviewed in Siebenlist et al., 1994; Baldwin, 1996; Barnes and Karin, 1997). The subunit composition of the NF-κB complex influences its subcellular localization, transactivation potential, and mode of regulation. For example, homodimers of NF-κB1 are primarily nuclear proteins with weak transactivation potential. They are activated by phorbol myristate acid (PMA) and peptide hormones to weakly activate gene transcription (Jamaluddin et al., 2000). Conversely, the prototypical NF-κB complex, composed of 50 kDa NF-κB1-65 kDa Rel A heterodimers, is primarily cytoplasmic. The complex is regulated by its association with a family of inhibitors, IκBs, whose members bind and specifically inactivate Rel A by masking its nuclear-localization sequence and preventing nuclear entry (Henkel et al., 1993; reviewed in Beg and Baldwin, 1993; Han et al., 1997). In response to cellular stimulation by cytokines and pathogens, Rel A-NF-κB1 enters the nucleus, where it becomes a potent transactivator.

The mechanism controlling nuclear translocation of Rel A-NF-κB1 has been intensively investigated. Rel A-NF-κB1 activation requires inducible proteolysis of the IκB inhibitors, a process initiated by IκBα phosphorylation (Brown et al., 1995). Intracellular NF-κB-activating signals converge on the multiprotein cytoplasmic IκB kinase complex (IKK), a complex that phosphorylates IκB on two serine residues (Ser32 and Ser36) in its NH2-regulatory domain (reviewed in Karin, 1999). The IKK is a multiprotein, cytosolic kinase of ~700 kDa, composed of catalytic basic helix-loop-helix-containing kinases, IKKβ and IKKα, and a regulatory subunit, IKKγ, required for coupling IKK to upstream activating kinases (Karin, 1999; Poyet et al., 2000; Zhang et al., 2000). The pathways for how the cytokines, interleukin-1 (IL-1) and tumor necrosis factor-alpha (TNFα), activated NF-κB through the IKK have been extensively investigated. IL-1 and TNFα cytokine receptors are single-pass, transmembrane receptors lacking intrinsic kinase activity. Upon binding ligand, the receptors form trimeric structures, resulting in intracellular aggregation of the cytoplasmic signaling domains. This results in the recruitment of a submembranous signaling complex activating IKK; in the case of the IL-1 signaling pathway, the signal adapters include MyD88, IL-1 receptor-associate kinase (IRAK), and the TNF receptor-associated factor 6 (TRAF6) (Cao et al., 1996). TRAF6 is required for coupling the activated IL-1 receptor to the IKK; recently, it was shown that TRAF6 is ligated by unique lysine 63-linked polyubiquitin chains, a modification that is required to activate the downstream MAPKK kinase, TGFβ-associated kinase (TAK1) (Ninomiya-Tsushima et al., 1999; Wang et al., 2001). TAK1, in turn, phosphorylates the first committed step in NF-κB activation, the NF-κB-inducing kinase (NIK), resulting in IKKβ phosphorylation and IKK activation (Malanin et al., 1997; Wang et al., 2001). In TNF signaling, TNF receptor-associated death domain (TRADD), TRAF2, and receptor-interacting protein (RIP) constitute the submembranous signaling complex (Hsu et al., 1996; Zhang et al., 2000). Here,
the RIP kinase plays an indispensable role in IKK activation, as targeted disruption of RIP abolishes NF-κB activation (Kelliher et al., 1998). RIP recently has been shown to recruit the IKK complex to the TNF receptor I by binding IKKγ, directly recruiting the cytosolic IKK to it (Zhang et al., 2000). IKKα and IKKβ then are phosphorylated, releasing the activated IKK complex back into the cytosolic fraction (Poyet et al., 2000; Zhang et al., 2000).

In the IKK complex, the catalytic IKKβ subunit is largely thought to be responsible for the site-specific serine phosphorylation of IκBα in its NH₂-terminal regulatory domain, resulting in the first rate limiting in the process of NF-κB activation. PhosphoIκB is specifically bound by the Skp1-cullin-F-box-type E3 ubiquitin ligase, E3RS, initiating IκB ubiquitination and proteolysis through the proteasome (Brown et al., 1995; Karin, 1999; Karin and Ben Neriah, 2000). A parallel pathway important in viral infection that produces IκB degradation through cytoplasmic calpains also has been identified (Jamaluddin et al., 1998; Han et al., 1999b). Following IκB proteolysis, liberated NF-κB enters the nucleus to activate target gene transcription.

B. MECHANISMS FOR NF-κB-DEPENDENT TRANSACTIVATION

Liberated, cytoplasmic NF-κB rapidly enters the nucleus through specific nuclear-importing signals located in its NH₂-terminal Rel homology domain and bind to specific DNA sites in the regulatory regions of inducible promoters. A nonbiased polymerase chain reaction (PCR)-binding site-selection assay has elucidated the high-affinity binding sites for the homodimeric NF-κB1 (p50) and Rel A (p65) proteins as 5’-GGGGATYCC-3’ and 5’-GGGRNTTTCC-3’, respectively (Kunsch et al., 1992). These binding sites match the empirically defined consensus binding sites of acute-phase reactant (angiotensinogen) (Brasier et al., 1996), cytokine (IL-8) (Garofalo et al., 1996; Brasier et al., 1998), and chemokine (RANTES, an acronym for regulated upon activation, normally T-cell expressed and presumably secreted) (Casola et al., 2000b) promoters.

Promoter-bound NF-κB activates transcription from inducible regulatory elements through the assembly of larger nucleoprotein complexes termed enhancosomes. Pioneering studies on the interferon (IFN)-β promoters have revealed that inducible transcriptional regulation is a multistep process involving the cooperative assembly of architectural (DNA remodeling), sequence-specific transcription factors, and coactivator/bridging proteins on a target enhancer. The phenomenon of binding cooperativity allows highly inducible genes to be expressed in the setting of limiting concentrations of transcription factors (reviewed in Carey, 1998). The IFN-β promoter contains a virus-inducible enhancer with three essential domains, termed positive regulatory domains (PRDs). In vitro, PRDII binds NF-κB, PRDIV binds activating transcription factor (ATF)-2/c-Jun, and PRDIV-1 binds interferon regulatory factor-1 (IRF-1).
Gene-transfer studies have shown that stimulus-specific activation of IFNβ is accomplished through initial recruitment of the architectural protein high-mobility group (HMG) I(Y), a minor groove DNA-binding protein recognizing AT-rich elements contained in the PRDII and PRDIV domains (Yie et al., 1999b). HMG I(Y) binding bends the DNA into a conformation that allows cooperative binding of the transcription factors NF-κB and ATF-2/c-Jun to their respective motifs in PRDII and PRDIV domains. Additional protein-protein interactions recruit IRF, general transcription factors, and the coactivator/pol holoenzyme cyclic AMP response binding protein (CBP)/pol II into the complex (Munshi et al., 1998). Additional modifications to DNA and histones result in an increased rate of preinitiation complex formation, allowing multiple rounds of transcription and reinitiation of the IFNβ transcripts (Yie et al., 1999b). Once assembled, the enhanceosome is an extremely stable nucleoprotein structure through protein-protein and protein-DNA interactions, resistant to oligonucleotide competition and detergent treatment (Yie et al., 1999a).

We have made similar observations on the NF-κB-dependent, highly inducible IL-8 and RANTES promoters (Casola et al., 2000a,b). For example, in the case of IL-8, although NF-κB is absolutely required for inducible transcriptional activation (mutation of the NF-κB-binding site renders the promoter inert to any stimulation), an additional upstream site containing the IRF-1/7 complex is required for activation of the gene (Casola et al., 2000a). Interestingly, gene-transfer experiments also demonstrate that the requirement for “ancillary” cis elements (e.g., the IRF-binding site in viral stimulation) is distinct from those required for activation by TNF (Casola et al., 2000a). Together, these data indicate that the actions of NF-κB are determined by presence of ancillary binding sites in the target promoter and constrained by the stimulus applied.

Enhanceosome formation involves coactivator recruitment, proteins required for control of core promoter activity at a distance. Coactivators are non-DNA binding proteins that allow transcription factors to couple with the basal transcriptional apparatus and induce chromatin remodeling (Blobel, 2000). Chromatin remodeling is an active process occurring as a first step in promoter activation, known as “de-repression” (Jones and Kadonaga, 2000). This adenosine triphosphate (ATP)-dependent event is mediated through enzymatic (histone acetyltransferase, HAT) activity contained in the p300/CBP and p300/CBP-associated factor (p/CAF) coactivators. HATs acetylate basic residues on core histones, weakening their binding, thereby allowing other components of the transcriptional machinery to access the target promoter (Korzus et al., 1998; Blobel, 2000). Although there is redundancy in coactivator activity, certain types of coactivators are known to preferentially mediate genomic actions of transcription factor classes. Of relevance here, Rel A recruits the p300/CBP coactivator through interaction with its COOH transactivation domain (Perkins et al., 1997; Wadgaonkar et al., 1999). Antibody-injection experiments have shown that the
HAT activity of the related coactivator p/CAF is also important in NF-κB transactivation. Moreover, the mechanism for coactivator recruitment appears to be distinct between NF-κB1 and Rel A-complexed DNA-binding sites. The steroid receptor coactivator 1 (SRC-1), originally thought to be a specific nuclear receptor coactivator, transactivates NF-κB1 binding sites in synergy with p300/CBP (Na et al., 1998). SRC-1 specifically binds NF-κB1 but not Rel A (Na et al., 1998), strongly suggesting that alternative pathways for coactivator recruitment and promoter activation are utilized, depending on the composition of the NF-κB complexes assembled on a specific site. The individual roles of the various coactivators have yet to be determined.

Some recent intriguing work in macrophages has indicated that chromatin organization controls the temporal access of NF-κB to its target promoters (Saccani et al., 2001). Chromatin ImmunoPrecipitation (ChIP) assays involve the specific immunoprecipitation of genes reversibly cross-linked to NF-κB that are detected by PCR. Using ChIP assays on lipopolysaccharide-stimulated macrophages, two distinct waves of NF-κB recruitment to target genes were observed. NF-κB was recruited rapidly (i.e., within 20 minutes of stimulation) to the IkBα, MIP-2, and manganese superoxide dismutase gene promoters, constituting the “constitutively and immediately accessible” (CIA) genes. CIA genes are stably associated with acetylated histone H4. Conversely, a second wave of NF-κB recruitment was observed on the RANTES, MCP, and IL-6 promoter, constituting the “regulated late accessibility” (RLA) genes. The RLA genes were not constitutively bound by acetylated histone H4 but were induced to bind acetylated H4 prior to detectable NF-κB recruitment (Saccani et al., 2001). Genes in the RLA group, although activated by NF-κB, apparently had to undergo chromatin remodeling event prior to NF-κB binding. The mechanisms and pathways controlling chromatin remodeling are incompletely understood; however, these workers observed that recruitment of activator protein-1 (AP-1) complexes temporally preceded that of NF-κB, suggesting that AP-1 may be inducing chromatin remodeling in a subset of RLA promoters. These observations suggest unanticipated complexity in NF-κB action by regulating its accessibility to target promoters to certain chromatin domains.

II. Approach for Experimental Identification of Genetic Networks Downstream of NF-κB

Although NF-κB-binding sites now can be identified with some degree of reliability, it is clear that the actions of NF-κB are controlled by the promoter context in which it binds, the stimulus applied, and the chromatin structure of the target promoter. Our general strategy was to employ a nonbiased, high-throughput methodology to identify genes downstream of NF-κB in response to a complex biological stimulus, schematically diagrammed in Figure 1. We rea-
soned that comparing high-density oligonucleotide arrays on control or stimulated cells in the absence or presence of a specific inhibitor of NF-κB activation would identify those genes downstream of NF-κB.

For this approach, we required a cell that could be manipulated to express a dominant-negative NF-κB inhibitor. Because NF-κB is required for normal cell growth and survival (NF-κB Rel A deficiency in mice is embryonic lethal due to massive hepatic apoptosis) (Beg et al., 1995) and adenovirus-mediated transduction could activate signaling pathways in infected cells, we developed a tetracycline-regulated cell system (Tet-Off) (Gossen and Bujard, 1992). This cell
system expresses the nondegradable, epitope-tagged IκBα (IκBα Ser32Ala/Ser36Ala, termed FLAG-IκBα Mut) under control of the Tet Operator sequences (TetO). FLAG-IκBα Mut contains site mutations in the serine phosphoacceptor sites of IKKβ and functions as a potent dominant-negative inhibitor of NF-κB activation (Thomas et al., 1998b). In the Tet-Off stable cells, we were able to inhibit expression of FLAG-IκBα Mut during the selection process by including low concentrations of doxycycline (Dox) in the selection medium. Dox binds the tetracycline transactivator (tTA) and inhibits its binding. Upon Dox withdrawal, tTA binds DNA (TetO) and activates expression of the target gene (Figure 2). Using this strategy, we successfully isolated stable clonal cell lines where FLAG-IκBα Mut expression was induced strongly after Dox withdrawal from the culture medium. For example, in Figure 3, a western immunoblot of cytoplasmic extracts was performed using antibody recognizing the FLAG

**FIG. 2.** Schematic view for regulated inhibition of NF-κB action. Inhibition of NF-κB is accomplished by a dual-selection procedure where the tetracycline-regulated transactivator (Tet-TA) regulates expression of the nondegradable FLAG-IκBα Mut inhibitor under control of the tetracycline operator. Tet-TA cannot bind DNA in the absence of tetracycline (doxycycline, Dox) and FLAG-IκBα Mut is not expressed at high levels. When Dox is withdrawn from the culture medium, FLAG-IκBα Mut is expressed, inhibiting NF-κB nuclear translocation.
epitope tag of the FLAG-\(\text{I} \kappa \text{B} \alpha\) Mut fusion protein. Highly inducible and tightly regulated expression of FLAG-\(\text{I} \kappa \text{B} \alpha\) Mut was seen with early passages after stable clones were isolated. Importantly, in the short durations used for expression and stimulation, these cells did not have detectable differences in spontaneous or viral-induced apoptosis (Tian et al., 2002). FLAG-\(\text{I} \kappa \text{B} \alpha\) Mut expression potently inhibited NF-\(\kappa \text{B}\)-dependent transcription in response to the prototypical NF-\(\kappa \text{B}\)-activating cytokine, TNF\(\alpha\), and in response to respiratory syncytial virus (RSV) infection (Figure 4). Figure 4A shows an electrophoretic gel mobility-shift assay (EMSA) of nuclear extracts taken from cells stimulated with TNF in the absence of Dox. (Data adapted from Tian B, Zhang Y, Luxon BA, Garofalo RP, Casola A, Sinha M, Brasier AR 2002 Identification of NF-\(\kappa \text{B}\) dependent gene networks in respiratory syncytial virus-infected cells. J Virol 76:6800–6814.)

FIG. 3. Regulation of FLAG-I\(\kappa \text{B} \alpha\) Mut expression by Dox. I\(\kappa \text{B} \alpha\) Mut-expressing cells isolated in the presence of Dox (2 \(\mu \text{g/ml in growth medium}) were transferred to Dox-free medium and cultured for 7 days. Cytoplasmic extracts were prepared and 150 \(\mu \text{g fractionated on 10% SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) and subjected to western immunoblot using anti-FLAG M2 monoclonal antibody-peroxidase conjugate. After FLAG detection, the immunoblot was reprobed for \(\beta\)-actin as a loading control (bottom). FLAG-I\(\kappa \text{B} \alpha\) Mut is strongly expressed in the absence of Dox. [Data adapted from Tian B, Zhang Y, Luxon BA, Garofalo RP, Casola A, Sinha M, Brasier AR 2002 Identification of NF-\(\kappa \text{B}\) dependent gene networks in respiratory syncytial virus-infected cells. J Virol 76:6800–6814.]

Expression of the dominant-negative I\(\kappa \text{B} \alpha\) (the slower migrating band) is induced strongly in the absence of Dox at levels comparable to endogenous I\(\kappa \text{B} \alpha\). Similarly, although endogenous I\(\kappa \text{B} \alpha\) is degraded proteolytically by TNF,
the FLAG-IκBα Mut protein is not. Finally, RNase protection assays demonstrated that sufficient NF-κB inhibition was achieved to interfere with expression of known NF-κB-dependent target genes (Figure 5). In this experiment, cells cultured in the absence or presence of Dox were infected with RSV and steady-state levels of RANTES, IP-10, and IL-8 were measured simultaneously.
by ribonuclease protection assay (RPA). The strong viral-inducible and time-dependent induction of RANTES, IP-10, and IL-8 was inhibited significantly in cells cultured in the absence of Dox. Control experiments, published earlier, indicated that similar levels of viral transcription was occurring in both cells (Tian et al., 2002). Together, these data indicated that we could achieve a robust and highly reproducible inhibition of NF-κB-dependent translocation and nuclear actions in response to a variety of stimuli.
With this information, high-density oligonucleotide arrays were used to assay for profiles of gene expression in response to RSV infection in the presence of FLAG-\(\text{IkB}\alpha\) Mut vs. those induced in its absence.

**III. Identification of NF-\(\kappa\)B-dependent Gene Network in Viral Infection**

We were interested in identifying NF-\(\kappa\)B-dependent genes in response to a complex biological stimulus, mucosal viral infection with RSV. RSV, a negative-sense RNA virus of the paramyxoviridea family, is the leading cause of epidemic bronchiolitis and pneumonia in children (Ruuskanen and Ogra, 1993). Lacking an effective vaccine, infection with this ubiquitous virus causes 40–60% of the bronchiolitis and 15–25% of the pneumonia cases in hospitalized children (Shay et al., 1999), accounting for \(\approx 100,000\) hospitalizations and \(\approx 500\) deaths annually in the United States (Shay et al., 2001). RSV productively replicates at high levels in a variety of airway epithelial cells, where it potently alters host gene expression. Our work using high-density oligonucleotide arrays has shown that at least 16 different C, CC, CXC, and CX3C chemokines are expressed by RSV-infected respiratory epithelial cells in three general expression profiles (Garofalo et al., 1996; Casola et al., 2000b; Zhang et al., 2001). Of these, several are known to be induced through an NF-\(\kappa\)B-dependent mechanism involving IkB\(\alpha\) proteolysis (Garofalo et al., 1996; Jamaluddin et al., 1998). Importantly, inhibition of NF-\(\kappa\)B prevents RSV-induced airway inflammation in experimental models of infection (Haeberle et al., 2002). Together, we interpret these data to indicate that the actions of NF-\(\kappa\)B are to induce the expression of genes important in airway inflammation.

To identify these potential pathology-inducing genes, we subjected RNA for high-density oligonucleotide array analysis extracted from control or infected cells (\(\pm\) Dox treatment). For comparison of the fluorescent intensity (average difference) values among multiple experiments, we normalized the data relative to the 2% trimmed mean, a global hybridization metric to correct for changes in hybridization efficiency (Zhang et al., 2001; Tian et al., 2002). The normalized average difference values were then subjected to a two-way analysis of variance (ANOVA with replications) to determine which genes were significantly influenced by either the RSV or Dox treatment. We found that RSV infection significantly changed the abundance of 1359 mRNAs, while FLAG-\(\text{IkB}\alpha\) Mut expression influenced only 380 gene-expression profiles. Comparison of the two groups found only 144 genes common to the two treatments. The proteins encoded by the genetic group sensitive to RSV + FLAG-\(\text{IkB}\alpha\) Mut expression were classified by their putative biological pathways (Table I). Notably, members of a wide variety of biological pathways were identified. Numerically, the largest groups included membrane proteins, metabolic enzymes, signaling molecules.
<table>
<thead>
<tr>
<th>Classification of NF-κB-dependent Gene Network</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Common name</td>
<td>GenBankID</td>
</tr>
<tr>
<td>Apoptosis</td>
<td></td>
</tr>
<tr>
<td>IAP-1</td>
<td>U45878</td>
</tr>
<tr>
<td>THBS2</td>
<td>L12350</td>
</tr>
<tr>
<td>Cell cycle</td>
<td></td>
</tr>
<tr>
<td>Rb</td>
<td>HG4036-HT4306</td>
</tr>
<tr>
<td>TSC1</td>
<td>AF013168</td>
</tr>
<tr>
<td>CETN1</td>
<td>U03270</td>
</tr>
<tr>
<td>ZNF211</td>
<td>U38904</td>
</tr>
<tr>
<td>Chemokines/cytokines</td>
<td></td>
</tr>
<tr>
<td>IL-8</td>
<td>M28130</td>
</tr>
<tr>
<td>IFN-gamma</td>
<td>J00219</td>
</tr>
<tr>
<td>MGS/A/Gro-alpha</td>
<td>X54489</td>
</tr>
<tr>
<td>RANTES</td>
<td>M21121</td>
</tr>
<tr>
<td>SCYA22</td>
<td>AC004382</td>
</tr>
<tr>
<td>Cytoskeletal</td>
<td></td>
</tr>
<tr>
<td>TUBG1</td>
<td>M61764</td>
</tr>
<tr>
<td>COL16A1</td>
<td>M92642</td>
</tr>
<tr>
<td>COL11A1</td>
<td>J04177</td>
</tr>
<tr>
<td>Merlin</td>
<td>L11353</td>
</tr>
<tr>
<td>BENE</td>
<td>U17077</td>
</tr>
<tr>
<td>NuP214</td>
<td>X64228</td>
</tr>
<tr>
<td>STXBPI</td>
<td>AF004563</td>
</tr>
<tr>
<td>Membrane protein</td>
<td></td>
</tr>
<tr>
<td>PRGP1</td>
<td>AF009242</td>
</tr>
<tr>
<td>Type II mem protein</td>
<td>AB015633</td>
</tr>
<tr>
<td>QA79 membrane protein</td>
<td>AJ007395</td>
</tr>
<tr>
<td>EPB41L1</td>
<td>AB002336</td>
</tr>
<tr>
<td>Common name</td>
<td>GenBankID</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>N33</td>
<td>U42360</td>
</tr>
<tr>
<td>SNAP25</td>
<td>D21267</td>
</tr>
<tr>
<td>EBNA1BP2</td>
<td>U86602</td>
</tr>
<tr>
<td>DNA maintenance</td>
<td></td>
</tr>
<tr>
<td>DNase gamma</td>
<td>U75744</td>
</tr>
<tr>
<td>UBE2V1</td>
<td>U49278</td>
</tr>
<tr>
<td>Growth factors</td>
<td></td>
</tr>
<tr>
<td>IGF-1b</td>
<td>HG2309-HT2405</td>
</tr>
<tr>
<td>Histocompatability and cell surface</td>
<td></td>
</tr>
<tr>
<td>GAGE-2</td>
<td>U19143</td>
</tr>
<tr>
<td>HLA-G</td>
<td>AL022723</td>
</tr>
<tr>
<td>MIC2</td>
<td>M16279</td>
</tr>
<tr>
<td>CD59</td>
<td>M84349</td>
</tr>
<tr>
<td>Ion channels and transporters</td>
<td></td>
</tr>
<tr>
<td>SCNN1B</td>
<td>X87159</td>
</tr>
<tr>
<td>Metabolism</td>
<td></td>
</tr>
<tr>
<td>GST</td>
<td>M16594</td>
</tr>
<tr>
<td>PSMD13</td>
<td>AB009398</td>
</tr>
<tr>
<td>Chol 25 OH lase</td>
<td>AF059214</td>
</tr>
<tr>
<td>ATPase homolog</td>
<td>AF020736</td>
</tr>
<tr>
<td>TYMS</td>
<td>X67098</td>
</tr>
<tr>
<td>CYP11B1</td>
<td>X55764</td>
</tr>
<tr>
<td>TIMP-2</td>
<td>U44385</td>
</tr>
<tr>
<td>SL15</td>
<td>AF038961</td>
</tr>
<tr>
<td>GALT</td>
<td>M60091</td>
</tr>
<tr>
<td>PSMB5</td>
<td>D29011</td>
</tr>
<tr>
<td>5' ALA synthase</td>
<td>Y00451</td>
</tr>
<tr>
<td>Common name</td>
<td>GenBankID</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Enolase 3</td>
<td>X56832</td>
</tr>
<tr>
<td>E2Ub-conj</td>
<td>AA883502</td>
</tr>
<tr>
<td>Selenium donor</td>
<td>U34044</td>
</tr>
<tr>
<td>TCP20</td>
<td>D78333</td>
</tr>
<tr>
<td>ALDH3A2</td>
<td>U46689</td>
</tr>
<tr>
<td>ACAAl</td>
<td>X14813</td>
</tr>
<tr>
<td>CANPX</td>
<td>AJ000388</td>
</tr>
<tr>
<td>UBE4B</td>
<td>AF043117</td>
</tr>
<tr>
<td>Beta glucuronidase</td>
<td>X75940</td>
</tr>
<tr>
<td>GGT</td>
<td>J04131</td>
</tr>
<tr>
<td>PIGCP1</td>
<td>AB000359</td>
</tr>
<tr>
<td><strong>Oncogenes</strong></td>
<td></td>
</tr>
<tr>
<td>PEP1</td>
<td>L16464</td>
</tr>
<tr>
<td>PLCB4</td>
<td>L41349</td>
</tr>
<tr>
<td><strong>Receptors</strong></td>
<td></td>
</tr>
<tr>
<td>FGFR4</td>
<td>L03840</td>
</tr>
<tr>
<td>CD160</td>
<td>AF060981</td>
</tr>
<tr>
<td>TLR-3</td>
<td>U88879</td>
</tr>
<tr>
<td>TSHR</td>
<td>S49816</td>
</tr>
<tr>
<td>Frizzled-1</td>
<td>AB017363</td>
</tr>
<tr>
<td>Frizzled</td>
<td>L37882</td>
</tr>
<tr>
<td>IL15RA</td>
<td>AF035279</td>
</tr>
<tr>
<td>Cadherin</td>
<td>AL031588</td>
</tr>
<tr>
<td><strong>RNA metabolism</strong></td>
<td></td>
</tr>
<tr>
<td>RAE1</td>
<td>U84720</td>
</tr>
<tr>
<td>RNAP A0</td>
<td>U23803</td>
</tr>
<tr>
<td>RNMT</td>
<td>AF009426</td>
</tr>
<tr>
<td>Common name</td>
<td>GenBankID</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Secreted proteins and peptides</td>
<td></td>
</tr>
<tr>
<td>IGF-BP6</td>
<td>M62402</td>
</tr>
<tr>
<td>hGH</td>
<td>J03071</td>
</tr>
<tr>
<td>Comp. B</td>
<td>L15702</td>
</tr>
<tr>
<td>Factor VIII</td>
<td>M14113</td>
</tr>
<tr>
<td>Signaling molecules</td>
<td></td>
</tr>
<tr>
<td>EphB4</td>
<td>U07695</td>
</tr>
<tr>
<td>PTPRR</td>
<td>D64053</td>
</tr>
<tr>
<td>MAPKK6</td>
<td>U39064</td>
</tr>
<tr>
<td>AKAP95</td>
<td>Y11997</td>
</tr>
<tr>
<td>RAGA</td>
<td>U41654</td>
</tr>
<tr>
<td>Ndr protein kinase</td>
<td>Z35102</td>
</tr>
<tr>
<td>PTPRB</td>
<td>X54131</td>
</tr>
<tr>
<td>Rap2</td>
<td>X12534</td>
</tr>
<tr>
<td>PTPN1</td>
<td>M33684</td>
</tr>
<tr>
<td>RAB4</td>
<td>M28211</td>
</tr>
<tr>
<td>Rar</td>
<td>U05227</td>
</tr>
<tr>
<td>Dmz-like 1</td>
<td>AJ005821</td>
</tr>
<tr>
<td>PRKWNK1</td>
<td>AB002342</td>
</tr>
<tr>
<td>Synaptotanin 2B</td>
<td>AF039945</td>
</tr>
<tr>
<td>CDC2L2</td>
<td>AL031282</td>
</tr>
<tr>
<td>GCN1</td>
<td>D50919</td>
</tr>
<tr>
<td>Transcription factors</td>
<td></td>
</tr>
<tr>
<td>PLA-1</td>
<td>AJ012214</td>
</tr>
<tr>
<td>STAT-1</td>
<td>M97936</td>
</tr>
<tr>
<td>IRF-7B</td>
<td>U53831</td>
</tr>
<tr>
<td>MITF</td>
<td>AB006909</td>
</tr>
</tbody>
</table>
(i.e., kinases, phosphatases), transcription factors, and those of unknown function.

To better visualize the gene-expression changes in the 144-member group regulated by both RSV and FLAG-IκBα Mut expression, hierarchical clustering was performed. In this technique, each gene-expression profile is grouped with its nearest neighbor and the mathematical proximity of this gene-expression profile is indicated by the height of a common line that connects the two nodes. A representative analysis is shown in Figure 6, where each treatment condition represents the average of the normalized average difference values from three independent experiments. From visual inspection of the dendogram, the gene-expression patterns are divided into two large clusters. The top cluster contains a group of genes whose constitutive (unstimulated) expression is high and falls with RSV infection. In this group, inhibition of NF-κB reduces the viral-induced inhibition of their expression. The bottom cluster contains a group of genes whose constitutive expression is low and is strongly increased by RSV infection. In this group, inhibition of NF-κB blocks the viral-induced activation of their expression.

### TABLE I

(continued)

<table>
<thead>
<tr>
<th>Common name</th>
<th>GenBankID</th>
</tr>
</thead>
<tbody>
<tr>
<td>IkB epsilon</td>
<td>U91616</td>
</tr>
<tr>
<td>IRF-9</td>
<td>M87503</td>
</tr>
<tr>
<td>NF-κB</td>
<td>X61498</td>
</tr>
<tr>
<td>NF-κB2</td>
<td>S76638</td>
</tr>
<tr>
<td>NR4A2</td>
<td>S77154</td>
</tr>
<tr>
<td>Nuclear Factor 1</td>
<td>HG4167-HT4437</td>
</tr>
<tr>
<td>Cdx2</td>
<td>U51096</td>
</tr>
<tr>
<td>TCEAL1</td>
<td>M99701</td>
</tr>
</tbody>
</table>

(The proteins encoded by the 144-member genetic network under control by respiratory syncytial virus (RSV) and FLAG-IκBα Mut expression were classified by putative biochemical activity. For each group, the common name is listed with the GenBank accession number. GenBank ID numbers of probe sets that could not be classified are not included. They are U51698, AF070549, AI538199, AL041663, AI687419, AL121073, W27666, AB029012, Y18504, AL096750, AI925946, AF035300, AB002308, AB014559, AB007975, AF035292, AB018324, AB023201, D31764, M92357, AB023215, AA846749, AB002375, AL050002, AB002304, AB002350, AI936826, AI741833, W27517, AB014528, D50911, AC005053, AI793265, AB01843, AA477898, AB020628, AB014608, AB018259, and L22342.)
FIG. 6. NF-κB-dependent gene networks identified by high-density microarrays. Clustering and heat map analysis of the RSV and FLAG-IκBα Mut-regulated dataset. Cells cultured in the presence or absence of Dox were exposed to RSV for 12 hours prior to RNA extraction and analysis by high-density microarrays. Agglomerative hierarchical clustering was performed on average fluorescence intensity values from three independent experiments using the unweighted pair group method with arithmetic mean (UPGMA) technique. A heat map for each gene for the three independent experimental datapoints is shown at right. The gradient used is dark gray (minimum value of 5 scaled fluorescence intensity units), light grey (middle value of 5000 scaled units), and white (maximum value of 10,000 scaled units). The bracket at right indicates the subset of highly inducible NF-κB genes shown in Table I and analyzed in Figure 8. [Data adapted from Tian B, Zhang Y, Luxon BA, Garofalo RP, Carola A, Sinha M, Brasier AR 2002 Identification of NF-κB dependent gene networks in respiratory syncytial virus-infected cells. J Virol 76:6800–6814.]
The genes with high constitutive activity whose expression fell with RSV treatment were classified into biochemical pathways to determine whether a single process was being influenced by NF-κB action. Disappointingly, approximately half of the genes were unknown and, of those that could be identified, no signal pathway emerged. For example, we found secreted peptides (growth hormone, Genbank Accession No. J03071), intracellular signaling proteins (tyrosine phosphatase, X54131; serine/threonine kinase, AB018324; RAB4, M28211), nucleic acid binding proteins (RNPA0, U23803), metabolic enzymes (enolase, X56832; selenium donor, U34044), transcription factors (micro-opthalmia-associated transcription factor, AB006909), extracellular structural proteins (type XI collagen, J04177; α1-type XVI collagen, M92642), and others (Tian et al., 2002). Control experiments also indicated that expression of other genes in this list was influenced by Dox treatment. The mechanisms for NF-κB-dependent gene downregulation will require independent analysis and further study.

A similar analysis was conducted for the genes whose constitutive activity was low and whose induction by RSV was blocked by FLAG-IκBα Mut expression. We identified secreted complement factors (complement B), transcription factors (the proteolytically processed NF-κB member, NF-κB2; the signal transducer and activator of transcription, STAT-1; the interferon regulatory factors, IRF-9 and IRF-7), metabolic enzymes (5'-aminolevulinate synthase); growth factor binding proteins (insulin-like growth factor binding protein 6), kinases in translational control (human GCN1), cytokines (RANTES, IL-8), and others (B94). The group indicated by the parentheses in Figure 6 is also tabulated (Table II). Again, no single biological pathway is easily seen, although identifying some of the previously known NF-κB-dependent genes (e.g., IL-8, RANTES) was an important validation of our system and analysis (Figure 5). Together, these data suggest that NF-κB plays a multifaceted role in control of constitutive gene expression, involvement in RSV-mediated downregulation of genes, as well as in mediating upregulation of a network of RSV-inducible genes.

IV. Insights into Biological Roles of NF-κB Through the Functions of Its Downstream Genes

RSV is a negative-sense RNA virus that is replicated in the cytoplasm through its own RNA-dependent RNA polymerase. In RSV infection, therefore, NF-κB has no role in viral transcription and its purpose is to activate expression of the host’s innate immune and inflammatory responses. A number of highly inducible genes encoding cytokine, chemokine, acute-phase reactant, and adhesion molecules contain NF-κB-binding sites in their proximal promoters and, in a certain number of cases, are induced by RSV (Li and Brasier, 1996; Brasier et al., 1998; Thomas et al., 1998a; Casola et al., 2001; reviewed in Zhang and Ghosh, 2001). However, whether NF-κB is truly a master regulator of gene
<table>
<thead>
<tr>
<th>Name</th>
<th>Similarity scores</th>
<th>Sequence</th>
<th>Experimental verification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Core</td>
<td>Matrix</td>
<td></td>
</tr>
<tr>
<td>MGSA/Gro-alpha</td>
<td>1</td>
<td>0.964</td>
<td>ccGGGAAttccccct</td>
</tr>
<tr>
<td>IL-8</td>
<td>0.863</td>
<td>0.92</td>
<td>cgTGGGAAttcccte</td>
</tr>
<tr>
<td>RANTES</td>
<td>0.955</td>
<td>0.974</td>
<td>tggaaaCTCCcct</td>
</tr>
<tr>
<td>IRF-9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRF-7B</td>
<td>0.955</td>
<td>0.941</td>
<td>ceggaaaCTCCCCgc</td>
</tr>
<tr>
<td>STAT1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chol 25 OH lase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYB11B1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAP-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF-κB2 (promoter 1)</td>
<td>1</td>
<td>0.991</td>
<td>GGGGAccccecc</td>
</tr>
<tr>
<td></td>
<td>0.955</td>
<td>0.906</td>
<td>ggggaacCTCCcgg</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.967</td>
<td>atgggaattTTCCcag</td>
</tr>
<tr>
<td>IkB epsilon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2Ub-conj</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCN1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp B</td>
<td>1</td>
<td>0.943</td>
<td>atgggaattTTCCcag</td>
</tr>
<tr>
<td>IGFBP6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II Mem protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear phosphoprotein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTPN1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’ALA synthase</td>
<td>1</td>
<td>0.911</td>
<td>cgGGGAAtccccgg</td>
</tr>
<tr>
<td>IL-15R alpha</td>
<td>1</td>
<td>0.935</td>
<td>gttgggaattTTCCCca</td>
</tr>
<tr>
<td>GAGE-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cadherin-like 22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Subgroup of highly inducible genes taken from hierarchical clustering (marked by parentheses, see Figure 6) were classified by primary biological function. Shown are the core and matrix similarities from the TRANSFAC analysis for NF-κB binding sites. Y = experimental verification of the NF-κB binding site was found; NF = not found.)
expression in mucosal viral infection and which genes it controls have not been subjected to direct experimental investigation. A surprising finding from our study is the relatively small number of genes that were sensitive to FLAG-\(\kappa B\) Mut expression. Of the 1359 RSV-inducible genes, only about 10\% were shown to be NF-\(\kappa B\) dependent. It is important to emphasize that our analysis is based on whether a given gene is induced by virus for us to identify it as an NF-\(\kappa B\) target. In this regard, a study using a similar approach to identify NF-\(\kappa B\)-dependent genes in rat pancreatic \(\beta\) cells identified 66 genes under NF-\(\kappa B\) control (Cardozo et al., 2001). The NF-\(\kappa B\)-dependent targets that this study identified in pancreatic \(\beta\) cells were markedly different from those found in our study. In \(\beta\) cells, the most highly upregulated genes are inducible peptide 10 kDa (IP-10), IL-15, cluster of differentiation-40 (CD-40), and major histocompatibility complex (MHC)-II-associated invariant chain. Like our findings, a number of genes were downregulated in an NF-\(\kappa B\)-dependent manner, including the Pdx-1 transcription factor, a \(\beta\)-cell differentiation factor that controls the expression of glucose transporters (Cardozo et al., 2001). These surprising differences between our studies indicate that the “NF-\(\kappa B\) network” is under cell type-specific control. Also, species-specific effects come into play as, for example, rats do not have an IL-8 gene.

In human epithelial-derived cells, our studies show that NF-\(\kappa B\) is important in RSV-inducible gene expression for a number of genes with apparently diverse functions, including chemokines, transcription factors (interferon regulatory factor (IRF)/signal transducers and activators of transcription (STATs) and NF-\(\kappa B\)/\(\kappa B\) members), proteins controlling translation/proteolysis, secreted proteins, cytoskeletal elements, and signaling proteins in those whose function is known (Table I). Clearly, these observations will have to be extended for other NF-\(\kappa B\)-activating stimuli and into other cell types. Our observations further suggest that not only does NF-\(\kappa B\) play a role in viral-dependent gene activation, it also plays a role in constitutive and viral-dependent inhibition of distinct genetic elements. In the following paragraphs, we comment on some of the biological insights that this study suggested (see Figure 7 for a schematic overview).

A. CONTROL OF NEUTROPHILIC AND EOSINOPHILIC CHEMOKINE EXPRESSION

Cellular recruitment into the virally infected lung is a multistep process involving adherence of circulating leukocytes to an activated endothelial surface, followed by their diapedesis into the inflamed tissue and migration toward chemical gradients of chemotactant peptides or antigens (reviewed in Springer, 1994). Recent attention has focused on the important role of chemokines in mediating leukocyte chemotaxis into the airways. Chemokines are a
superfamily of proteins divided into four distinct groups — C, CC, CXC, and CX₃C — based on the number and spacing of highly conserved NH₂-terminal cystine residues (reviewed in Oppenheim et al., 1991; Baggioni et al., 1997). This grouping has functional significance because the spacing of the NH₂-terminal cystine residues influences the type of cell-surface leukocyte receptors that they bind to produce leukocyte activation and chemotaxis. Although a number of cell types can inducibly secrete chemokines, the airway epithelium is thought to play a central role in initiating pulmonary inflammation by RSV because the epithelium is the first target for RSV replication and the mucosa is the only tissue that allows productive viral replication. Our recent studies using high-density oligonucleotide arrays have shown that at least 16 different C, CC, CXC, and CX₃C chemokines are expressed by RSV-infected respiratory epithelial cells in three general expression profiles (Zhang et al., 2001). The molecular mechanisms controlling expression of CXC- and CC-type chemokines in RSV-infected airway epithelial cells have been investigated extensively (Garofalo et al., 1996; Brasier et al., 1998; Thomas et al., 1998a; Casola et al., 2000b).
studies here show that the NF-κB-dependent chemokine group contains the CXC chemokines GRO-α and IL-8 and the CC chemokine RANTES. The CXC chemokine class is potent chemotactic cytokines for activated neutrophils and is functionally subdivided into two groups, based on the presence or absence of a signature ELR (Glu-Leu-Arg) motif upstream of the canonical CXC motif. The ELR-containing group of CXC chemokines includes IL-8, GRO-α, and ENA78. These cytokines primarily activate the bacteriocidal activity and chemotaxis of neutrophils, an abundant cell type found in the bronchoalveolar lavage of intubated, RSV-infected children (Everard et al., 1994). The ELR CXC chemokines can activate other target cells. For example, IL-8 activates T cells and eosinophils (Matsushima and Oppenheim, 1989; Murphy, 1994), while GRO-α/β/γ activate basophils (Geiser et al., 1993). This may account for some of the spectrum of cellular infiltration in RSV-infected lungs and presence of cell-specific degranulation products in nasopharyngeal secretions in patients with naturally acquired RSV infection (Garofalo et al., 1992).

The CC chemokine RANTES, originally described as a T cell-derived chemokine, is now known to be highly expressed in stimulated airway epithelial cells, where it may serve to recruit T lymphocytes, monocytes, basophils, and eosinophils into the virally infected lung (Alam et al., 1994). RANTES has been shown to be highly NF-κB dependent by our studies and those of others (Thomas et al., 1998a; Casola et al., 2000b). Of relevance, RANTES has been shown to be present in nasal washes of children infected with RSV, where its levels and those of its cellular target, eosinophilic degranulation products, correlate with disease severity (Garofalo et al., 1994; Teran et al., 1996). Our studies suggest that NF-κB controls the expression of chemokines important in neutrophilic and eosinophilic recruitment into the airways.

B. CONTROL OF IFN SIGNALING PATHWAY

We were surprised that the NF-κB-dependent gene list contained important members of two transcription factor families, STAT-1 and IRF-9 and IRF-7B. STATs are cytosolic proteins contained in high molecular mass complexes (“statosomes”) (Ndubuisi et al., 1999) activated by tyrosine phosphorylation mediated by the interferon and growth factor receptor-associated kinases, Jak and Tyk (reviewed in Schindler and Darnell, 1995). Activated STATs form homo- or heterodimers through intermolecular SH₂ phosphotyrosine interactions and subsequently are translocated into the nucleus in distinct binding complexes dictated by the nature of the stimulus and target promoters (Taniguchi et al., 2001). The IRFs are a family of nine constitutive and inducible helix-turn-helix transcription factors important in innate defense through their ability to mediate viral-inducible transcription of cytokine and chemokine genes (reviewed in Taniguchi et al., 2001). IRF-9 is a DNA-binding component of the ISGF3 complex induced
by α-IFN. ISGF3 contains STATs — 1 and 2 — as well as IRF-9. The observation that two important components of ISGF3 are controlled by NF-κB highlights an interdependent relationship between these transcription factor families. Perhaps not surprisingly, others have shown the IRF-7 promoter contains an NF-κB-binding site, mediating its response to cytokine stimulation (Lu et al., 2002). STAT1, to our knowledge, is not known to be NF-κB dependent. However, our findings that the STAT/IRF family members are downstream of NF-κB in viral infection suggests an important mechanism for how NF-κB plays an important role in innate immunity by facilitating signaling through the interferon pathway.

C. NF-κB AUTOREGULATION

Autoregulation recently has been recognized to be an important controller of network stability. Studies of synthetic gene circuits have shown, for example, that negative feedback reduces the variability in target gene expression by limiting the range over which the concentrations of the signaling molecules operate (Becskei and Serrano, 2000). Remarkably, in our analysis, members of the NF-κB activation pathway themselves are NF-κB dependent. Of the members of the NF-κB pathway identified in this study, the inhibitor of apoptosis (IAP-1)/cIAP-2 is an upstream regulator of caspase activation that associates with the tumor necrosis factor receptor 1 (TNFR1) and death domain-containing receptors. Inducible expression of IAP-1 may be one mechanism through which NF-κB exerts its antiapoptotic effect. In support of our analysis, IAP-1 previously has been shown to be transcriptionally regulated by NF-κB (Hong et al., 2000). The DNA-binding subunit, NF-κB2, is an NF-κB1 homolog encoded by a large, 105-kDa precursor that must be processed into its 50-kDa binding form. Like IAP-1, NF-κB2 is known to be regulated by NF-κB (Liptay et al., 1994). Although both NF-κB1 and NF-κB2 are inducible genes, in RSV-infected epithelial cells, expression of NF-κB2 is predominately NF-κB dependent. Unlike NF-κB1, NF-κB2 is a potent transcriptional activator in heterodimeric form with Rel A (Schmid et al., 1991). These observations suggest that epithelial stimulation by NF-κB-activating agents changes the distribution of Rel A heterodimers, allowing the cell to respond differently to subsequent signals. Another well-established NF-κB-dependent target is that of the IκB inhibitors themselves (Han et al., 1999a; Brasier et al., 2001). Activation of the IκB members is an autoregulatory feedback loop where NF-κB induces the synthesis of its own inhibitor to terminate its action. Previously, we showed that the NF-κB-IκBα autoregulatory loop was impaired in RSV-infected cells. This finding explains why IκBα was not identified in the present analysis (Jamaluddin et al., 1998). IκBε, in contrast to IκBα, is strongly upregulated by RSV infection and suggests stimulus-specific differences in the NF-κB-dependent expression
control of individual IκB subunits. Notwithstanding, the induction of IκBε may compensate for a relative deficiency in IκBα. The existence of multiple, independent NF-κB-IκB and NF-κB-BCL-3 inhibitory loops suggests that unregulated NF-κB activation is highly deleterious (Jamaluddin et al., 1998; Han et al., 1999a; Brasier et al., 2001).

D. OTHER FUNCTIONS

Our study has identified a number of other highly NF-κB-dependent genes that do not easily fit into a single biological pathway (Table I). The functional consequences of enhanced expression of 5′-aminolevulinate synthase, a rate-limiting enzyme in heme biosynthesis, is unknown to us but suggests a role for NF-κB in viral-regulated heme metabolism in nonerythroid cells. Conversely, the NF-κB dependence of the E2-ubiquitin-conjugating enzyme suggests that NF-κB activation may have an important role in determining cellular capacity to break down proteins regulated through the ubiquitin-proteasome pathway, a process important in cell-surface presentation of viral antigens in the context of MHC class I molecules. We were surprised to identify a human homolog of the Saccharomyces cerevisiae GCN1 gene, a protein controlling translational efficiency through modifying upstream activation of the eIF2 protein kinase (Marton et al., 1997). Gene expression/regulation studies of hGCN1 in the setting of viral infection have not been reported to our knowledge, although viral infections are known to profoundly influence translational regulation. Complement factor B, a hepatic acute-phase response factor important in the alternative complement pathway, is well known to be NF-κB inducible (Nonaka and Huang, 1990). However, viral induction of the alternative complement pathway and its role in response to infection have not been investigated. The induction of insulin-like growth factor binding protein 6 (IGF-BP6) suggests that viral-infected cells exert paracrine control on the mitogenic actions of IGF-II (Gabbitas and Canalis, 1997). Perhaps IGF-BP6 expression is beneficial to prevent local cellular proliferation in the presence of an infecting viral agent. Upregulation of the α subunit of the IL-15 receptor may suggest that viral-infected cells have distinct signaling phenotypes to cytokines as a result of NF-κB action. Finally, B94 was identified as a highly cytokine- and lipopolysaccharide (LPS)-inducible transcript (Sarma et al., 1992) that may play a role in angiogenesis, spermatogenesis, or myelogenesis. Our data suggest that B94 is also a viral-inducible transcript in epithelial cells through an NF-κB-dependent mechanism. The function of B94 in this context is unknown.

Recent work suggests that NF-κB has genomic actions, even in the absence of exogenous stimuli. This “constitutive” NF-κB activation may be important to inhibit apoptosis (reviewed in Barkett and Gilmore, 1999). For example, NF-κB appears to be required to maintain low levels of the Bcl-2 protein, A1, at levels
to prevent loss of mitochondrial transmembrane potential and apoptosis in macrophages (Pagliari et al., 2000). Alternatively, constitutive NF-κB activity may be important in cellular immortalization (Arsura et al., 2000). Although our experimental design cannot distinguish these or other potential roles for constitutive NF-κB activity, we interpret our data to mean that constitutive NF-κB appears to downregulate expression of the collagen genes and others. The mechanism (transcriptional or post-transcriptional) by which NF-κB influences the abundance of these genes will require further investigation. In this regard, we note that a recent study has implicated NF-κB in the post-transcriptional control of MyoD mRNA abundance (Guttridge et al., 2000); perhaps collagen is regulated in a similar way.

V. Expression Kinetics of the NF-κB-inducible Genetic Network

We preliminarily analyzed the kinetics of gene expression for the highly upregulated NF-κB-dependent genes taken from Table I. For this analysis, microarray data from our previously established database of RSV infection was taken corresponding to type II-like alveolar epithelial cells (A549) (Zhang et al., 2001) and laryngeal carcinoma cells (Hep2) (B. Tian, unpublished data). A hierarchical clustering/heat map analysis was performed and is presented in Figure 8. In Figure 8A, the A549 time course shows that several of the NF-κB-dependent genes exhibit anomalous behavior. For example, an “unknown” gene, cytochrome P (CYP)11B1, melanoma antigen-encoding gene (GAGE)-2, and tyrosine phosphatase were not expressed by A549 cells. In addition, the gene encoding a putative membrane protein was paradoxically regulated, being downregulated by RSV infection in A549 cells. The remaining genes were activated from 6–12 hours after RSV infection, a time when induction of NF-κB binding can be demonstrated by EMSA and immunofluorescence analysis (Garofalo et al., 1996; Tian et al., 2002). Figure 8B shows a similar analysis for the kinetics of NF-κB-dependent gene induction in Hep2 cells. Like A549 cells, the genes “unknown,” p450, and GAGE-2 were not expressed; unlike A549 cells, however, the melanoma growth stimulatory activity/growth-regulated oncogene (MGSA/GRO)-α and cholesterol hydroxylase genes were not expressed in the Hep2 cells. The remaining genes were regulated by RSV infection in a manner quite similar to that seen in A549 (compare Figure 8B with 8A). Together, these data indicate that cell type has a strong influence on the composition of a genetic network.

VI. Promoter Analysis of NF-κB-inducible Genes

The identification of NF-κB-inducible genes allows us to glean insights into the actions of this master regulator of inflammation. That these genes are all
downstream of the same transcription factor and have similar patterns of expression indicated to us that further analysis of their promoter organizations may yield some unifying insights into the mechanism for NF-κB-dependent gene control. One successful approach to identification of common regulatory sequences has been to analyze empirically groups of promoters based on similarities in expression patterns (Tavazoie et al., 1999; Pilpel et al., 2001). In this example, a K-means algorithm was used to cluster groups of genes whose expression was similar in data derived from cell-cycle progression in Saccharomyces. A local alignment tool identified 17 motifs in 12 different clusters, over half of which corresponded to known DNA-binding sites, suggesting that expression coclustering identifies promoters with common regulatory motifs (Tavazoie et al., 1999). To address whether similarities could be found in the NF-κB-inducible genes, we retrieved the promoter sequences of the most tightly coclustering expression group (Figures 8A and B). The promoter sequences were annotated and analyzed for the presence of transcription factor-binding sites. Although a number of methods for identifying the presence of transcription factor-binding sites exist, recent algorithms employing positional weight matrices (PWM) have significantly improved the reliability of computational prediction of transcription factor-binding sites. PWMs have improved the accuracy of prediction of specific transcription factor-binding sites by accommodating for the large, base-to-base variability in target binding sites. PWMs assign a weighted score for each nucleotide in the promoter sequence and generate an overall score as the sum of the matrix values. Moreover, the weighted matrix score is related to the Gibbs free energy of binding (Berg and von Hippel, 1987; Stormo, 2000) and therefore has biological significance.

For analysis of factors binding the proximal promoters, binding sites for all eukaryotic transcription factors were predicted using the high-stringency cutoff, to minimize the false positive and negative matches from the TRANSFAC database (Heinemeyer et al., 1999). To correlate these data with experimental determination of binding sites, we compared the TRANSFAC output with our in vitro binding studies on the IL-8 and RANTES promoters. For example, the analysis correctly predicted the known AP-1 (Vlahopoulos et al., 1999) and NF-κB binding sites (Brasier et al., 1998) of the IL-8 gene. Similarly, the known NF-IL6 and NF-κB binding sites from the RANTES promoter were predicted (Casola et al., 2000b). Computer-identified NF-κB binding sites are displayed in Table II. Here, we also searched the literature for previously identified NF-κB binding sites for experimental verification where possible. We were able to confirm the putative NF-κB-binding sites with the published literature for the GRO-α (Wood and Richmond, 1995), IL-8 (Brasier et al., 1998), RANTES (Casola et al., 2000b), IRF-7B (Lu et al., 2002), IAP-1 (Hong et al., 2000), NF-κB2 (Liptay et al., 1994), and IL-15Rα (Mariner et al., 2001) promoters. In the other genes, the absence of NF-κB sites may indicate that the binding site is
located outside of the 700 bp of 5′ flanking sequence that was analyzed or that the gene is indirectly regulated by NF-κB. More investigation will have to be done to distinguish between these possibilities.

We then analyzed matches for all predicted eukaryotic transcription factor-binding sites using the most stringent cutoffs for matrix similarity score in the

<table>
<thead>
<tr>
<th>Time: 0</th>
<th>6</th>
<th>12</th>
<th>24</th>
<th>36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>CYP1B1</td>
<td>CYP1B1H</td>
<td>GAGE-2</td>
<td>Unknown</td>
</tr>
<tr>
<td>Membrane protein</td>
<td>Tyrosine phosphatase</td>
<td>Unknown</td>
<td>Nuclear phosphoprotein</td>
<td>GCN-1</td>
</tr>
<tr>
<td>STAT1</td>
<td>IRF-7B</td>
<td>STAT1</td>
<td>IRF-9</td>
<td>Complement B</td>
</tr>
<tr>
<td>RANTES</td>
<td>chol 25 hydroxylase</td>
<td>5′ALA</td>
<td>RANTES</td>
<td>Unknown</td>
</tr>
<tr>
<td>Unknown</td>
<td>RANTES</td>
<td>E2-Ub conjugating</td>
<td>IκB</td>
<td>NF-κB</td>
</tr>
<tr>
<td>NF-κB2</td>
<td>NF-κB2</td>
<td>IGFBP-6</td>
<td>NF-κB2</td>
<td>IL-8</td>
</tr>
<tr>
<td>IL-8</td>
<td>MGSAn</td>
<td>IAP-1</td>
<td>IAP-1</td>
<td>B64</td>
</tr>
</tbody>
</table>

FIG. 8. Kinetics of NF-κB-dependent gene expression in epithelial cells. (A) A549 cells. Clustering and heat map analysis of the kinetics of NF-κB-dependent gene expression in type II-like alveolar cells following RSV infection. Data represent the mean of three independent time courses analyzed by high-density microarrays. Agglomerative hierarchical clustering was performed on average fluorescence intensity values from three independent experiments using the UPGMA technique. A heat map for each gene for the three independent experimental datapoints is shown at right. The color gradient used is dark gray (minimum value of 5 scaled fluorescence intensity units), light gray (middle value of 5000 scaled units), and white (maximum value of 10,000 scaled units). CYP, cytochrome P; GAGE, melanoma antigen-encoding gene; MGSAn, melanoma growth stimulatory activity. [Data from Zhang Y, Luxon BA, Casola A, Garofalo RP, Jamaluddin M, Brasier AR 2001 Expression of RSV-induced chemokine gene networks in lower airway epithelial cells revealed by cDNA microarrays. J Virol 75:9044–9058.]
TRANSFAC database. In the sequences analyzed, the predicted binding site density was one transcription factor for approximately every eight base pairs (7.9 ± 1.8). Unfortunately, this density of transcription factor-binding data made it difficult to see relationships between the various promoters. To help identify

FIG. 8. (B) Hep2 cells. Clustering and heat map analysis of the kinetics of NF-kB-dependent gene expression in Hep2 laryngeal carcinoma cells following RSV infection. [Data analysis as in Figure 8A.]
and visualize binding patterns, we developed an interactive, versatile data visualization/mining tool that we call GeneRep. GeneRep displays the location of transcription factor-binding sites as colored rectangles along their linear sequence. The closeness of the match (the “similarity score”) is indicated by the height of the rectangle. To help identify the location of the NF-κB binding sites, they appear as solid black in Figure 9. The NF-κB, IL-8, IGF-BP6, and Gro-α promoters all contained high-affinity binding sites in the proximal 100 bp of the TATA box. Preliminary analysis did not reveal any consistent relationship of the NF-κB-binding site with binding sites for AP-1, IRF, or NF-IL6. More systematic analysis will be required to identify any common features for this subclass of inducible promoters.

VII. Conclusions

NF-κB is a highly inducible transcription factor that controls hepatic acute-phase response, innate and adaptive immunity, and cellular survival through the induction of genetic networks. We reviewed our work using a high-density microarray analysis of a tightly regulated cell model expressing a specific NF-κB inhibitor to identify its downstream gene network. Our findings

FIG. 9. Promoter analysis of NF-κB-dependent network. Graphical display for NF-κB-dependent promoter. Transcription factor binding sites are represented by shaded rectangles whose location is dependent upon distance from the transcription start site. The height of the rectangle is directly proportional to the TRANSFAC matrix similarity score, so the significance of the match can be easily determined by visual inspection. The rectangles are shaded, based on the DNA composition for each site. The transcription start site is indicated by a vertical arrow and the location in nucleotides is indicated below each arrow. The high-stringency NF-κB binding sites are indicated by double asterisks.
suggest that NF-κB regulates expression of distinct genetic networks of constitutive genes whose expression are inhibited further by viral infection as well as controlling a distinct subset of those that are viral inducible. Moreover, these data suggest that NF-κB is a upstream regulator of RSV-inducible gene expression through controlling expression of the proteins involved in interferon signaling (STAT/IRF), perhaps providing insights into mechanisms of how NF-κB controls the innate immune response. More work will be required to understand the cell type-specific influences on expression of NF-κB networks, so that members of this network can be comprehensively identified. Further analysis is required to understand whether a common promoter architecture of an NF-κB-dependent gene can be identified, which genes are indirectly controlled by NF-κB, and, in these, the mechanisms for this regulation.

ACKNOWLEDGMENTS

The authors would like to thank M. Gossen and H. Bujard for tTA plasmids (University of Heidelberg) and the University of Texas Medical Branch Genomics Core Laboratory (T. Wood, Director) for performing the arrays. This project was supported by grant R21 AI48163 from the National Institute of Allergy and Infectious Diseases (NIAID) and, in part, by R01 AI40218 (to A.R.B.) and AI 15939 (to R.P.G.), Child Health and Human Development (R30HD 27841), and #P30 ES06676 from the National Institute of Environmental Health Sciences (to R.S. Lloyd, University of Texas Medical Branch).

REFERENCES

Barkett M, Gilmore T 1999 Control of apoptosis by Rel/NF-kB transcription factors. Oncogene 18:6910–6924


Carey M 1998 The enhanceosome and transcriptional synergy. [Review] [20 refs]. Cell 92:5–8


Hsu H, Huang J, Shu HB, Baichwal V, Goeddel DV 1996 TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. Immunity 4:387–396


Karin M 1999 The beginning of the end: IkB kinase (IKK) and NF-kB activation. J Biol Chem 274:27342


Li J, Brasier AR 1996 Angiotensinogen gene activation by Angiotensin II is mediated by the Rel A (NF-kB p65) transcription factor: one mechanism for the renin angiotensin system (RAS) positive feedback loop in hepatocytes. Mol Endocrinol 10:252–264


Marton MJ, Vazques de Aldana CR, Qium H., Chakraburtty K, Hinnebusch AG 1997 Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosome in activation of eIF2 alpha kinase GCN2. Mol Cell Biol 17:4474–4489

Matsushima K, Oppenheim JJ 1989 Interleukin 8 and MCAF: novel inflammatory cytokines inducible by IL-1 and TNF. Cytokine 1:2–33


Perkins ND, Felzien LZ, Betts JC, Leung K, Beach DH, Nabel GJ 1997 Regulation of NF-kB by cyclin-dependent kinases associated with the p300 coactivator. Science 275:523–527


Teran LM, Noso N, Carroll M 1996 Eosinophil recruitment following allergen challenge is associated with the release of the chemokine RANTES into asthmatic airways. J Immunol 157:1806–1812


Wang C, Deng L, Hong M, Akkaraju GR, Inoue J, Chen ZJ 2001 TAK1 is a ubiquitin-dependent kinase of MKK and IKK [see comments]. Nature 412:346–351

Yie J, Merika M, Munshi N, Chen G, Thanos D 1999a The role of HMG I(Y) in the assembly and function of the IFN-beta enhanceosome. EMBO J 18:3074–3089
Zhang SQ, Kovalenko A, Cantarella G, Wallach D 2000 Recruitment of the IKK signalosome to the p55 TNF receptor: RIP and A20 bind to NEMO (IKKgamma) upon receptor stimulation. Immunity 12:301–311
Role of Defective Apoptosis in Type 1 Diabetes and Other Autoimmune Diseases

TAKUMA HAYASHI AND DENISE L. FAUSTMAN

Immunobiology Laboratory, Massachusetts General Hospital, and Harvard Medical School, Charlestown, Massachusetts 02129

ABSTRACT

Lymphocyte development, selection, and education are strictly controlled to prevent autoimmunity, with potentially autoreactive cells being removed by apoptosis. Dysregulation of apoptosis is a central defect in diverse murine autoimmune diseases. In murine models of autoimmune lupus, for example, mutations in the death receptor Fas (CD95) or in its ligand, FasL (CD95L), have been identified and shown to render lymphoid cells resistant to apoptosis. In contrast, select lymphoid subpopulations of mice with autoimmune diabetes manifest an increased susceptibility to apoptosis as a result of impaired activation of the transcription factor nuclear factor-kappa B (NF-κB), which normally protects cells against tumor necrosis factor-alpha (TNF-α)-induced apoptosis. The genetic basis of this defect in NF-κB activation is a mutation in the promoter-enhancer region of a gene that encodes an essential subunit (LMP2) of the proteasome. Although no specific genetic defects have been identified in most common forms of human autoimmune disease, functional assays consistently demonstrate heightened apoptosis attributable to multiple death signaling pathways.

I. Introduction

Autoimmunity encompasses a diverse group of diseases that are defined clinically by the target organ or tissue destroyed. Rheumatoid arthritis and type 1 diabetes mellitus (also known as insulin-dependent or juvenile-onset diabetes), for example, result from a presumed T-cell attack on the joints and insulin-secreting beta (β) cells of the pancreas, respectively. Although the clinical manifestations of each autoimmune disease are distinct, the underlying genetics of these conditions are similar, with most showing an association with the human leukocyte antigen (HLA; also known as the human major histocompatibility complex, or MHC) region of the genome or with nearby non-HLA loci (Becker et al., 1998).

Apoptosis may play a role in two different aspects of autoimmune disease. First, controlled apoptotic cell death contributes to normal T-cell selection and education. Thus, interruption of this process might result in the generation of...
autoreactive cells. Second, apoptosis might represent a lymphocyte-independent mechanism of organ or tissue destruction. To date, most experimental data as well as identified genetic defects that promote or impair apoptosis have implicated abnormal T-cell selection and development in autoimmunity. Although a target cell apoptotic defect, possibly involving the Fas death receptor, has been proposed to affect the pancreatic islets of individuals with type 1 diabetes (Chervonsky et al., 1997; Itoh et al., 1997; Amrani et al., 1999; Suarez-Pinzon et al., 1999), other studies have suggested that apoptosis is not a major mechanism of β-cell destruction (Kang et al., 1997, 1998; Kim et al., 1999; Pakala et al., 1999; Thomas et al., 1999; Kim et al., 2000; Restifo, 2000). This chapter will focus on the role of apoptotic defects that affect education of the lymphoid system in autoimmunity.

A prominent feature of autoimmunity is the failure of autoreactive cells, either during development or subsequently, to undergo negative selection and die. Such apoptotic defects in humans and mice result in autoreactivity and may lead to marked lymphoproliferation. In certain instances, these defects have been attributed to mutations in the genes for proteins that function in apoptotic signaling pathways. One such example is the lpr/lpr mouse, a model of human systemic lupus erythematosus (SLE), in which defective apoptosis results in lymphoproliferation and generalized autoimmunity. These animals harbor a spontaneous mutation in the gene for Fas (Watanabe-Fukunaga et al., 1992; Watson et al., 1992; Mountz et al., 1996), a cell-surface molecule also known as CD95 that belongs to the tumor necrosis factor receptor (TNF-R) superfamily. Similarly, the gld/gld mouse, which also manifests a lupus-like autoimmune disease, harbors a point mutation in the intracellular domain of the Fas ligand (FasL) (Allen et al., 1990; Lynch et al., 1994; Ramsdell et al., 1994; Takahashi et al., 1994). The identification of these autoimmunity-associated defects in the Fas signaling pathway stimulated a search for similar mutations in humans with lupus. However, only individuals with a rare form of lupus associated with diffuse lymphoproliferation have been shown to possess a mutation in the FasL gene (Wu et al., 1996a). Only patients with the rare Canale-Smith syndrome or autoimmune lymphoproliferative syndrome have been found to harbor a Fas mutation (Rieux-Laucat et al., 1995; Drappa et al., 1996). Not unexpectedly, the lymphoproliferation apparent in these patients resembles that in lpr/lpr and gld/gld mice and is thought to result from the failure of select lymphocyte populations to undergo apoptosis. Most individuals with lupus do not appear to harbor mutations in the Fas or FasL genes. Indeed, lymphocytes from such individuals manifest an increased susceptibility to apoptosis in vitro as well as increased FasL expression (Emlen et al., 1994; Mysler et al., 1994; Desai-Mehta et al., 1996; Koshy et al., 1996; Wu et al., 1996a; Kovacs et al., 1997; Lorenz et al., 1997; Wong et al., 1999).
In most spontaneous forms of human or murine autoimmunity, severe lymphoproliferation is not a prominent feature of the disease. Indeed, we have shown that the pathogenic cells may manifest an increased susceptibility to apoptosis. In the nonobese diabetic (NOD) mouse, for example, a spontaneous model of human type 1 diabetes, lymphocytes are more susceptible to TNF-α-induced apoptosis than are lymphocytes from control animals. This results from a defect in the activation of nuclear factor-kappa B (NF-κB) (Hayashi and Faustman, 1999), a transcription factor that protects against TNF-α-induced cell death. In addition to the accelerated apoptosis, there is increased FasL expression exhibited by peripheral blood lymphocytes from humans with lupus in vitro (Wong et al., 1999). The genetic basis of these human defects remains unknown.

Members of the TNF-R superfamily appear to play an important role in autoimmune disease. These proteins comprise an extracellular domain consisting of cysteine-rich motifs, a transmembrane domain, and a cytoplasmic tail (Liang and Fesik, 1997; Wallach et al., 1999).

Activation of NF-κB protects cells against TNF-α-induced apoptosis but this transcription factor also contributes to cell death mediated by Fas (Quaaz et al., 1999), another TNF-R family member. In addition, NF-κB activation in response to TNF-α may contribute to FasL expression (Hsu et al., 1999). The interplay between these various overlapping apoptotic pathways may explain why the apoptotic defects associated with autoimmune disease confer phenotypes of enhanced or diminished T-cell selection.

II. Genetic Risk Factors for Type 1 Diabetes Located in the MHC Region of the Genome

Genetic risk factors for type 1 diabetes map to the MHC region of the genome. In both human type 1 diabetes and two rodent models of this disease (the NOD mouse and BB rat), pancreatic β cells are selectively destroyed as a result of a chronic autoimmune reaction (Figure 1A and B) (Crisa et al., 1992; Rabinovitch and Skyler, 1998). The MHC region of the genome contains immune response genes that are important for T-cell education and for antigen presentation by both MHC class I and class II molecules. Studies of both humans and rodents have suggested that the centrally located MHC class II genes confer the greatest statistical risk for autoimmune disease. However, functional derangement of MHC class II genes has not been demonstrated in humans with autoimmune disease. In contrast, cellular abnormalities in expression of maturation markers or in antigen presentation have been detected in both NOD mice and diabetic humans. These defects include reduced expression of the maturation antigen CD45 and a reduced abundance of conformationally correct complexes of MHC class I molecules and self-peptides on the cell surface (Faustman et al., 1989,1991; Smerdon et al., 1993; Jansen et al., 1995).
Evidence based on functional assays suggests that human autoimmune diseases are associated with impairment of antigen processing controlled by the MHC. Thus, cytosolic extracts of lymphocytes from either humans with type 1
diabetes or NOD mice exhibit altered patterns of cleavage of test substrates by the proteasome. This results in the generation of peptides that are poorly suited for assembly with MHC class I molecules (Faustman et al., 1989, 1991; Smerdon et al., 1993; Jansen et al., 1995). In addition, lymphocytes of individuals with diverse autoimmune diseases — including type 1 diabetes, multiple sclerosis, and rheumatoid arthritis — manifest a reduced expression of peptide-loaded MHC class I molecules on their surface (Faustman et al., 1991; Fu et al., 1993; Li et al., 1995). Moreover, clinical studies have shown that the antigen presentation defect correlates with disease expression in identical twins with type 1 diabetes (Faustman et al., 1991). The genes responsible for antigen processing map to the MHC region of the genome, suggesting that abnormalities in this region might underlie these various conditions.

Candidate genes in the MHC region of the genome in humans and rodents that might be responsible for the antigen presentation defects associated with autoimmune disease include those for the TAP peptide transporters and the LMP proteasome subunits. Thus, for example, both LMP2 and LMP7 are encoded by genes located in the MHC region of the genome (Figure 2). These proteins are expressed constitutively in most cell types but their expression is markedly increased in antigen-presenting cells (APCs) or lymphoid cells in response to exposure to interferon-gamma (γ) (Fruh et al., 1992; Van Kaer et al., 1994; Hisamatsu et al., 1996; Griffin et al., 1998). Knockout (KO) mice that lack specific TAP or LMP genes exhibit abnormal T-cell selection and autoreactivity against transplants of syngeneic normal tissue (Aldrich et al., 1994; Glas et al., 1994; Van Kaer et al., 1994; Wakatsuki et al., 1994).

Ubiquitin-dependent proteolysis mediated by the proteasome, a multisubunit adenosine triphosphate (ATP)-dependent protease, plays important roles in various cellular processes, including cell-cycle progression, gene transcription, and signal transduction (Goldberg, 1995; Coux et al., 1996). In many instances, the target protein is marked for degradation or processing by both phosphorylation and ubiquitination. Cleavage of endogenous proteins by the proteasome also generates small peptide fragments that contribute to T-cell education as a result of their presentation by MHC class I molecules. Although, in general, the proteasome exhibits minimal variability in substrate selectivity and subunit composition, incorporation of the LMP2 and LMP7 subunits during assembly of the proteasome changes its specificity for self-proteins in such a manner that the suitability of the generated peptides for presentation in the peptide-binding groove of MHC class I molecules is increased (Belich et al., 1994; Gaczynska et al., 1996). The abundance of LMP2 mRNA in lymphocytes derived from NOD mice is reduced, compared with that in lymphocytes from control animals (Figure 2) (Yan et al., 1997), which likely explains, at least in part, the altered T-cell education toward self apparent in these mice.
III. The NOD Mouse: A Spontaneous Model of Type 1 Diabetes

Type 1 diabetes usually is caused by T-cell-mediated autoimmunity, with a prediabetic state characterized by the production of autoantibodies specific for proteins expressed by pancreatic β cells, including insulin. In general, the autoantibodies recognize intracellular proteins and likely are generated in response to islet death. The NOD mouse frequently is studied as a rodent model of human type 1 diabetes. The etiology of diabetes in the NOD mouse is complex and multifactorial (Delovitch and Singh, 1997; Rabinovitch, 1998; Atkinson and Leiter, 1999). Both CD4+ and CD8+ T cells mediate the autoimmune response, with underlying functional defects being present in bone marrow-derived APCs. Many CD4+ and CD8+ T-cell lines and clones with diabetogenic potential that are targeted to a variety of identified and unidentified antigens have been established from both the islets and spleen of NOD mice. Destruction of
pancreatic β cells appears to be mediated by both necrotic and apoptotic death triggered by invasion of islets by leukocytes, a process referred to as insulitis (Rabinovitch, 1998). Although insulitis is not apparent in NOD mice up to 3 weeks of age, its prevalence increases in both female and male animals after 5 weeks of age. A clear sex difference is observed with respect to the onset of diabetes, however (Figure 1C). In NOD females, the onset of diabetes occurs as early as 10 weeks, with the number of affected animals increasing with age (Makino et al., 1980). The cumulative prevalence of diabetes in NOD females by 50 weeks of age is ≈ 70–80%. In contrast, only about 20% of NOD males are affected by diabetes at this age. The large numbers of leukocytes apparent in the islet infiltrates of NOD mice are suggestive of lymph node formation around islets (Figure 1A and B). A strain-specific characteristic of NOD mice is the accumulation of many T lymphocytes in peripheral lymphoid organs, the pancreas, and submandibular salivary glands. This T-cell accumulation may reflect low interleukin (IL)-2 concentrations and the resistance of thymocytes and peripheral T cells to the induction of apoptosis. Such apoptotic resistance may be an early phenotype of lymphoid lineages prior to disease initiation (Lamhamedi-Cherradi et al., 1998).

Type 1 diabetes in the NOD mouse, like that in humans, exhibits a marked genetic component that maps to the MHC region of the genome. We have identified a specific proteasome defect in NOD mouse lymphocytes that results from downregulation of expression of the LMP2 proteasome subunit (Figures 2 and 3) (Hayashi and Faustman, 1999), which is encoded by a gene located in the MHC genomic region. This defect both prevents the proteolytic processing required for the production and activation of NF-κB, which plays an important role in immune and inflammatory responses, and increases the susceptibility of the affected cells to apoptosis induced by TNF-α (Figure 4). The proteasome dysfunction in NOD mice is both tissue and developmental stage specific; it is not apparent in islet cells.

IV. Defects in Proteasome-mediated NF-κB Activation and T-cell Education in NOD Mice

The proteasome mediates the processing and activation of the transcription factor NF-κB (Figure 5). NF-κB is activated in response to various extracellular stimuli, including IL-1, lipopolysaccharide, and TNF-α (Thanos and Maniatis, 1995; Verma et al., 1995; Baeuerle and Baltimore, 1996; Baldwin, 1996). It contributes to regulation of the gene expression for cytokine production, cell adhesion, lymphocyte maturation, and protection from TNF-α-induced apoptosis, as well as antigen processing and presentation by MHC class I molecules (Bohnline et al., 1988; Cross et al., 1989; Tan et al., 1992; Beg and Baltimore, 1996; Van Antwerp et al., 1996). Insights into the various biological functions of
NF-κB have been provided by the generation and characterization of KO mice lacking either subunits of this protein or associated regulatory factors (Burkly et al., 1995; Kontgen et al., 1995; Weih et al., 1995; Franzoso et al., 1997; Bushdid et al., 1998; Caamano et al., 1998; Kanegae et al., 1998; Hu et al., 1999; Li et al., 1999a,b; Takeda et al., 1999).

Active NF-κB exists predominantly as a heterodimer composed of p65 (RelA) and either p50 or p52 subunits. The p50 and p52 subunits are generated constitutively but their abundance is increased markedly by various extracellular stimuli, including IL-1 and TNF-α. These proteins are generated as a result of the proteasome-mediated removal of the carboxyl termini of p105 and p100 precursors, respectively (Fan and Maniatis, 1991; Schmid et al., 1991; Palombella et al., 1994; Coux and Goldberg, 1998; Lin et al., 1998; Sears et al., 1998). In resting cells, NF-κB is sequestered in the cytoplasm as a result of its association with IκBα or other members of the IκB family of inhibitory proteins (Ghosh and Baltimore, 1990; Hayashi et al., 1993a,b). Cell stimulation results in the phosphorylation of IκBα by the IκB kinase (IKK) complex and its degradation by the ubiquitin-proteasome pathway, thereby allowing the p50–p65 or p52–p65 heterodimer to translocate to the nucleus and initiate transcription of target genes (Figure 3) (Ghosh and Baltimore 1990; Oeri et al., 1991; Palombella et al., 1994;
Complexes of p65 and p105 also have been detected but these do not appear to translocate rapidly to the nucleus in response to cell stimulation (Sun et al., 1994; Lin et al., 1998).

Our laboratory has sought to understand why, in type 1 diabetes, T cells treat pancreatic β cells as foreign. We therefore have attempted to understand the process of T-cell education to self-antigens and how this process is altered in individuals with type 1 diabetes. T-cell education requires the presentation of self-antigens, a task that is undertaken by “professional” APCs such as macrophages, dendritic cells, and B cells. Until recently, it was thought that autoimmunity results from the inappropriate activation of T cells by foreign antigens (e.g., viral proteins) that generate cross-reactivity with self-antigens, which was considered an MHC class II defect. However, we proposed, and presented evidence for the notion, in both NOD mice and humans with type 1 diabetes, that interruption of the presentation of self-antigens by MHC class I molecules underlies the development of autoimmune disease (Faustman et al., 1991). This proposal was based on the contention that such MHC class I-mediated presentation of self-peptides is essential for the development of normal tolerance. Previously, MHC class I proteins were thought to function primarily in the

![Image of Figure 4: Impaired granulocyte-macrophage (GM) colony formation and increased sensitivity to TNFα-induced apoptosis in NOD mouse spleen cells.](image-url)
presentation of peptides derived from foreign intracellular proteins, especially viral proteins, for the generation of cytotoxic T cells. Subsequent studies in transgenic mice deficient in chaperone proteins required for the intracellular assembly of MHC class I complexes confirmed the importance of self-peptide presentation by MHC class I molecules in T-cell education to self (Aldrich et al., 1994; Glas et al., 1994; Van Kaer et al., 1994).

In our attempt to discover the basis for the impairment in presentation of self-peptides by MHC class I molecules in the NOD mouse, we found that the abundance of LMP2 mRNA in lymphoid cells from these animals was markedly reduced, compared with that in control animals. This defect in LMP2 expression

FIG. 5. Impaired expression of LMP2 in NOD mouse splenocytes. (A) Schematic representations of 26S and 20S proteasomes. (B) Lysates of spleen cells from adult male (M) or female (F) BALB/c or NOD mice were subjected to immunoblot analysis with antibodies specific for the indicated 20S proteasome subunits or, as controls, with antibodies to various cyclin-dependent kinases (CDKs) or to the transcriptional factor TAF150.
in the NOD mouse was shown to be attributable, at least in part, to a specific mutation in the shared bidirectional promoter-enhancer region of the LMP2 and TAP1 genes in the MHC class II region of the genome (Figure 2). The reduced abundance of LMP2 interrupts the proteasome-mediated generation of self-peptides for presentation by MHC class I molecules and the consequent development of T-cell tolerance to self-antigens (Yan et al., 1997). It also prevents the processing of NF-κB precursor proteins and the degradation of IκBα required for activation of NF-κB (Hayashi and Faustman, 1999), events important for T-cell maturation and normal immune and inflammatory responses. The LMP2 expression defect in NOD mice is specific for lymphoid lineage cells and becomes apparent after 10 weeks of age (Hayashi and Faustman, 1999).

The interruption by the LMP2 defect in NOD mice of both self-peptide presentation by APCs as well as normal T-cell development — two phenotypes we had established as important in both murine and human autoimmune diabetes — suggests that the onset of LMP2 downregulation is an essential trigger for disease initiation. The expression of MHC class I molecules in islets is upregulated early during islet invasion by T cells in both humans and NOD mice with type 1 diabetes. This phenomenon probably defines target selection by augmenting self-antigen presentation, thereby promoting cytotoxic T-cell attack mediated by poorly educated, LMP2-deficient T cells.

V. Increased Sensitivity of NOD Mouse Lymphocytes to TNF-α-induced Apoptosis

Recent reports indicate that NF-κB is an important protector of cells from TNF-α-induced apoptosis (Beg et al., 1995). Embryos of mice lacking the NF-κB p65 subunit, IKKβ or IKKγ, manifest marked hepatic apoptosis that appears to result from the associated defects in NF-κB activation (Beg and Baltimore, 1996; Li et al., 1999b, Rudolph et al., 2000). The activation of NFκB by the ubiquitin-proteasome pathway also is thought to protect cells from TNF-α-induced cell death (Figure 3) (Beg and Baltimore, 1996; Van Antwerp et al., 1996; Wang et al., 1996; Wu et al., 1996b). The antiapoptotic effect of NF-κB is likely mediated by the activation of genes that encode cell survival-promoting factors.

We investigated the effect of TNF-α on the viability of adult NOD mouse lymphocytes, in which TNF-α-induced activation of NF-κB is impaired. Whereas incubation of BALB/c mouse splenocytes with various concentrations (2–20 ng/ml) of TNF-α for 24 hours had virtually no effect on cell survival, TNF-α induced a dose- and time-dependent decrease in the survival of splenocytes derived from male or female NOD mice (Hayashi and Faustman, 1999; Hayashi et al., 2000). Similarly, whereas incubation of BALB/c mouse splenocytes with TNF-α (10 ng/ml) for up to 48 hours had no effect on cell viability,
the survival of NOD splenocytes already was reduced markedly after incubation with the same concentration of TNF-α for only 12 hours (Hayashi and Faustman, 1999; Hayashi et al., 2000). The toxic effect of TNF-α on NOD mouse lymphocytes appeared more pronounced for female than for male animals. Exposure of lymphocytes from LMP2 KO mice to TNF-α also resulted in marked cell death (Hayashi and Faustman, 1999; Hayashi et al., 2000). Agarose gel electrophoresis confirmed that TNF-α induced a pattern of internucleosomal DNA fragmentation characteristic of apoptosis in lymphocytes from NOD mice and LMP2 KO, whereas it did not induce DNA fragmentation in those from BALB/c mice (Hayashi and Faustman, 1999). It is thus likely that the toxicity of TNF-α for NOD mouse lymphocytes is attributable to the NF-κB inactivation due to defective proteasome function.

TNF-α also reduced the viability of spleen cells derived from 7-day-old NOD mice but to a lesser extent than it did in cells derived from adult animals. It had no effect on the viability of spleen cells derived from 7-day-old BALB/c mice. Whereas TNF-α had no effect on the viability of cultured macrophages derived from 13.5-day BALB/c or NOD mouse fetal liver, it induced a dose- and time-dependent decrease in the viability of such cells derived from LMP2 KO mouse fetal liver at the same stage of development (Hayashi and Faustman, 1999). Similarly, TNF-α had no effect on the viability of cultured BALB/c or NOD mouse embryonic fibroblasts, whereas TNF-α treatment of such cells derived from LMP2 KO mice resulted in prominent cell death (Hayashi and Faustman, 1999, 2000). Although disruption of the NF-κB p65, IKKβ, or IKKγ genes is associated with marked abnormalities in liver development (Beg et al., 1995; Beg and Baltimore, 1996; Li et al., 1999b; Rudolph et al., 2000), hematoxylin-eosin staining of liver sections from 6-week-old NOD mice did not reveal any apparent defects (Hayashi and Faustman, 1999).

VI. Impaired Granulocyte-Macrophage Colony Formation by NOD Mouse Spleen Cells

NF-κB also plays an important role in the maturation of lymphocytes and monocytes. We therefore examined the development of the granulocyte-macrophage (GM) cell lineage with splenocytes isolated from 6-week-old NOD and BALB/c mice. Colony-formation assays revealed that, whereas GM-colony-stimulating factor (CSF) induced the formation of clusters of mature GMs in BALB/c mouse splenocytes, the formation of such clusters was impaired in splenocytes from NOD mice (Figure 4, A-D). Furthermore, whereas exposure of GM-CSF-treated spleen cell cultures from BALB/c mice to TNF-α had no effect on cell viability or colony development, TNF-α induced the death of all cells in NOD mouse cultures (Figure 4, E-H).
The specificity of the developmental defect and cytotoxic effect of TNF-α in the GM lineage of NOD mice was investigated by examining colony-forming units (CFUs) of erythrocytes in cultures of spleen cells derived from 6-week-old BALB/c and NOD animals. Erythrocyte colony formation appeared normal in erythropoietin-supplemented cultures of NOD mouse spleen cells, compared to that observed in spleen cells from BALB/c mice (Hayashi and Faustman, 1999). Moreover, TNF-α had no effect on erythrocyte colony formation, which is known to require NF-κB, in spleen cells from either BALB/c or NOD mice. These results suggest that a lack of NF-κB activation in GM precursors derived from NOD mice at 6 weeks of age impairs the maturation of these cells and renders them susceptible to the cytotoxic effect of TNF-α. In contrast, NF-κB appears to be functional in the erythrocyte lineage of these mice, which seem to develop normally and be resistant to TNF-α-induced apoptosis. Given that TNF-α had no effect on the viability of cultured macrophages derived from 13.5-day BALB/c or NOD mouse fetal liver, the proteasome defect in NOD mice appears to be specific for both cell type and developmental stage.

VII. Gender, Age, and Tissue Specificity of Proteasome Dysfunction and Disease Expression in NOD Mice

The prevalence of diabetes is markedly greater in NOD females than in NOD males. Most human autoimmune diseases also are expressed preferentially in females. Consistent with a role for defective proteasome activity and consequent impaired NF-κB function in NOD mouse diabetes, cytosolic extracts of splenocytes from male NOD mice were able to convert a small proportion of recombinant NF-κB p105 to p50. However, the product of this reaction appeared to differ in size slightly from that of the p50 subunit produced by extracts of BALB/c mice (Hayashi and Faustman, 1999). Splenocyte extracts from NOD females did not generate any detectable p50 protein in this assay. Furthermore, as mentioned previously, both the time course and dose-response relation for the effect of TNF-α on cell viability revealed that the sensitivity of splenocytes from NOD females to this cytokine was greater than that of cells from NOD males (Hayashi and Faustman, 1999; Hayashi et al., 2000).

The characteristics of KO mice that lack NF-κB subunits or LMP2 overlap partially with those of NOD mice (Van Kaer et al., 1994; Burkly et al., 1995; Kontgen et al., 1995; Weih et al., 1995; Horwitz et al., 1997). However, LMP2-deficient mice do not develop diabetes by 32 weeks of age (D.L. Faustman, unpublished observation), consistent with the contribution of multiple chromosomal regions to disease penetrance in both NOD mice and humans. The homogeneous nature of the gene defect in all tissues of LMP2 KO mice differs from the apparent developmental stage and tissue specificity of the proteasome defect in NOD mice, which might underlie target selection in disease expression.
LMP2-deficient and other KO mice with defects in the assembly of MHC class I molecules with self-peptides destroy transplanted syngeneic tissues from control animals (Li and Faustman, 1993; Vidal-Puig and Faustman, 1994; Freeland et al., 1998). Target cell loss thus might result from preferential direct attack by cytotoxic T lymphocytes in the early stages of autoimmune disease.

The marked proapoptotic effect of TNF-α in NOD mouse lymphocytes also suggested a possible role for this cytokine in early β-cell destruction in these animals. Such a mechanism of β-cell death would require that β cells exhibit the same proteasome defect as that apparent in NOD mouse lymphocytes. This defect is characterized by loss of LMP2 expression, aberrant NF-κB activation, increased sensitivity to the cytotoxic effect of TNF-α, and reduced expression of peptide-filled MHC class I molecules on the cell surface. However, one of the early pathological features of autoimmune diabetes in both humans and rodent models is hyperexpression of correctly assembled MHC class I molecules on the surface of β cells (Foulis, 1987; Ono et al., 1988; Weringer and Like, 1988; Hanafusa et al., 1990; Kay et al., 1991; Vivés-Pi et al., 1996; Stephens et al., 1997), a phenomenon that requires intact proteasome function. Studies of both humans and animals with diabetes or other autoimmune diseases suggest that discordance in the regulation of MHC-linked genes between tissues might confer target specificity for attack by cytotoxic T lymphocytes (Hayashi and Faustman, 1999).

Macrophages and fibroblasts derived from 13.5-day NOD mouse embryos exhibited normal cell growth and resistance to TNF-α cytotoxicity. In contrast, TNF-α exhibited a marked proapoptotic effect in the corresponding cell types derived from LMP2 KO mice (Hayashi and Faustman, 1999, 2000; Hayashi et al., 2000). TNF-α also induced a relatively small decrease in the viability of spleen cells derived from 7-day-old NOD mice but had no such effect on the corresponding cells from BALB/c mice. In contrast, lymphoid cells of splenic origin, lung macrophages (Kupffer cells), and GMs from 6- to 8-week-old NOD mice exhibit reduced LMP2 expression, impaired NF-κB activation, and increased sensitivity to the cytotoxic effect of TNF-α (Hayashi and Faustman, 1999). Furthermore, consistent with a role for the proteasome and NF-κB in normal cell growth, culture of spleen cells from 6-week-old NOD mice with GM-CSF failed to induce normal expansion of the GM cell lineage. The islets of Langerhans, liver, and erythrocytes of 6- to 8-week-old NOD mice appear normal. The ability of NOD mouse macrophages to activate regulatory T cells in an autologous mixed lymphocyte reaction also has been shown to be impaired (Atkinson and Leiter, 1999).

The age-dependent proteasome defect in the macrophages of NOD mice likely explains some of the important features of disease development in these animals. Thus, female NOD mice show no signs of autoimmunity up to 3 weeks of age. At 5 weeks and older, insulitis begins to appear. By 8 weeks of age,
autoantibodies are detectable. The insulitis gradually increases in intensity, with complete destruction of islets usually apparent by 30 weeks of age (Makino et al., 1980). Furthermore, the outcomes of various interventions and treatments in NOD mice are age dependent. For instance, the administration of TNF-α to animals older than 6 weeks sometimes prevents the development of diabetes, whereas the same treatment in animals younger than 4 weeks has no effect or a detrimental effect (Yang et al., 1994). Therefore, both the time course of the histopathology of autoreactivity and the paradoxical responses to TNF-α treatment parallel the altered developmental regulation of LMP2 expression and NF-κB activity in these animals.

VIII. Defective Proteasome Function and Autoimmunity

The ubiquitin-proteasome pathway plays an essential role in many important biological processes (Maniatis, 1999). Protein degradation by this pathway thus generates peptides for presentation by MHC class I molecules and either activates or inactivates transcription factors. In general, proteasome subunit composition varies minimally among eukaryotic cells. However, the interferon-γ-induced expression of the MHC-encoded proteasome subunits LMP2 and LMP7 is thought to promote the generation of endogenous peptides compatible with the peptide-binding cleft of MHC class I molecules (Akiyama et al., 1994; Belich et al., 1994). The MHC-encoded proteasome subunits also play a role in general proteasome function, including the processing and activation of NF-κB.

The defect in proteasome function in NOD mouse splenocytes is attributable to a loss of expression of the LMP2 subunit and was evident from the impaired proteolytic processing of the p105 precursor of the NF-κB subunit p50 in vitro as well as from the lack of degradation of phosphorylated IκBα in these cells in response to TNF-α. This defect confers sensitivity on the affected cells to the apoptotic action of TNF-α (Figure 6). The role of LMP2 in NF-κB activation was confirmed by observations that 1) cytosolic extracts of lymphocytes from LMP2 KO mice also failed to convert p105 to p50 and 2) only NOD mouse tissues that lack LMP2 subunit showed impaired activation of NF-κB and sensitivity to TNF-α-induced apoptosis (Hayashi and Faustman, 1999, 2000). The defect in LMP2 protein production in NOD mice is both developmental stage (age) and tissue specific. Dysfunction of a gene in the MHC region of the genome thus virtually abolishes the activity of a transcription factor that plays important roles in both immune and nonimmune cellular functions. The NOD mouse therefore represents a newly defined mosaic model of discordant MHC gene expression that exhibits marked proteasome dysfunction in an age- and tissue-specific manner.

The delayed maturation of lymphocytes and cytokine abnormalities apparent in NOD mice that spontaneously develop type 1 diabetes are mirrored, in part, by
the phenotypes of KO mice lacking NK-\(\kappa\)B subunits or LMP2 (Van Kaer et al., 1994; Sha et al., 1995; Beg and Baltimore, 1996; Snapper et al., 1996; Franzoso et al., 1997; Horwitz et al., 1997; Iotsova et al., 1997; Caamano et al., 1998; Tanaka et al., 1999). The clinical relevance of the phenotypes of the NOD mouse and of these various KO animals to human disease is supported by the existence of nearly identical cytokine and lymphocyte maturation defects in humans with type 1 diabetes.

In conclusion, we have demonstrated the existence of a marked defect in proteasome function in lymphocytes from autoimmune diabetes-prone NOD
mice. This defect results from a deficiency of the LMP2 subunit, which is encoded by a gene located in the MHC region of the genome. It results in both impaired processing of self-peptides for presentation by MHC class I molecules as well as the inability to activate NF-κB. A similar age-related defect in GMs is proposed to confer target specificity in autoimmunity toward tissues with intact LMP2 expression. Abnormal processing of intracellular proteins thus may contribute to the pathogenesis of type 1 diabetes.

REFERENCES


Bohnline E, Lowenthal JW, Siekevitz M, Franza BR, Greene WC 1988 The same inducible nuclear protein regulates mitogen activation of both the interleukin-2 receptor-α gene and type 1 HIV. Cell 53:827–836


Caamano JH, Rizzo CA, Durham SK, Barton DS, Raventos-Suarez C, Snapper CM, Bravo R


Cross SL, Halden NF, Lenardo M, Leonard WJ 1989 Functionally distinct NF-κB binding sites in the immunoglobulin κ and IL-2 receptor α chain genes. Science 244:466–468


Goldberg AL 1995 Functions of the proteasome: the lysis at the end of the tunnel [see comment]. Science 268:522–523


Hayashi T, Faustman D 1999 NOD mice are defective in proteasome production and activation of NF-κB. Mol Cell Biol 19:8646–8659


Horwitz BH, Scott ML, Cherry SR, Bronson RT, Baltimore D 1997 Failure of lymphopoiesis after adoptive transfer of NF-κB-deficient fetal liver cells. Immunity 6:765–772


Kang SM, Schneider DB, Lin Z, Hanahan D, Dichek DA, Stock PG, Bækkeskov S 1997 Fas ligand expression in islets of Langerhans does not confer immune privilege and instead targets them for rapid destruction [see comments]. Nat Med 3:738–743

Kang SM, Lin Z, Ascher NL, Stock PG 1998 Fas ligand expression on islets as well as multiple cell lines results in accelerated neutrophilic rejection. Transplant Proc 30:538

Kay TWH, Campbell II, Oxbrow I, Harrison LC 1991 Overexpression of class I major histocompatibility complex accompanies insulitis in the nonobese diabetic mouse and is prevented by anti-interferon-γ antibody. Diabetologia 34:779–785
Li Q, Van Antwerp D, Mercurio F, Lee KF, Verma IM 1999b Severe liver degeneration in mice lacking the IkB kinase 2 gene [see comments]. Science 284:321–325


Palombella V, Rando OJ, Goldberg AL, Maniatis T 1994 The ubiquitin-proteasome pathway is required for processing the NF-κB1 precursor protein and the activation of NF-κB. Cell 78:773–785


Stephens LA, Thomas HE, Kay TW 1997 Protection of NIT-1 pancreatic beta-cells from immune attack by inhibition of NF-κB. J Autoimmunity 10:293–298

Suarez-Pinzon W, Sorensen O, Bleackley RC, Elliott JF, Rajotte RV, Rabinovitch A 1999 Beta-cell destruction in NOD mice correlates with Fas (CD95) expression on beta-cells and proinflammatory cytokine expression in islets. Diabetes 48:21–28

Sun SC, Ganchi PA, Beraud C, Ballard DW, Greene WC 1994 Autoregulation of the NF-κB transactivator RelA (p65) by multiple cytoplasmic inhibitors containing ankyrin motifs. Proc Natl Acad Sci USA 91:1346–1350


Vidal-Puig A, Faustman DL. 1994. Tolerance to peripheral tissue is transient and maintained by tissue specific class I expression. Transplant Proc 26:3314–3316


Yan G, Fu Y, Faustman DL 1997 Reduced expression of Tap1 and Lmp2 antigen processing genes in the nonobese diabetic (NOD) mouse due to a mutation in their shared bidirectional promoter. J Immunol 159:3068–3080

Genomic Analysis of Glucocorticoid-regulated Promoters in Murine T-lymphoma Cells

LUCHEH,∗ † CELESTE FINNERTY,∗ WILLIAM C. GUSTAFSON,∗‡ CRAIG R. BUSH,‡
PING CHI,* HUIPING GUO,* BRUCE LUXON,†‡ ALAN P. FIELDS,*†
AND E. AUBREY THOMPSON* †

∗Sealy Center for Cancer Cell Biology; †Department of Human Biological Chemistry and Genetics; ‡Bioinformatics Program; †Department of Pharmacology and Toxicology,
University of Texas Medical Branch, Galveston, Texas 77555

ABSTRACT

We have undertaken a high-throughput analysis to identify targets of glucocorticoid regulation in P1798 murine T-lymphoma cells. G1/S-arrested cultures were treated for 8 hours with 0.1 μM dexamethasone (dex) in the presence and absence of 1 μg/ml cycloheximide. Untreated cultures and cultures exposed to cycloheximide alone were prepared as controls. RNA was isolated and gene expression analyzed using Affymetrix MG-U74A oligonucleotide arrays (Gene Chips®). Three independent experiments were performed. The data were analyzed using a variety of statistical and analytical approaches in order to identify primary transcriptional targets of the glucocorticoid receptor. We identified 44 genes that increase by >2-fold in both dex-treated and dex + cycloheximide-treated cultures (relative to control and cycloheximide-treated cultures) in three replicate experiments. Statistical analysis of control data indicate that the probability that a given probeset would, as a result of random error, increase >2-fold both in the presence and absence of cycloheximide in two independent experiments is approximately 7 × 10⁻¹⁹. We have retrieved from the Celera mouse genomic sequence 8 kb of promoter sequence, spanning 4 kb either side of the 5'-end of the cDNA from eight of the induced genes. These sequences were analyzed for potential glucocorticoid receptor binding sites. Five of these genes contain the sequence ACAnnnTGTnCT within 4 kb of the presumptive transcriptional start site. Eight control genes were selected at random and analyzed for the sequence ACAnnnTGTnCT. Two control genes had such sequences within 4 kb of the transcriptional start site.

I. Introduction

The glucocorticoid receptor is one of the best-characterized mammalian ligand-dependent transcription factors. Prevailing wisdom holds that the genomic effects of the glucocorticoid receptor result from interaction with gene promoters. This causes either an increase or decrease in the frequency of initiation of transcription of associated genomic sequences. Genes that are regulated in this fashion are said to be target genes and, more specifically, primary targets. In theory, primary targets are defined as genes that are regulated by direct interac-
tion with the receptor. In practice, they are defined as genes that are regulated in the presence of protein synthesis inhibitors. Primary targets may be either induced or repressed by virtue of interaction with the activated glucocorticoid receptor. Regulation of target promoters may result from direct binding of the receptor to its cognate DNA-binding elements. Promoter elements that function in this fashion have come to be called simple response elements (Diamond et al., 1990). The prototypic simple response element is the nucleotide sequence ACAnnnTGTnCT, first identified in the long terminal repeat of the mouse mammary tumor virus (MMTV) (Groner et al., 1983; Ringold et al., 1983; Beato, 1987). There is a second class of composite response elements (Diamond et al., 1990; Miner and Yamamoto, 1991) in which regulation occurs by indirect interaction that involves binding of the receptor to unrelated DNA-binding proteins (Gronemeyer, 1992; Miner and Yamamoto, 1992; Herrlich, 2001). These then target the receptor and its associated coactivators and corepressors to the transcriptional machinery.

Considerable complexity can be imagined — and, indeed, has been demonstrated — in the interactions between the glucocorticoid receptor and the glucocorticoid response elements with which it interacts. The molecular details of these interactions have been deduced in a number of cases and are actively studied in many laboratories. However, many important questions remain to be addressed concerning the role of glucocorticoid-regulated gene expression in cellular physiology. At the genomic level, it would be of interest to know whether direct interaction of the glucocorticoid receptor with a simple response element, along the model of the MMTV long terminal repeat, is the predominant mechanism for gene regulation, or whether composite-type interactions, involving indirect association with the promoter via other DNA-binding proteins, are more common among primary target promoters. It would be of interest to know whether all primary targets are regulated in all cell types, an idea that seems unlikely. If not, then what features of the promoters predict cell-specific versus ubiquitous responsiveness?

These questions require a complete definition of the primary glucocorticoid targets, ideally in several different cellular and physiological contexts. In addition, a detailed knowledge of the promoter structures of all these target genes will be essential for a complete understanding of how glucocorticoids regulate different genes under different circumstances. We are near to having some of this information in hand. High-throughput gene-profiling technology, although in its infancy, holds the promise of being able to identify all the glucocorticoid target genes. Recent release of a draft of the mouse genomic sequence provides a first opportunity to examine the promoters of primary glucocorticoid response genes.

We have undertaken an initial analysis of glucocorticoid-regulated gene expression in murine T-lymphoma cells. Affymetrix Gene Chips® were used to measure expression of 12,422 genes in cell cycle-arrested cells exposed to
dexamethasone (dex) in the presence and absence of cycloheximide. A statistical evaluation of the data was carried out to estimate the level of confidence that one might have in using this technology to identify target genes. The Celera® mouse genomic sequence was used to discover the promoters for the target genes and to ascertain to what extent these promoters contain potential glucocorticoid receptor binding sites. The results we have obtained are provocative but must be considered only a first step in the process of defining the entire repertoire of glucocorticoid-regulated genes in murine T-lymphoma cells. The clearest lesson that we have learned is that the technology is adequate to provide some initial insight into many of the questions raised above. However, several significant limitations in the theory and technology of genomic analysis must be overcome before a more-complete understanding will emerge.

II. Results

A. EVALUATION OF AFFYMETRIX GENE CHIP® TECHNOLOGY AS A MEANS OF IDENTIFYING GLUCOCORTICOID TARGET GENES

Our initial objective was to evaluate the Gene Chip® technology to determine whether this approach was sufficiently reliable and reproducible to permit identification of glucocorticoid-regulated genes in lymphoid cells of thymic origin. Addition of glucocorticoids to such cells causes apoptosis and/or cell-cycle arrest, depending upon the cell line and culture conditions employed. We were concerned about eliminating secondary changes in gene expression that might result from incipient cell death or from changes in cell-cycle distribution rather than from direct interaction between the glucocorticoid receptor and target genes. We have analyzed gene expression in G1-arrested P1798 T-lymphoma cells, which neither die nor undergo cell-cycle redistribution under the conditions we have employed (Rhee et al., 1995). Late G1-arrested cultures were treated for 8 hours with 0.1 μM dex in the presence or absence of 1 μg/ml cycloheximide. RNA was extracted from control cells (hereafter designated C), dex-treated cells (designated D), cycloheximide-treated cells (designated X), and cells treated with both dex and cycloheximide (DX).

This protocol was repeated three times; the individual experiments are identified as G49, G95, and G116. Within a given experiment, C49 refers to the control sample from the G49 experiment, X116 to the cycloheximide-treated sample for the G116 experiment, and so on.

The ability to identify large numbers of glucocorticoid-regulated genes depends upon the reproducibility of the analytical system. Consequently, our initial efforts focused upon evaluation of the extent to which reproducible results were obtained from replicate experiments. Our first objective was to identify and analyze genes that were not regulated by glucocorticoids. We used two Affy-
metrix parameters to identify such genes. The Affymetrix MicroArray Suite software (MAS 4.0) reports a number of parameters related to fluorescent intensity of hybridization of labeled RNAs. *Average Difference* corresponds more or less to the intensity of the signal, whereas *Absolute Call* is derived from an algorithm that purportedly designates individual RNAs as present, marginal, or absent. These parameters are used in reference to approximately 12,500 probesets, each of which corresponds to a known gene or expressed sequence tag (EST) sequence printed onto the murine MG-U74A gene chips.

An average of 5533 probesets (standard deviation (SD) = 594) were scored as present on each of the 12 chips, with a range of 4232 probesets present (DX95) to 6255 probesets present (D116). We initially excluded probesets that were scored as absent in all four chips from a given experiment (G49, G95, or G116). Spotfire was used to identify probesets that were increased by > 2.0 in dex-treated samples or decreased by > 0.5 in dex-treated samples. Probesets that conformed to these conditions in all three experiments were segregated as potential glucocorticoid-regulated genes. The remaining data were then sorted for probesets that were present on all 12 chips. The result was a dataset of 3170 probesets that were scored as positive on all 12 chips and did not appear to be reproducibly induced or repressed by glucocorticoids. These 3170 probesets formed our control dataset, which we analyzed to determine the amount of random variation in average difference for a given probeset in replicate analyses.

Figure 1A contains the results of three-dimensional (X,Y,Z) linear regression analysis in which the average difference (i.e., roughly the intensity of the hybridization signal) was plotted for each probeset in the control dataset of 3170 probesets from three control samples (C49, C95, and C116). As can be seen from the correlation coefficients (r²), the data exhibited a high degree of correspondence to a linear relationship, with very little scatter around the regression line for the data. This outcome indicates a high degree of reproducibility. More specifically, the data indicate that the average difference measured for a given probeset in one control experiment has a strong predictive value for the average difference of the same probeset in a replicate control experiment. The same relationship was obtained within the dex-treated, cycloheximide-treated, and dex + cycloheximide-treated samples (data not shown).

We felt that the most important use of the data from these 3170 control probesets was to analyze the degree of random variation within the average differences measured for a given probeset on the 12 chips. We wrote a program to calculate the probability that any two of the 12 measurements might vary by a user-defined amount (e.g., 2.0-fold up or down) through pairwise comparison of the average differences for a probeset in each of the 12 experiments by that in every other experiment. The regression line for this calculation, applied to 3170 probesets, is shown in Figure 1B. The equation that defines this line predicts that the probability that any pair of probesets will vary by > 2.0-fold, either up or
FIG. 1. Statistical analysis of average-difference data from control probesets. As described in the text, 3170 probesets were selected as controls, based on presence in all 12 chips and lack of response to dexamethasone (dex). Panel A contains XYZ linear regression of 3170 probesets from the control samples of the G49, G95, and G116 experiments. The coefficient of linear correlations, $r^2$, was calculated using the SigmaPlot 2001 statistics package and is shown for each comparison. Panel B contains the regression line calculated by determining the probability that any two pairs of a given probeset from a control dataset would vary by a given fold change, either up or down. Panel C illustrates the regression line calculated for the probability that for any given probeset, the average differences would vary by a given fold change in the presence and absence of cycloheximide in one experiment (filled circles), any two of three experiments (open circles), or in all three experiments (filled triangles).
down, in the control dataset was 0.096. That number predicts that, for a given pair of chips, having something on the order of 5000 probesets scored as present in the sample, random variation would result in about 500 probesets that would vary by > 2-fold. However, if variation is random, the probability that a given probeset will vary by a defined amount in two separate experiments would be multiplicative. Thus, the probability that a given control probeset would vary by > 2.0-fold in both control (C) versus dex (D) and in cycloheximide (X) versus dex + cycloheximide (DX) for the same experiment would be (0.096)^2. We calculated the probability that any control probeset would vary by a user-defined amount in both C versus D and X versus DX for a given experiment. The regression line that describes that relationship is shown in Figure 1C (filled circles). The equation that defines this curve predicts that the probability that a control probeset will vary by > 2.0-fold in both C versus D and X versus DX in a single experiment approaches zero at ≈ 2.5-fold. We also calculated the probability that a control probeset would vary by a given amount in C versus D and X versus DX in any two of three experiments (open circles) or in all three experiments (filled triangles). As can be observed from visual examination of the curves in Figure 1C, the probability that a control probeset will vary in C versus D and X versus DX in two of three experiments approaches zero around 1.8-fold change, whereas the probability is nearly zero that a control probeset will vary by > 1.5-fold in C versus D and X versus DX in all three experiments (filled triangles).

This kind of analysis predicts that when one analyzes an experiment (e.g., G49 or G95) consisting of four chips with average datasets of 5000–6000 probesets present, perhaps 50–100 probesets will vary by > 2.0-fold in both dex and dex + cycloheximide-treated samples. The probability that the same probeset will yield the same outcome, as the result of random variation, in two separate experiments (e.g., that D49/C49 > 2.0 AND DX49/X49 > 2.0 AND D95/C95 > 2.0 AND DX95/X95 > 2.0 or D49/C49 < 0.5 AND DX49/X49 < 0.5 AND D95/C95 < 0.5 AND DX95/X95 < 0.5) would be (0.096)^8 or about 8.5 × 10^-5. This probability predicts that something on the order of one probeset will, as a result of random variation, change by > 2-fold in both dex-treated and dex + cycloheximide-treated samples in two independent experiments. The probability that a given probeset will behave in this fashion in three independent experiments, as a result of random variation, is ≈ 7 × 10^-9, far less than one probeset per dataset of 5000 probesets present.

This initial evaluation of the data from three experiments of four chips each suggests two important considerations in the design and interpretation of experiments of this sort. Initially, comparing three sets each of control data — dex, cycloheximide, and dex + cycloheximide-treated — indicates that the results are highly reproducible. In this regard, it should be kept in mind that these three experiments were done over ≈ a 6-month period. The second consideration that
one must deal with in this kind of experiment is the degree of random variation. Some sense of the degree of random variation must be made if one is to have confidence that the number of replicate experiments is sufficiently great to allow statistically significant conclusions to be drawn. In our experience, random variation, defined as the probability that a given probeset will vary by > 2.0-fold in a pair of control experiments, varies from \( \approx 14\% \) to \( \approx 4\% \), depending on the cells and the conditions under which they are analyzed.

**B. IDENTIFICATION AND ANALYSIS OF GLUCOCORTICOID-REGULATED PROBESETS**

Our initial analysis of these data was performed using Affymetrix MAS 4.0 software, which has several problems that we needed to overcome. We developed a query, illustrated in Figure 2, to accommodate present/absent calls and negative average differences. Primary-induced probesets were identified as those in which, for a given experiment, \( C > 0 \) AND \( X > 0 \) AND \( D/C > 2.0 \) AND \( DX/X > 2.0 \) AND the probeset was scored as present in the D sample AND the probeset was scored as present in the DX sample. The rationale was that we would reject a probeset that appeared to be induced (\( D/C > 2 \)) if it was scored as absent in the samples in which it was supposed to be induced (\( D = \) absent).

Using this query, we identified 41 probesets that were induced in each of three experiments. We constructed an additional query (illustrated in Figure 2) to identify probesets that had positive average difference in the D and DX samples (i.e., \( D > 0 \) AND \( DX > 0 \)) AND were scored as present in D and DX AND had negative average differences in C or X (i.e., \( C < 0 \) or \( X < 0 \)). The query was constructed to allow any combination of these conditions. Using this query, we identified three additional probesets that have negative average differences in control and cycloheximide-treated samples but were induced by dex in the presence and absence of cycloheximide in each of three experiments. Thus, we identified 44 probesets that were induced by > 2.0-fold in both the presence and absence of cycloheximide in each of three experiments.

**C. ANALYSIS OF GENOMIC SEQUENCES OF TARGET GENES**

One obvious question that arises is whether the promoters of glucocorticoid-regulated genes contain recognizable features that might predict, *a priori*, that a given promoter would be induced or repressed by the glucocorticoid receptor. In an initial attempt to answer this question, we set about to examine the recently released draft of the Celera mouse genomic sequence to determine whether 1) we could identify promoters that correspond to the glucocorticoid-induced genes that we have described and 2) such promoters contain canonical glucocorticoid receptor binding sites. We excluded from this analysis any gene for which there was ambiguity concerning the 5′ end of the cDNA sequence. We included only...
those genes for which we could obtain cDNA sequence including at least 50 bp 5' to the ATG start codon of the mouse cDNA. If there were fewer than 50 bp in the mouse cDNA sequence, then there must be ≥ 50 bp of 5' untranslated
region (UTR) in the corresponding human cDNA and the ATG start codon must be in exon 1 of the human gene. In addition, we accepted only those genes that contained four or more contiguous, ordered exons including exon 1. Finally, we accepted only promoters with no string of NNN(N)x (with x being a variable and indeterminate number) between the putative GRE and the transcriptional start site. If no GRE was found and there was no string of NNN(N)x within 4 kb of the transcriptional start site, the promoter was scored as having no GRE. We have attempted to emphasize stringency in our initial analysis, at the expense of excluding some genes that clearly contained potential receptor binding sites but were otherwise ambiguous in their sequence or organization in the Celera draft database.

Forty-four genes were induced by glucocorticoids in each of three independent experiments. Of these, only eight genes corresponded to the strict genomic criteria that we defined to verify promoter structure. Five of these promoters were associated with low-abundance mRNAs: Src-suppressible C kinase (i.e., protein kinase C) substrate (SSeCKS), acid phosphatase 5, RhoB, eIF2a kinase, and phosphatidic acid phosphatase (Figure 3A). Three promoters were associated with high-abundance mRNAs: L29441, 70zpep, and TDAG8 (Figure 3B). TaqMan® probes and primer sets were designed for these eight genes and mRNA abundance was assayed by real-time polymerase chain reaction (PCR). Two internal standards were used, beta-actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Triplicate samples were analyzed using the ΔΔCt approach that yields mRNA abundance relative to the internal standard. As shown in Figure 3, the mRNAs corresponding to all eight genes were induced by dex in the presence and absence of cycloheximide.

Five of these eight promoters contained the sequence ACAnnnTGTnCT within 4 kb of the presumptive transcriptional start site, as shown in Table I. The Gene Chip® expression data for these eight genes are shown in Figure 4, which displays the mean and standard deviation of the average differences from three chips for control, dex, cycloheximide, and dex + cycloheximide. Also shown in Figure 4 are expression data for eight control genes for which we could obtain reasonable data concerning the transcriptional start site. As shown in Table I, two of the eight contained presumptive receptor binding sites.

III. Discussion

If one is to identify all of the genes that respond to a given stimulus, it will be necessary to develop reliable procedures for simultaneously measuring expression of many transcripts. The two techniques that offer most promise in this effort are serial analysis of gene expression (SAGE) (Velculescu et al., 1995) and high-density microchip arrays. Neither technique currently is constrained by widespread appreciation of the limitations. No significant consensus exists on
FIG. 3. Real-time polymerase chain reaction (PCR) analysis of glucocorticoid-induced genes. TaqMan® analysis was used to measure mRNAs corresponding to the eight glucocorticoid-induced genes shown in Table I. Experiments were performed in triplicate and ΔΔCt values calculated relative to beta-actin or glyceraldehyde-3-phosphate dehydrogenase (GAPDH), as described in the “Materials and Methods” section. Means plus standard deviations (SD) are shown for RNA extracted from midlog-phase P1798 cells and from cells that had been treated with dex for 24 hours. [For abbreviations, see Figure 4 legend.]
<table>
<thead>
<tr>
<th>Description</th>
<th>Sequence</th>
<th>Location</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Induced probeset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95022_at</td>
<td>AB020886: SSeCKS</td>
<td>ACAGCTTGCT</td>
<td>−2286</td>
</tr>
<tr>
<td>100151_at</td>
<td>L29441: Overexpressed in testicular tumor</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>92356_at</td>
<td>M90388: Protein tyrosine phosphatase (70zpep)</td>
<td>ACACCTGTCTCT</td>
<td>+1349</td>
</tr>
<tr>
<td>98859_at</td>
<td>M99054: Acid phosphatase 5, tartrate resistant</td>
<td>ACAGCTTGCTCT</td>
<td>−1097</td>
</tr>
<tr>
<td>96553_at</td>
<td>U39827:G protein-coupled receptor TDAG8</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>101030_at</td>
<td>X99963: rhoB</td>
<td>ACAAATGTAC</td>
<td>−3721</td>
</tr>
<tr>
<td>94941_at</td>
<td>AI243533: GCN2</td>
<td>ACAATATGTAC</td>
<td></td>
</tr>
<tr>
<td>98508_s_at</td>
<td>D84376: Phosphatidic acid phosphatase</td>
<td>ACAAAATGTACT</td>
<td>+3575</td>
</tr>
<tr>
<td>Control probeset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100144_at</td>
<td>X07699: Nucleolin</td>
<td>ACAGTCTTGCT</td>
<td>+1661</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACACCTTGACT</td>
<td>+1333</td>
</tr>
<tr>
<td>100156_at</td>
<td>D26090: CDC46</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>100131_at</td>
<td>X15830: Secretory granule neuroendocrine protein 1</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>100600_at</td>
<td>M58661: CD24a</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>104606_at</td>
<td>M55561: CD80</td>
<td>ACAAGCTTGCTCT</td>
<td>−648</td>
</tr>
<tr>
<td>94837_at</td>
<td>U67328: NIP1-like protein (NIP1L (A3))</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>94892_r_at</td>
<td>M27938: Male-enhanced antigen 1</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>94896_at</td>
<td>D90151: CArG-binding factor-A</td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

[Genes were selected from the Celera mouse genomic database, based upon the criteria described in the text. Sequences from 4 kb upstream to 4 kb downstream of the presumptive transcription start site were imported into Vector NTI and scanned for the sequence ACAnnnTGnCT.]
FIG. 4. Expression data for glucocorticoid-induced and control promoters. Using the Celera database, we were able to recover detailed and complete promoter sequence for eight glucocorticoid-induced genes and eight control genes, which were selected more or less at random. The mean and standard deviation (SD) average differences for control, dex, cycloheximide, and dex + cycloheximide-treated samples were calculated by Excel, as indicated. The genes are identified according to the following abbreviations: (Panel A) PA p’tase, phosphatidic acid phosphatase; eIF2a kinase, GCN2 eIF2alpha kinase; rhoB, small GTPase rhoB; SSeCKS, src-suppressible C kinase substrate; acid p’tase 5, tartrate-resistant acid phosphatase 5; TDAG8, putative G protein-coupled receptor; 70zpep, protein tyrosine phosphatase; and L29441, a protein of unknown function that is overexpressed in testicular tumors. The control genes shown in Panel B include SGNEP1, secretory granule neuroendocrine protein 1; CD80; CAg-BF-A, CAg-binding factor-A; MEA-1, male enhanced antigen 1; CDC46; NIPIL A3, NIP1-like protein; nucleolin; and CD24a.
how reliably either technique may be applied to identifying target genes. Our experience indicates that two related issues must be resolved if one wishes to use gene chips to this end. Initially, one must have a good idea of the extent of random variation within the assay. Statistical analysis is the only obvious way to address this question, which means that a large dataset must be obtained and analyzed to ascertain the frequency with which signal intensities from individual probesets vary at random. Having some appreciation of the variability of the data, one then can determine how many replicates are needed to achieve an acceptable level of assurance that a given probeset will respond in a predictable fashion. In our experiments, the probably that any pair-wise comparison of a single probeset will yield two numbers that vary by > 2-fold is almost 10%. Thus, for a given probeset, the probability that $D/C > 2.0$ or $< 0.5$ is $\approx 0.1$. Likewise, the probability that $DX/X > 2.0$ or $< 0.5$ is $\approx 0.1$. Since these probabilities are multiplicative, the probability that $D/C > 2.0$ and $DX/X > 2.0$ or $D/C < 0.5$ and $DX/X < 0.5$ is $0.01$ for a single experiment. The probability that this condition will prevail for two experiments is $0.1 \times 0.1 \times 0.1 \times 0.1$, or $1 \times 10^{-4}$. For three independent experiments, the probability would be $1 \times 10^{-4} \times 0.1 \times 0.1$ or $1 \times 10^{-6}$. So, from a statistical perspective, we have a very high degree of confidence that the genes that we have identified will behave in a reproducible fashion.

Many of the genes that we have identified contain canonical glucocorticoid response elements within 4 kb of the presumptive transcriptional start site. Five out of eight of the genes that were induced by > 2-fold contained such sequences, whereas two of eight control genes, selected at random, contained presumptive receptor binding sites. If we accept promoters that contain long strings of Ns, we have detected potential receptor binding sites in nine of 13 target genes. More will be said about the reliability of this analysis but one must have a reasonable degree of confidence in assigning as a target any gene that is induced > 2-fold in three independent experiments and contains a canonical glucocorticoid response element within the promoter. For those genes, the majority, for which promoter sequence is unavailable, we must rely on the very high degree of reproducibility of the data. We have confirmed eight out of eight genes by quantitative reverse transcription (qRT)-PCR but this is not a practical approach to screening expression of hundreds of genes. Therefore, it will be necessary to rely heavily on the gene chip data.

At this point, it is appropriate to comment on the Celera mouse genomic database. In our experience, this database must be approached with considerable caution. The filters that we have applied require that the genomic sequence must contain a minimum of four exons in sequence and must be devoid of long strings of unassigned bases (identified by the letter N, hence strings of Ns) within 4 kb of the presumptive start site. Slightly less than 20% of the sequences that we have identified meet these criteria. We feel quite confident in those promoters that we have been able to analyze but we were disappointed that we could not generate
a larger and more-reliable dataset. The majority (five of eight) of the induced
genes contain MMTV-like glucocorticoid receptor binding sequences. Four of
the six presumptive GREs were upstream of the transcriptional start site, none
closer than −1 kb, and two were downstream at +1.3 and +3.5 kb, respectively.
The GRE sequences were more or less equally associated with the + and −
strands of the gene. GRE sequences were found at a much-lower frequency
among control genes (2/8). One control gene, nucleolin, contained two GREs
downstream of the transcriptional start site and the CD80 gene contained a GRE
at −648 bp. We have previously shown that nucleolin expression is inhibited by
glucocorticoids in P1798 cells (Suzuki et al., 1992), almost certainly due to a
delayed, secondary effect that would not be apparent under the conditions used
in the present study. Glucocorticoids do not affect CD80 expression in dendritic
cells (Vieira et al., 1998). It will be interesting to determine why these genes are
not induced by glucocorticoids. However, it is clear from our analysis that the
presence of the sequence ACAnnnTGTnCT is not sufficient to convey induction
by glucocorticoids. It remains to be proven that the presumptive response
elements that we have identified in five of the eight glucocorticoid-induced genes
actually mediate the response. Although we were somewhat surprised that such
a high percentage of our glucocorticoid-induced genes contain GREs, we feel
that this observation must be interpreted with caution, since it remains to be seen
whether this kind of correlation will be maintained as we refine the analysis to
include more promoters. We are particularly interested in analyzing promoters
that are repressed by glucocorticoids. Unfortunately, we could identify only three
such promoters, using the criteria defined previously. None of these contained
GREs; however, the sample size is not sufficient for any conclusion to be drawn
from this result. For the present, there is not much more that can be made of the
database.

Several interesting points can be made with respect to the genes identified in
Table I. Two of these, rhoB and the putative G protein-coupled receptor TDAG8,
are known to be involved in apoptosis (Choi et al., 1996; Liu et al., 2001) and
have been reported to be induced by glucocorticoids in other cell lines (Choi et
al., 1996; Koukouritaki et al., 1999). Apoptosis is the normal fate of glucocortic-
oid-treated T cells. P1798 cells are unusual in that they do not undergo
apoptosis when treated with glucocorticoids in medium containing serum growth
factors, although such cells die rapidly when treated with glucocorticoids in
serum-free medium (Thompson, 1991). We are in the process of examining gene
expression profiles in G1-arrested cells exposed to dex in serum-free medium. It
is possible that we may, by comparing genomic responses to glucocorticoids in
the presence and absence of serum, identify downstream targets of serum growth
factors that attenuate the apoptotic response. Such principles could prove to be
important therapeutic targets to increase the sensitivity of malignant T cells to
glucocorticoid-mediated apoptosis.
One of the target genes in Table I encodes SSeCKS. SSeCKS is a cytoplasmic scaffolding protein (Wassler et al., 2001) involved in nuclear cytoplasmic trafficking of D-type cyclins (Lin et al., 2000). Overexpression of SSeCKS causes G1 arrest (Lin et al., 2000), which is the cellular phenotype of glucocorticoid-treated P1798 cells (Thompson, 1991; Rhee et al., 1995). Glucocorticoids also induce a phospholipid phosphatase, which may be an important target for glucocorticoids during lung maturation (Snyder et al., 1981) and in hepatocytes (Pittner et al., 1985).

Phosphatidic acid phosphatase generates diacylglycerol, the activator of classical and novel isozymes of protein kinase C (PKC). The prototypic classical member of this family is PKC alpha (PKCα), which was induced by glucocorticoids (data not shown). PKCα is known to be involved in proliferative control and may be responsible for activation of SSeCKS (Lin et al., 1996). Our data hint at potential cross-talk between protein kinase cascades and nuclear hormone receptor signaling pathways, whereby glucocorticoids stimulate transcription of 1) a phosphatidic acid phosphatase, thereby increasing diacylglycerol synthesis; 2) PKCα, which is stimulated by diacylglycerol; and 3) SSeCKS, which, when phosphorylated by PKCα, causes cytoplasmic sequestration of D-type cyclins and G1 arrest. This mechanism remains to be proved. It will be necessary to confirm that the relevant proteins are induced but abundant evidence exists that cross-talk between activator protein-1 (AP-1), a critical PKC target, and the glucocorticoid receptor is important in glucocorticoid signaling (Miner and Yamamoto, 1991; Herrlich, 2001). Our data suggest that there may be significant cross-talk between the glucocorticoid and PKC signaling pathways upstream of AP-1.

Kofler and coworkers have published a similar analysis of glucocorticoid regulation of gene expression in proliferating and cell cycle-arrested human CCRF-CEM cells (Tonko et al., 2001). There are several very important differences between their experimental approach and ours. They used Incyte chips, which contain significantly fewer probes than the Affymetrix chips (7074 versus about 12,500). They used a human cell line that responds much more slowly to glucocorticoids and undergoes apoptosis when treated with dex for long periods of time. Although they analyzed genes that were rapidly induced in G1-arrested cells, they did not use cycloheximide to block secondary effects. So, it would not be surprising to find that the genes that they identified are not identical to those we have identified. They identified only eight genes that were either induced or repressed when glucocorticoids were added to both asynchronous and G1-arrested cultures of CCRF-CEM cells. Presumably, these would include — but probably not be limited to — primary transcriptional targets in these cells. Not one of these eight genes was identified in our analysis. Furthermore, only one of the genes listed in Table I (acid phosphatase 5) was identified in any of their analyses.
Brad Thompson recently completed an analysis of CCRF-CEM cells. A preliminary comparison of his results indicates that there are a few genes that are induced in both P1798 and CCRF-CEM cells. However, it seems clear CCRF-CEM cells are very different from P1798 T-lymphoma cells in their glucocorticoid response. Perhaps one should not be surprised, in light of the very different glucocorticoid response phenotypes of these two cell lines: CEM cells die rather slowly but continue to proliferate to a considerable extent in dex, whereas P1798 cells immediately withdraw from the cell cycle and do not die when treated with dex in the presence of fetal bovine serum. However, this result would appear to speak to the question of whether or not there is a subset of primary transcriptional targets that always is regulated in every cell type. The answer to this question awaits additional analysis of glucocorticoid target genes in cell lines and primary cells. The data presented here are an initial step in this direction and represent only a subset of the data that will be required to define the transcriptional targets of the glucocorticoid receptor.

IV. Materials and Methods

A. CELL CULTURE

P1798 T-lymphoma cells were maintained in midlog-phase growth in RPMI1640 containing 2% fetal bovine serum. Initial G0 arrest was achieved by adding 0.1 μM dex for 24 hours. Under these conditions, P1798 cells do not die but instead undergo complete G0 arrest (Thompson, 1991). G0-arrested cells were washed with complete medium to remove dex and suspended in complete medium containing 2 mM thymidine. Under these conditions, 100% of the cells, relieved from the inhibitory effects of dex, will exit G0, traverse G1, and arrest at the G1/S interface due to the presence of 2 mM thymidine (Rhee et al., 1995). G1/S-arrested cells were exposed to 0.1 μM dex for 8 hours in the presence or absence of 1 μg/ml cycloheximide. No cell death occurs under these circumstances and since the cells are arrested at the G1/S interface, no cell-cycle redistribution can occur. Total RNA was extracted using RNAqueous kits from Ambion, according to the manufacturer’s protocols.

B. GENE CHIP® ANALYSIS

First-strand cDNA synthesis was performed using total RNA (10–25 μg), a T7-(dT)24 oligomer (5’ GCCAGTGAATTGTAATACGACTCACTATA-GGGAGGCGG-dT24 3’) and SuperScript II reverse transcriptase (Invitrogen). The T7 promoter, introduced during first-strand cDNA synthesis, directed the synthesis of cRNA using bacteriophage T7 RNA polymerase. The cRNAs were labeled with biotin during the T7 transcription. Biotin-labeled target RNAs were
fragmented to a mean size of 200 bases to facilitate their hybridization to probe sequences on the Gene Chip® (Affymetrix) array. Each target RNA sample initially was hybridized to a test array. This array contains a set of probes representing genes commonly expressed in the majority of cells (e.g., actin, GAPDH, hexokinase, 5S rRNA, B1/B2 repetitive elements). Test arrays confirmed the successful labeling of the target RNAs and precluded the use of degraded or nonrepresentative target RNA samples.

Hybridization was performed at 45°C for 6 hours in 0.1 M morpholeno-ethane sulfonic acid (MES), pH 6.6, 1 M sodium chloride, 0.02 M ethylenediaminetetraacetic acid (EDTA), and 0.01% Tween 20. Four prokaryotic genes (bio B, bio C, and bio D from the E. coli biotin synthesis pathway and cre, the recombinase gene from P1 bacteriophage) were added to the hybridization cocktail as internal controls. These control RNAs were used to normalize expression levels between experiments. Because they are added at varying copy numbers (Bio B, 1.5 pM; Bio C, 5 pM; Bio D, 25 pM; cre, 100 pM), they may be used to estimate relative abundance of RNA transcripts in the sample. Arrays were washed using both nonstringent (1 M sodium chloride (NaCl), 25°C) and stringent (1 M NaCl, 50°C) conditions prior to staining with phycoerythrin streptavidin (10 µg/ml). Gene Chip® arrays were scanned using a Gene Array Scanner (Hewlett Packard) and analyzed using Affymetrix MicroArray Suite 4.0 software.

C. DATA ANALYSIS

Data from individual chips were analyzed separately, using a combination of programs. However, chip-to-chip comparison via MAS 4.0 was not used, since this approach limits comparisons to pairs of samples. Initially, data were imported into Excel files. Affymetrix controls were removed and absolute calls were converted to a numerical value (absent = 0, marginal = 1, present = 2) to facilitate quantitative assessment of presence or absence in multiple samples. The data from three replicates of four chips each were combined in a single file, which was queried using SQL as follows.

For induced probesets from an individual experiment (e.g., G49, G95, or G116):

```sql
SELECT *
FROM g49
WHERE (c49 > 0 and x49 > 0 and d49/c49 = 2 and dx49/x49 = 2
and d49ac = 2 and dx49ac = 2)
  or (c49 < 0 and d49 < 0 and x49 > 0 and dx49/x49 = 2 and d49ac = 2
and dx49ac = 2)
  or (c49 < 0 and d49 > 0 and x49 < 0 and dx49 > 0 and d49ac = 2
and dx49ac = 2)
```
or (c49 > 0 and x49 < 0 and dx49 > 0 and d49/c49 > 2 and d49ac = 2 and dx49ac = 2); The following query was used to combine induced probesets in all three experiments:

```sql
SELECT *
FROM allthree
WHERE probeset in (select probeset from g49twoac2new)
and probeset in (select probeset from g95twoac2new)
and probeset in (select probeset from g116twoac2new); (Note: g49twoac2new is the result of first query.) The following query was used to select repressed probesets from individual experiments:

```sql
SELECT *
FROM g49
WHERE (c49 > 0 and c49ac = 2 and x49 > 0 and x49ac = 2 and d49 < = 0.5*c49 and dx49 < = 0.5*x49);  
```

D. REAL-TIME PCR

Applied Biosystems assays-by-design 20× assay mix of primers and TaqMan® MGB probes (FAM™ dye-labeled) were prepared for all target genes and mouse beta-actin. Primers were designed to span exon-exon junctions, to not detect genomic DNA. All primers and probes sequences were subject to BLAST search against the Celera mouse genome to confirm specificity. TaqMan® rodent GAPDH with VIC™ dye-labeled probe also was used as an internal control. The sequences of primers and probes of these genes may be obtained by contacting Huiping Guo (huiguo@utmb.edu). TaqMan® one-step RT-PCR master mix reagent kit was used. A validation experiment was performed to test the efficiency of the target amplification and the efficiency of the reference amplification for all primers and probes. All absolute values of the slope of log input RNA versus C(T) were < 0.1. Separate tubes (singleplex) one-step RT-PCR was performed using 5 ng of RNA. The cycling parameters for one-step PCR were RT 48°C for 30 minutes, AmpliTaq activation 95°C for 10 minutes, denaturation 95°C for 15 seconds, and annealing/extension 60°C for 1 minute (repeat 40 times) on ABI7700. Triplicate C(T) values were analyzed using the comparative C(T)(ΔΔC(T)) method, as described by the manufacturer.

ACKNOWLEDGMENTS

This work was supported in part by grant CA24347 to E.A.T. from the National Institutes of Health. The investigators wish to express their particular appreciation to Dr. Thomas G. Wood and Ms. Michelle Guigneaux of the University of Texas Medical Branch molecular genomics core facility for help with gene chip analysis. We are also grateful to Dr. E. Brad Thompson and Dr. Olivera Nesic for helpful discussions concerning these data.
REFERENCES

Gronemeyer H 1992 Control of transcription activation by steroid hormone receptors. FASEB J 8:2524–2529
Herrlich P 2001 Cross-talk between glucocorticoid receptor and AP-1. Oncogene 20:2465–2475
Koukouritaki SB, Gravanis A, Stournaras C 1999 Tyrosine phosphorylation of focal adhesion kinase and paxillin regulates the signaling mechanism of the rapid nongenomic action of dexamethasone on actin cytoskeleton. Mol Med 5:731–742
Thompson EA 1991 Insensitivity to the cytolytic effects of glucocorticoids in vivo is associated with a novel “slow death” phenotype. Cancer Res 51:5544–5550

Wassler MJ, Foote CI, Gelman IH, Shur BD 2001 Functional interaction between the SSeCKS scaffolding protein and the cytoplasmic domain of β1,4-galactosyltransferase. J Cell Sci 114:2291–2300
Regulation of a Distinctive Set of Genes in Glucocorticoid-evoked Apoptosis in CEM Human Lymphoid Cells

E. BRAD THOMPSON AND BETTY H. JOHNSON

Department of Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston, Texas 77555-0645

ABSTRACT

Gene expression was evaluated in clones of the acute lymphoblastic leukemic cell line CEM that were sensitive or resistant to apoptosis evoked by the glucocorticoid, dexamethasone (Dex). Founding clones CEM-C7 (glucocorticoid sensitive) and CEM-C1 (glucocorticoid resistant) were subcloned to maximize uniformity of each population studied. Among subclones of C1, our original pseudodiploid clone of glucocorticoid-resistant cells, we found a high proportion of hyperploid clones. Most C1 subclones were glucocorticoid resistant but two C1 subclones were found to be revertants to glucocorticoid sensitivity. Glucocorticoid receptor content of the C1 subclones varied almost 5-fold but higher quantity of receptors did not guarantee steroid sensitivity. Gene expression analysis was carried out on microchips containing representations for ~12,600 human genes. When a group of four subclones from C1 (three glucocorticoid-resistant and one glucocorticoid-sensitive revertant) were compared with the glucocorticoid-sensitive subclone CEM-C7-14 for basal gene expression, the four C1 subclones clustered closely and far from C7-14. Thus, basal gene expression in the C1 subclones differed for a large number of genes from that in the C7 subclone. Reversion to glucocorticoid sensitivity did not cause a major shift in basal gene expression to a more C7-like state. Three clones (one revertant glucocorticoid sensitive from C1 subclone, one C7 sensitive subclone, and one C1 glucocorticoid-resistant subclone) were compared for the genes regulated by treatment for 20 hours with 10^{-6} M Dex. This interval brings the cells to a point just before the onset of apoptosis. We tested the hypothesis that a distinctive set of genes would be regulated in the glucocorticoid-sensitive clones. This proved to be so. In three experiments, at our chosen levels of discrimination, 39 genes were consistently induced ≥2.5-fold and 21 genes were consistently reduced ≥2-fold in glucocorticoid-sensitive clones but not in the glucocorticoid-resistant clone. The glucocorticoid-resistant clone showed induction or reduction of 88 genes different from those regulated in the glucocorticoid-sensitive clones. These data support our hypothesis and further show that the glucocorticoid-resistant clone is capable of responding to steroid but with a different set of genes. We propose that a general metabolic switch accounts for the alteration.

I. Introduction

Glucocorticoids are a mainstay of therapy for many lymphomas and leukemias, due to the ability of such steroids to prevent the growth and to cause the apoptotic death of these malignant cells. While some effects of steroids may be
derived from their ability to block or lower the production of various growth factors upon which lymphoid cells depend, malignant lymphoid cells frequently have achieved independence from exogenously produced factors. These cells either make and use the factors themselves in intracrine, autocrine, or paracrine pathways or become independent of them altogether. Glucocorticoids can act directly on such malignant cells to bring about apoptosis. Evidence that this is so originally came from mouse and human malignant cell lines (Baxter et al., 1971; Norman and Thompson, 1977) and subsequently was supported by studies of freshly isolated human leukemic cells (Kaspers et al., 1994).

The complete mechanism by which steroids bring about apoptotic death of leukemic lymphoid cells is unknown but several basic facts have been established, largely through the use of cultured cells. The first basic requirement for glucocorticoid-evoked apoptosis is a functional glucocorticoid receptor (GR). Cells containing nonfunctional GRs and those with very low normal GR content are invariably resistant. GRs are transcription factors, activated by their ligands, the steroids. Synthetic steroids, such as dexamethasone (Dex), frequently are used therapeutically in place of the natural human glucocorticoid, cortisol. In the absence of ligand, the GR is located primarily in the cytosolic compartment, in association with a complex of several other proteins (Housley et al., 1985; Sanchez et al., 1987; Pratt and Toft, 1997; Silverstein et al., 1999). When the ligand binds to its high-affinity site in the carboxy-terminal portion of the GR, the receptor presumably reconfigures, dissociates from its heteromeric binding partners, and translocates to the nucleus. There, in association with a variety of other transcription factors, coactivators, or corepressors, the GR acts to regulate the transcription of specific sets of genes (Ivarie and O'Farrell, 1978; O'Farrell and Ivarie, 1979; Young et al., 1981; Harrigan et al., 1989; Horwitz et al., 1996; McKenna et al., 1999; Holter et al., 2001).

Though necessary, the GR alone is not sufficient for producing apoptosis. Many leukemic cells contain plentiful quantities of normal GRs but are nevertheless strongly resistant to glucocorticoid-evoked apoptosis. In principle, these cells could fail to die in the presence of agonist glucocorticoids for many reasons. The steroid ligands could be blocked from passage through the plasma membrane, pumped rapidly from the cells, and destroyed metabolically or biochemically conjugated so that their function is blocked. Though some of those mechanisms may occur occasionally (Johnson et al., 1984), they seem to be used rarely in leukemic systems. A second theoretical group of mechanisms includes resistant cells that have genetically or phenotypically altered the response systems to glucocorticoids so as to resist their lethal effect. Among these possibilities are 1) loss or critical reduction in quantity of one or more factors with which the GR must interact, 2) development of a dominant-negative form of such a factor, or 3) improper posttranslational modifications of the GR or an interactive factor (Gruol et al., 1986). A third list of mechanistic possibilities
involves changes that affect the general pathways for apoptosis: 1) alterations in
the balance of pro- and antiapoptotic members of the Bcl2 family of proteins;
2) loss of or inactivating mutations in caspases or other lethal proteases; and
3) changes in one or more critical protease substrates, rendering it/them resistant.
A final group of mechanisms could evoke alterations in specific genes’ abilities
to be regulated by ligand-driven GR: large-scale or selective heterochromatiza-
tion or methylation of DNA (Gasson and Bourgeois, 1983; Gasson et al., 1983)
or “insulation” of patches of genes could prevent the GR from properly affecting
their transcription (Bell and Felsenfeld, 1999).

A second known basic requirement for GR-evoked apoptosis of lymphoid
cells is cellular gene transcription and translation. In normal rodent thymocytes,
as well as rodent and human leukemic cell lines, blocking cellular transcription
or translation prevents the advent of the classic morphological and biochemical
events in the apoptotic pathway (Wyllie et al., 1981,1984; Van den Bogert et al.,
1988; Migliorati et al., 1994; Ramdas and Harmon, 1998; Mann et al., 2000).
Although, given enough time, such inhibitors are themselves lethal, the tight
correlations between the classic biochemical events preceding apoptosis and
outright eventual cell death lead one to conclude that cellular macromolecular
synthesis is necessary for glucocorticoid-evoked apoptosis. This is consistent
with the fact that the GR is a transcription factor and also with the timing of cell
kill. Unlike the rapid onset of apoptosis seen when the stimulatory agent simply
causes activation of a caspase cascade (Pinkoski and Green, 1999), glucocorti-
coid-evoked apoptosis is initiated only after a delay. In freshly isolated mouse
thymocytes, this delay is 1–2 hours, while in malignant human lymphoid lines,
the time is much longer, requiring 24 hours or more (Thompson, 1999). Only
after this delay interval do the markers accompanying apoptosis begin to be seen,
followed later by the breakdown of cellular macromolecular synthetic pathways
and outright cell death.

During the lag period, if the steroid is removed or replaced on the GR by an
antagonist ligand, apoptosis does not occur. In other words, the agonist steroid
must be continually present until the protease cascade is initiated and mitochon-
drial membrane disruption occurs with the consequent release of protease-
activating factors (Thompson, 1999).

As to the precise genes whose expressions are altered in such a way to cause
apoptosis, only a few landmarks have been identified. Several efforts to identify
the critical genes by older technologies designed to locate mRNAs that differed
in quantity after steroid treatment failed to identify the essential genes (Harrigan
et al., 1989,1991; Briehl et al., 1990; Baughman et al., 1991; Caron-Leslie et al.,
1991; Pease et al., 1992; Chapman et al., 1995; Cidlowski et al., 1996).
Consequently, the problem has remained unsolved and only recently has the
advent of gene microarray technique permitted a search for the detailed data
necessary to resolve the gene expression changes that lead to leukemic cell
apoptosis. We have chosen to address this issue using clones of cells established from the CEM line.

II. Clonal CEM Cell Lines as Models for Glucocorticoid-sensitive and Glucocorticoid-resistant Human Lymphoid Leukemia

Most early studies of glucocorticoid-induced lymphoid cell death were carried out in rodents, either treating the whole animal with steroid or isolating thymocytes for in vitro treatment. Later, it was shown that some malignant lymphoid cells from these laboratory animals could be killed by corticosteroids (Baxter et al., 1971). Based on this knowledge, lines of cells such as S49 (Sibley and Tomkins, 1974; Sibley and Yamamoto, 1979), WEHI7 (Bourgeois and Newby, 1977), and P1798 (Thompson, 1991) were developed. Each of these and several other cell lines have proven very useful in studying glucocorticoid action.

It has long been recognized that glucocorticoids were effective as chemotherapeutic agents against several types of human leukemias and lymphomas but not against others. It also became obvious during the era of single-drug chemotherapy for leukemia that resistance often developed after treatment of the sensitive malignancies with glucocorticoids alone. How well the rodent models explained these observations in humans was dubious. In the early 1970s, the first few human cell lines of T-cell origin were established. Since, in general, primitive T cells (i.e., young thymocytes not yet differentiated so as to be only CD4 or CD8 surface antigen positive) were likely to be sensitive to glucocorticoid-evoked apoptosis, we screened the leukemic cell lines of T-cell derivation then available for sensitivity to cell death when exposed to Dex. This synthetic glucocorticoid agonist was chosen because of its high potency, use in therapy, and resistance to metabolic inactivation. Among the lines tested, we noted that CEM, derived from a 3-year-old female patient with late-stage acute lymphoblastic leukemia, was somewhat sensitive to Dex (Foley et al., 1965). Reasoning that the partial effect was likely due to the presence of a mixture of sensitive and resistant subpopulations, we immediately cloned the line and screened a number of clones for Dex sensitivity. As anticipated, these showed pure sensitivity or resistance. Two were chosen for detailed study: CEM-C7 as the prototypical sensitive clone, killed by GR-occupying concentrations of Dex or other agonist glucocorticoids, and clone CEM-C1 as the prototypical resistant cells, able to grow in up to $10^{-5}$ M Dex, a concentration 100-fold greater than that required to fully saturate the GR (Norman and Thompson, 1977).

Many properties of the two clones were similar. Each was pseudodiploid, with a characteristic pericentric inversion of chromosome 9 and an extra chromosome 22 (Moore et al., 1985). They contained similar concentrations of GR (12,000–14,000 GR sites/cell, $K_d \approx 20$ nM Dex), showed the same weak ability to metabolize Dex, and both demonstrated induction of glutamine syn-
The latter result suggested that the resistance of C1 cells to Dex-evoked apoptosis did not represent a global inability to respond to glucocorticoid but rather to some change that blocked the pathway to apoptosis. Both clones later were shown to be haploid for the normal GRα gene, with one normal and one mutant allele for the GRα gene. In both, the mutant allele contained a single amino acid substitution (Leu753Phe) in the ligand-binding domain (Ashraf and Thompson, 1993; Hillmann et al., 2000). This renders the GR produced from that allele incapable of retaining bound Dex under conditions that alter the receptor to its transcriptionally active form. Eventually, it was found that the Leu753Phe mutation had existed in the cells of the patient and was not a consequence of tissue culture conditions (Hillmann et al., 2000). Somatic cell hybrids were prepared between C1 cells and a completely Dex-resistant subclone of C7 (ICR-27) that had lost essentially all functional GR (GR−). The hybrids in mass culture and after cloning all were sensitive to Dex-evoked apoptosis (Yuh and Thompson, 1987). Restoring normal GR to the GR− subclone by transfection caused it to regain Dex sensitivity; thus the GR was both necessary and sufficient to restore the sensitive phenotype (Harbour et al., 1990). From the fact that the C1 × ICR-27 hybrid cells underwent apoptosis when exposed to Dex, we concluded that 1) the C1 cells supplied sufficient active GR to deliver the apoptotic signal from Dex in the hybrids and 2) C1 cells do not contain a dominant-negative function that blocks the apoptotic response. A later clue to the nature of the lesion in C1 cells came when we discovered that treating them with forskolin (to activate the cyclic AMP (cAMP) signaling pathway) restored Dex sensitivity (Medh et al., 1998). We believe that clones of the CEM line represent a useful model system in which to study the mechanism for glucocorticoid-induced leukemic lymphoid cell apoptosis.

The virtues and limitations of tissue-culture models for in vivo cellular behavior are well known but a few points bear repeating. By using a tissue-culture system, we are able to study in detail the responses of a defined population of human leukemic cells to steroids and other agents. With this system, we can carry out experiments difficult or impossible to conduct in vivo. We can isolate the direct effects of the steroids on the leukemic cells, without the many complicating influences of secondary effects due to the actions of the agents on other cells and tissues. The use of clonal populations increases the uniformity of response and permits powerful genetic experiments. Obviously, the specifics of response uncovered in a particular clone of cells must be compared with other cells and in vivo systems before they can be considered to be generally applicable. Limits notwithstanding, the use of cultured cells — or better yet, clones of cultured cells — has provided enormously useful information about biochemical mechanisms. This has been true for systems from bacteria through yeast and higher eukaryotes.
III. Basic Responses of the Sensitive Clone CEM-C7 to Glucocorticoid

When Dex (or another glucocorticoid agonist) is constantly present in the culture medium, the population-doubling time of the cells is maintained for at least 24 hours but, over several days afterward, increasing numbers of cells arrest in the G1/G0 phase of the cell cycle. Concomitantly, there is an increasing loss of clonogenicity (Harmon et al., 1979) and accumulative apoptosis. Time-lapse photomicrography of Dex-sensitive cells has shown that, following the 24-hour lag period, increasing numbers of cells change from their usual spheroidal shape to a collapsed, shrunken form. This appears to occur stochastically. For each cell, once the apoptotic process begins, the alteration events occur rapidly, beginning with violent membrane bubbling followed by shrinkage and collapse without immediate lysis (E.B. Thompson and B.H. Johnson, unpublished results). Over a longer period of hours, the collapsed cells lyse.

Biochemically, several major steps in the pre-apoptotic and apoptotic periods have been identified and are summarized in Figure 1. The earliest change in gene expression we have documented is a reduction in c-myc mRNA and protein, due to an inhibition of c-myc transcription (Zhou et al., 2000a). The reduction of mRNA levels begins as early as 1 hour after addition of Dex (Thulasi et al., 1993). Since the half-life of c-myc mRNA is \( \approx 30 \) minutes, the block of c-myc transcription must occur shortly after addition of Dex. The reductions are seen only in sensitive cells; resistant clones, whether or not they contain an active form of the GR, maintain normal cMyc levels in the presence of Dex. Interest-

![FIG. 1. Timing of some major events in CEM-C7 cells after addition of dexamethasone (Dex).](image-url)
ingly, when forskolin is used to stimulate the cAMP pathway in the GR$^+$ resistant clone C1, addition of Dex kills the cells and cMyc is suppressed (Medh et al., 1998). Constitutive expression of ectopic c-myc delays glucocorticoid-evoked apoptosis in C7 cells (Medh et al., 2001). The importance of c-myc downregulation in the death of lymphoid cells has been reviewed (Thompson, 1998).

Following in time, c-jun becomes chronically induced (Zhou et al., 2000b) and ornithine decarboxylase transcription is reduced (Miller et al., 2002). Eversion of phosphatidyl serine in the mitochondrial membrane begins $\approx 20$ hours after addition of Dex, while the first indications of increased caspase activity occur only after $\approx 24$ hours. DNA scission into very large fragments starts first after about 24 hours of Dex treatment and later after about 36 hours into $\approx 200$ base-pair units (Johnson et al., 1997). Both caspase activity and DNA lysis start slowly, then increase exponentially, consistent with the recruitment of increasing numbers of shrunken, apoptotic cells, noted microscopically. These cells eventually go on to lyse, leaving debris and some bits of DNA still contained within membranes (karyorrhexis).

IV. Quantity of GR Does Not Explain the Resistance of CEM-C1 Subclones to Glucocorticoid

When the apoptosis-resistant clone C1 was originally established, careful comparison with the apoptosis-sensitive clone C7 indicated that, quantitatively and qualitatively, the GR content of the two was equivalent (Zawydiwski et al., 1983). Together with the somatic cell hybridization studies mentioned earlier, it was clear that lack of GR per se in C1 cells was not the explanation of their lack of apoptosis. Yet, the apoptosis-inducing function of the steroid-activated GR in C1 cells could be restored by forskolin treatment of the cells. Thus, it seemed that some interactive pathway was lacking. One report, using C1 cells presumably originally supplied from our laboratory, found that the population studied contained far fewer GR sites than did C7 cells. Transfection of these C1 cells with a plasmid designed to express normal GR both raised cellular GR levels and rendered several subclones apoptosis sensitive to Dex (Geley et al., 1996).

In considering this question, we suspected that the C1 cells employed in the study had undergone phenotypic drift and, as a consequence, had acquired a predominant population with low GR content. To test this possibility, we recloned our C1 cell population. Examination of the subclones showed a variety of karyotypes and GR phenotypes. The diversity of subclones we have subsequently uncovered within the C1 population would suggest that the cells in the above-cited study represented an overgrowth of subpopulations with the phenotype described. Some of these are listed in Table I. We noted a strong tendency for the C1 cells to develop hyperploidy, a phenomenon often seen in lymphoid malignancies
In chromosome spreads of the uncloned C1 population, many hyperploid metaphases were noted. Among the clones we chose at random for karyotyping, 75% were subtetraploid. This is thought to arise because of nondisjunction at metaphase. Although it has been reported that, in general, hyperploidy correlates clinically with more treatment-sensitive behavior, ploidy in these clones did not generally correlate with resistance/sensitivity to glucocorticoids.

Low quantity of GR, as measured by specific, high-affinity cellular binding sites for Dex, did not consistently explain the resistance of C1 subclones to steroid-evoked apoptosis. GR quantity in the C1 clones totally resistant to 1 μM Dex varied widely from 7000 to 33,500 GR sites/cell (Table I). Thus, the latter clone contained three times more GR than the sensitive clone C7-14, without showing Dex sensitivity.

Conceivably, the low-GR clones (e.g., C1-11, with ≈ 7000 GR sites and relatively short doubling times) could overgrow the population under some conditions. Two subclones, C1-6 and C1-8, were found to have reverted to high sensitivity to Dex-evoked apoptosis. Both clones were hyperploid and contained higher levels of GR than the prototypical, pseudodiploid sensitive C7 clones. However, the fact that other hyperploid clones (e.g., C1-16) contained as much or more GR than clones C1-16 and C1-8, yet continued to be resistant, suggested

| TABLE I | Physiochemical Characteristics of CEM-C1 Clones |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| CEM-C1 (1982)   | Res             | 10,000 ± 700    | 9 ± 5.0         | 25.7 ± 1.8      | 48              |
| Subclones^5     |                 |                 |                 |                 |
| C1-15           | Res             | 9,900 ± 2,600   | 12 ± 3.2        | 21.9 ± 1.5      | 0               |
| C1-11           | Res             | 6,800 ± 1,700   | 7 ± 0.2         | 21.4 ± 0.6      | 0               |
| C1-4            | Res             | 12,250 ± 2,800  | 8 ± 2.3         | 25.4 ± 2.0      | 84              |
| C1-16           | Res             | 33,600 ± 7,600  | 9 ± 0.4         | 21.9 ± 1.5      | 72              |
| C1-6            | Sens            | 27,200 ± 3,200  | 7 ± 1.7         | 25.4 ± 2.9      | 84              |
| C1-8            | Sens            | 27,900 ± 1,000  | 7 ± 0.2         | 23.8 ± 2.7      | 72              |

[^1]Resistant or sensitive to apoptosis in 10^{-6} M Dex. 2Average number of Dex binding sites/cell and Kd ± standard deviation (SD) from two to four multipoint Scatchard analyses. 3Average log-phase population-doubling time in hours ± SD, n = 3. 4Average percent hyperploidy in 25 chromosome spreads on each of three separate slides. 5Subclones were obtained by the soft agar method (Harmon et al., 1985).]
that quantity of GR alone was not the explanation for the reversions. Gene-array analysis data (below) confirmed this suggestion.

The overall picture presented by the clonal CEM cell system, therefore, is that by activating the GR, Dex starts a process in which the ligand-activated receptor must participate continually for some 20–30 hours. This process culminates in activation of caspases (and possibly other proteases), with the consequence of relatively rapid and irreversible apoptosis. Resistance to this process cannot — in most cases — be explained by simple lack of GR. Consequently, we framed the following hypothesis: During the lag before overt apoptosis, an interactive network of genes undergoes altered expression. Some are controlled directly by the activated GR, others, indirectly, as a result of the changes in primary GR targets.

V. Gene-array Analysis of Glucocorticoid Effects in Clones of Human Leukemic Cells

We have begun to test the above hypothesis, employing several of these closely related clones of CEM cells.

A. BASAL GENE EXPRESSION IN CEM-C1 SUBCLONES

Gene-expression analysis was carried out by use of Affymetrix HG.U95Av2 chips, which carry sets of oligonucleotides capable of identifying ~12,600 human genes for which full-length cDNAs have been reported. The identity and functions of many of these genes are known. RNA samples were prepared from four of the C1 subclones, during midlog growth in RPMI 1640 medium supplemented with 5% heat-inactivated fetal calf serum. Approximately half of the 12,600 genes were found to be expressed in the basal state in each of the clones. Cluster analysis showed that while the expressed gene sets of the four resistant clones generally were similar and quite different from sensitive clone C7-14, certain differences in expressed gene levels divided the four clones into two pairs (Figure 2). Clones C1-6 and C1-16 clustered together, as did the other pair, C1-15 and C1-12. C1-6 is Dex sensitive and C1-16 is resistant; therefore, their relative clustering in the basal state does not correlate with either phenotype. One obvious correlation that could be behind this clustering is that C1-6 and C1-16 are hypotetraploid clones, whereas the other clones are both pseudodiploid (Table I).

B. COMPARISON OF APOPTOSIS-SENSITIVE WITH APOPTOSIS-RESISTANT CLONES

Three of the freshly isolated clones were selected for closer analysis (Figure 3). C7-14 is a subclone of CEM-C7 and, like its parent, contains 11,000 GR
sites/cell and is sensitive to glucocorticoid-evoked apoptosis. C1-15 and C1-6 are subclones of C1. Clone C1-15 closely resembles the original properties of C1 in karyotype, strong Dex resistance, and GR content (statistically not different from C7-14). Clone C1-6 is interesting because it represents a rare revertant to sensitivity towards Dex-evoked apoptosis. In routine culture over its years of use, clone CEM-C1 never has been overgrown by a sensitive population and, among the clones chosen for study, only C1-6 and C1-8 showed high sensitivity.

Based on our original hypothesis as to the mechanism by which glucocorticoids cause apoptosis, we posed a subsidiary hypothesis, to be tested by gene-array analysis: A distinctive network of genes is regulated by glucocorticoids in the apoptosis-sensitive clones. These genes will not be regulated in the glucocorticoid-apoptosis-resistant cells. However, the functional GR of the cells resistant to glucocorticoid-evoked apoptosis still will be capable of regulating expression of some genes.

A time point of 20 hours following addition of Dex to the cultures was chosen, to be near the end of the lag period. In preliminary experiments, we confirmed that, after 20 hours in 10^{-7} M Dex, the sensitive C7-14 and C1-6 cells had not entered apoptosis (i.e., they still excluded propidium iodide vital dye). Between 24 and 48 hours, they showed increasing apoptosis, including overt cell death, just as the original CEM-C7 cells behaved. Clone C1-15 cells were
completely Dex resistant in that they continued to grow at the same rate and showed no apoptosis to 1 \( \mu \)M Dex over a period of 4 days.

For gene-array studies, the same protocol was carried out on three occasions over the course of a year. To cells in midlog growth, Dex was added to a final concentration of \( 10^{-6} \) M. Control cultures received an equal volume of the ethanol vehicle used to dissolve the Dex. Twenty hours later, the cells were collected and their RNA extracted. This late time point in the preapoptotic, reversible phase of Dex treatment was chosen to emphasize the accumulated changes in gene expression just preceding onset of full apoptosis. By repeating the experiment three times over a long period, we could test the constancy of the observed changes in gene expression and gain some statistical confidence of their relevance. Time-matched controls were included in each experiment. This approach is much safer than experiments involving a single sample at one time point and controls at only one time. The RNA was provided to the University of Texas Medical Branch Genomics Core Facility, where it was processed and analyzed by a single individual, using the methods prescribed by Affymetrix. The basic validity of the results was verified in several ways. We noted that, in the array results, as we and others had shown earlier by Northern blot analysis, c-myc mRNA consistently was reduced and GR\( \alpha \) was induced. Second, we selected three genes that, according to the array data, were induced in the sensitive clones. Northern blots for those three mRNAs in cellular RNA extracts confirmed their clear induction (Table II). Third, Affymetrix chips often contain multiple probes
for the same gene at various locations on the chip. Several of the genes regulated in both sensitive clones appeared twice on our chips and, at both sites, gave comparable data and additional validation of the results. Induced genes were transforming growth factor beta (TGFβ) receptor 2, interleukin (IL)-7 receptor, α tubulin, absent in melanoma 1 (AIM 1), inositol polyphosphate-1-phosphatase (INPP 1), and an unknown protein KIAA0878. The one repressed gene was c-myc.

For a first level of data analysis, we chose relatively stringent criteria for inclusion of a gene in the category “regulated by Dex.” These criteria were greater than a 2.5-fold increase or greater than a 2-fold decrease in mRNA level in at least two of the three experiments. Validation studies in our Genomics Core Facility had shown that when a single RNA preparation was evaluated on three independent chips, less than 0.2% of the genes differed in expression by more than 2-fold. The higher limit of 2.5-fold was chosen to give an additional level of confidence that the changes seen were not due to random causes. On the other hand, we chose 2-fold reduction of mRNA level as the downward limit because, unlike fold induction, which theoretically can be infinite, fold reductions in mRNA level asymptotically approach zero. A 2-fold reduction in mRNA represents a 75% reduction in the pool for that mRNA, a quantitative reduction likely to have functional significance.

When these criteria were applied, 39 genes were found to be induced greater than 2.5-fold in at least two of the three experiments in both C7-14 and C1-6 cells and were never or only once induced in C1-15 cells. A set of 21 genes distinctive to the two apoptosis-sensitive clones was also found to be repressed by 2-fold (Table III). In these two gene sets, 22 genes were induced and four were repressed exclusively in both sensitive clones in all three experiments. When we examined the signal intensities for each of the genes that, in the three experiments, twice exceeded the chosen limits but once showed less than the cutoff regulation, the great majority in that experiment showed regulation in the correct direction, though quantitatively less than our chosen limits.

Examination of Table III also shows that other sets of genes were induced exclusively in clone C1-6 or C7-14, while a few others were induced in one of

| TABLE II |
| Fold Induction by Dex | |
|----------------------|------------------|------------------|------------------|
| Gene                | GenBank #        | C7-14            | C1-6             | C1-15            |
| BTG-1 (X61123)      | 7.4              | 8.5              | 0.3              |
| DSCR1 (U85267)      | 8.4              | 12.2             | 0.8              |
| NFIL3, E4BP4 (X64318) | 21.1            | 7.9              | 0.4              |

[1] By image analysis of Northern blot.]
these clones as well as the apoptosis-resistant clone C1-15. Interestingly, clone C1-15 repeatedly showed unique induction of 62 genes. Similar patterns occurred for downregulated genes, though the numbers were smaller and, in several cases, zero. These data confirm our hypothesis that a distinctive set of genes is regulated in cells that show Dex-evoked apoptosis. Those that are found to be regulated in both sensitive clones are strong candidates for important players in the preapoptotic sequence. The specific sets of genes identified here are unlikely to include all those that are essential to glucocorticoid-evoked apoptosis. The chosen limits were deliberately made rather stringent and it is very likely that other genes, whose expression is altered to a lesser degree, will prove important to the process. For example, two genes that we know to be Dex regulated in CEM-C7 cells, c-jun and odc, fell outside our chosen limits.

Application of other statistical screening procedures (e.g., analysis of variance, ANOVA) can result in larger gene sets. The use of each particular statistical method has its own merits and limitations. For example, ANOVA evaluation of all three clones with a cutoff of 99% probability that any given gene was altered significantly differently from the rest identified some 350 genes regulated by Dex distinctive to the apoptosis-sensitive cells. ANOVA assumes that the populations considered follow a gaussian distribution. The fold-induction data actually do not do so. Thus, some of the genes accepted as induced (or deinduced) by ANOVA may be random variations. However, the conclusion that a unique set of genes is altered in the apoptosis-responsive clones is not changed. It should be recalled that the HG.U95Av2 gene chip only contains about one third of the genes expressed in human cells; hence, there are undoubtedly additional genes to be discovered that are relevant to this system. Unfortunately, many of the genes not found on this chip are only expressed sequence tags (ESTs), so much work will

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>Distinctive Gene Sets That Are Dexamethasone Regulated in Clones of Cells Sensitive or Resistant to Glucocorticoid-induced Apoptosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single clone only</td>
<td>Two clones</td>
</tr>
<tr>
<td>C1-6</td>
<td>C7-14</td>
</tr>
<tr>
<td>A</td>
<td>71</td>
</tr>
<tr>
<td>B</td>
<td>124</td>
</tr>
</tbody>
</table>

[A = number of mRNAs induced > 2.5 fold in ≥ two of three experiments. B = number of mRNAs repressed > 2.0 fold in ≥ two of three experiments. Each column gives the number of genes unique to that column (i.e., the 71 genes induced in C1-6 are uniquely so in C1-6, not in any other clones or combination).]
### TABLE IV

**Genes Induced or Deinduced by Dexamethasone in Both Apoptotic-sensitive Clones**

<table>
<thead>
<tr>
<th>Induced gene</th>
<th>Average-fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR4, CC-CCR-4, K5-5, CMKBR4, CKR4, ChemR13, C-C chemokine receptor-4</td>
<td>465.2</td>
</tr>
<tr>
<td>BIMEL, BCL2L11, BOD, BIM, BIML, BAM</td>
<td>224.0</td>
</tr>
<tr>
<td>KIAA0620 protein</td>
<td>215.0</td>
</tr>
<tr>
<td>Protein kinase H31</td>
<td>214.3</td>
</tr>
<tr>
<td>TUBA1, FLJ30169, alpha tubulin 1</td>
<td>112.8</td>
</tr>
<tr>
<td>TIP3, JAK, JAB, SOCS1, SOCS-1</td>
<td>34.4</td>
</tr>
<tr>
<td>BTG1, B-cell translocation gene 1</td>
<td>20.7</td>
</tr>
<tr>
<td>IL7R, CD127, interleukin-7 receptor</td>
<td>12.1</td>
</tr>
<tr>
<td>AP1G2, G2AD, gamma2-adaptin</td>
<td>11.4</td>
</tr>
<tr>
<td>SLC18A2, SVM1, VAT2, VMAT2, synaptic vesicle amine transporter</td>
<td>7.4</td>
</tr>
<tr>
<td>Hypothetical protein MGC17330, clone DJ515N1</td>
<td>7.4</td>
</tr>
<tr>
<td>DKFZP586A0522 protein</td>
<td>7.2</td>
</tr>
<tr>
<td>BIRC3, HAIPI, MIHC, HAIPI, API2, CIAP2, inhibitor of apoptosis protein 1</td>
<td>7.2</td>
</tr>
<tr>
<td>CD69, CD69 antigen</td>
<td>7.1</td>
</tr>
<tr>
<td>DSCR1, MCIP1, ADAPT78, down syndrome candidate region 1</td>
<td>7.1</td>
</tr>
<tr>
<td>NFIL3A, NF-IL3A, IL3BP1, E4BP4, nuclear factor interleukin 3</td>
<td>6.8</td>
</tr>
<tr>
<td>seb4B, RNP1, DJ80021.2, HSRNASEB, SEB4D</td>
<td>5.8</td>
</tr>
<tr>
<td>JAM3, FLJ14529</td>
<td>5.7</td>
</tr>
<tr>
<td>DFNA5, ICERE-1, deafness autosomal dominant 5</td>
<td>5.5</td>
</tr>
<tr>
<td>KIAA0386, C6orf32, DIFF40, DIFF48, PL48</td>
<td>4.8</td>
</tr>
<tr>
<td>KIAA0878 protein</td>
<td>4.5</td>
</tr>
<tr>
<td>GR alpha, NR3C1, GRL, GR, GCR, glucocorticoid receptor</td>
<td>4.2</td>
</tr>
<tr>
<td>TGFB1, HNPPC6, transforming growth factor beta receptor II</td>
<td>4.2</td>
</tr>
<tr>
<td>SLIM1, SLIMMER, FHL1, KYO-T</td>
<td>4.1</td>
</tr>
<tr>
<td>AIM1, absent in melanoma 1</td>
<td>3.9</td>
</tr>
<tr>
<td>PTK2B, PTK, FAK2, PKB, CADTK, CASK, PYK2, RAFTK, cell adhesion kinase beta</td>
<td>3.5</td>
</tr>
<tr>
<td>TXNIP, VDUP1, THIF, EST01027, HHCPA78</td>
<td>3.4</td>
</tr>
<tr>
<td>WFS1, WFRS, DFNA38, DFNA14, WFS, DFNA6, DIDMOAD, Wolfram syndrome 1</td>
<td>3.4</td>
</tr>
<tr>
<td>INPP1, inositol polyphosphate 1-phosphatase</td>
<td>3.4</td>
</tr>
<tr>
<td>MAD-3, NFKBI, IKBA, NFKBIA</td>
<td>3.1</td>
</tr>
<tr>
<td>KIAA0540 protein</td>
<td>3.0</td>
</tr>
<tr>
<td>KIAA0920, paralemmin 2, A kinase (PRKA) anchor protein 2, AKAP-KL</td>
<td>3.0</td>
</tr>
<tr>
<td>Induced gene</td>
<td>Average-fold change</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>CD53 glycoprotein, MOX44</td>
<td>2.9</td>
</tr>
<tr>
<td>UQCRFS1, GRAP2</td>
<td>2.7</td>
</tr>
<tr>
<td>MAP1A, MAP1L, microtubule-associated protein 1A</td>
<td>2.7</td>
</tr>
<tr>
<td>PRG1, PRG, hematopoetic proteoglycan core protein</td>
<td>2.7</td>
</tr>
<tr>
<td>BTG2, TIS21, PC3</td>
<td>2.6</td>
</tr>
<tr>
<td>PTPRM, HR-PTPU, PTPRL1, RPTPM, protein tyrosine phosphatase receptor type M</td>
<td>2.6</td>
</tr>
<tr>
<td>SLA1, SLAP, SLA, Src-like adapter protein</td>
<td>2.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deinduced gene</th>
<th>Average-fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD17 (S. pombe) homolog</td>
<td>392.3</td>
</tr>
<tr>
<td>RAG1, recombination activating protein</td>
<td>7.2</td>
</tr>
<tr>
<td>TLE4, transducin-like enhancer of split 4, homolog of Drosophila E(sp 1)</td>
<td>6.7</td>
</tr>
<tr>
<td>GPAT, PPAT, glutamine PRPP amidotransferase, phosphoribosyl pyrophosphate</td>
<td>4.5</td>
</tr>
<tr>
<td>amidotransferase</td>
<td></td>
</tr>
<tr>
<td>Hypothetical protein MGC2574</td>
<td>4.0</td>
</tr>
<tr>
<td>HRY, hairy (Drosophila) A85 homolog</td>
<td>3.2</td>
</tr>
<tr>
<td>HMGCS1, 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1</td>
<td>3.2</td>
</tr>
<tr>
<td>MYC, c-myc</td>
<td>3.0</td>
</tr>
<tr>
<td>KIAA0830 protein</td>
<td>2.7</td>
</tr>
<tr>
<td>DKFZP564C186 protein</td>
<td>2.6</td>
</tr>
<tr>
<td>ITGA4, integrin alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)</td>
<td>2.6</td>
</tr>
<tr>
<td>Protein predicted by clone 23733, HSU79274</td>
<td>2.5</td>
</tr>
<tr>
<td>Hypothetical protein HSPC111</td>
<td>2.4</td>
</tr>
<tr>
<td>LTB, lymphotixin beta, TNF superfamily member 3</td>
<td>2.4</td>
</tr>
<tr>
<td>MEGF2, CELSR3, cadherin, EGF LAG seven-pass G-type receptor 3, flamingo</td>
<td>2.3</td>
</tr>
<tr>
<td>(Drosophila) homolog</td>
<td></td>
</tr>
<tr>
<td>DKFZp5860720</td>
<td>2.2</td>
</tr>
<tr>
<td>PPP3CA, calcineurin A1, protein phosphatase 3 (formerly 2B) catalytic subunit-</td>
<td>2.1</td>
</tr>
<tr>
<td>alpha isoform (calcineurin A alpha)</td>
<td></td>
</tr>
<tr>
<td>Hypothetical protein FLJ10534</td>
<td>2.1</td>
</tr>
<tr>
<td>PRMT3, protein arginine N-methyltransferase 3</td>
<td>2.1</td>
</tr>
<tr>
<td>DNA from chromosome 19-specific cosmid R30923, similar to RIKEN cDNA</td>
<td>2.1</td>
</tr>
<tr>
<td>2410153K17 gene</td>
<td></td>
</tr>
<tr>
<td>DTK, TYRO3, protein tyrosine kinase</td>
<td>2.0</td>
</tr>
</tbody>
</table>
be necessary to identify their full sequences, identities, and functions. Nonetheless, our initial, limited gene set already shows genes with some interesting properties, worthy of investigation for their relevance to the initiation of apoptosis.

Table III shows another interesting contrast between the two apoptosis-sensitive clones and C1-15, the resistant clone. Clone C1-15 also has a distinctive set of genes that are uniquely and consistently regulated by Dex. These data clearly indicate that there is a functional GR in C1-15 cells. One possible mechanism to explain the resistance of C1-15 cells is that a gene strongly protective against apoptosis is induced or that an antiapoptotic gene is deinduced. A search through the expressed gene arrays for the obvious candidates in these categories has not revealed any clear candidates. Thus, C1-15 cells contain an active GR that regulates many genes but these genes differ for the most part from those regulated by the GR in the clones that eventually undergo apoptosis after the addition of the GR agonist ligand, Dex.

Most of the regulated genes unique to clones C1-6 and C7-14 have been identified. Among these, several have shown antigrowth, proapoptotic, or pro-differentiation activities. Another group of genes was involved in various signal transduction or cell regulatory functions. The remainder fell into various groups, including a few structural genes. Few, if any, genes of general cell metabolism turned up in the set identified. Quite a few have received multiple working names or designations. These genes are listed by some of their names in Table IV.

These data demonstrate that the ligand-activated GR acts as both an inducer and a repressor of gene transcription. At least some of the regulated genes are controlled at the primary level by the GR. For example, inhibition of c-myc transcription is a primary effect of the activated GR, requiring no additional macromolecular synthesis (Medh et al., 1998; Zhou et al., 2000a). Not all of the genes need be regulated directly by the GR, however. The observed regulation of some genes is likely to be the consequence of GR-driven alterations in mRNA and protein levels of macromolecules that themselves serve regulatory functions. For example, we have shown that the GR-driven reduction in cMyc leads to a later reduction in ornithine decarboxylase, whose promoter contains a cMyc-binding site. At that site, cMyc is a potent inducer of transcription. Thus, the loss of cMyc due to GR inhibition of c-myc transcription eventually results in a reduction of ornithine decarboxylase mRNA and protein. For other genes, GR-driven loss of a repressor of transcription could result in greater transcription of the gene secondarily regulated. In analogous fashion, GR-driven inductions could, by secondary or later sequential effects, produce either induction or repression of genes not under primary GR control. The GR itself could participate in these derivative regulations, for genes where the GR acts only in a heteromeric complex with other transcription factors.
Hypothesis for the underlying mechanism of resistance or sensitivity to glucocorticoid-evoked apoptosis: We propose that the critical difference between the sensitive and resistant clones is in the nature of a switch, one that shifts the sets of responsive genes from one group to another. One group of regulated genes culminates in turning on apoptosis, while the other does not. The nature of the proposed switch is unknown; it could be genetic, epigenetic, or purely regulatory.

VI. Comparisons with Earlier Studies of Related Systems

In a previous attempt to identify genes relevant to glucocorticoid-evoked apoptosis in CEM cells, the effect of Dex on growth-arrested compared to growing cells was studied (Tonko et al., 2001). A growing CEM-C7 subclone (CCRF-CEM-C7H2) was exposed to $10^{-7}$ M Dex and single samples were taken for analysis at 0, 2, and 8 hours. Against this were compared results from CCRF-CEM-C7H2-GE2-p16 cells treated with doxycycline to cause growth arrest. These cells had been prepared from the H2 clone by expressing in them the cyclin-dependent kinase inhibitor p16 (INK4A) gene, causing G0/G1 arrest. This had been found to heighten their sensitivity to Dex-evoked apoptosis (Ausserlechner et al., 2001). The primary array analysis was carried out by a commercial concern and the data subsequently analyzed by the authors. Despite its many limitations (e.g., one set of data at each time point, no time-matched controls, no evaluation of the effects on gene expression of doxycycline itself or of the introduction of the INK4A gene, and no comparison of the same cells (i.e. the INK4A-expressing cells, treated by Dex when growing and growth arrested)), the paper does show that a limited number of genes are altered by Dex treatment. The authors found that only eight genes were altered beyond the chosen limits, in both the growth-arrested and growing cells, at the times chosen for comparison. It is not made clear why one should expect the same genes to be altered in the growing cells many hours before their growth arrest as those in previously growth-arrested cells. We found none of these eight genes to be altered in our clones under our experimental conditions.

A second study from the same laboratory compared the effects of Dex treatment on CEM-C7 cells with those on another human leukemic line, Jurkat (Obexer et al., 2001). The Jurkat cells were of two sorts, one (clone A11-1) transfected so as to stably express a high level of wild-type (WT) rat GR, the other transfected with the GR-LS7 mutant gene (clone F6-1). The mutant GR (Miesfeld et al., 1987; Helmberg et al., 1995) was about half as potent overall in transactivating genes as WT GR. Both the WT and the mutant GR repressed genes overall to about the same extent. An early-version Affymetrix chip Hu6800/HuGeneFL containing probes for about 5600 genes was employed. Samples were taken at 0, 3, and 8 hours after addition of Dex. Generous criteria were chosen for accepting an expression change as valid, making it difficult to be
sure whether some of the changes seen are due to random variations. This problem is compounded by analyzing only one experiment per time point. Nevertheless, several genes were found to be regulated similarly at one or more time points in the CEM and Jurkat cells.

Our data were obtained at a much later time, close to when apoptosis begins in CEM cells. We sought to determine the culminating gene changes that bring on that crisis. We employed Affymetrix chips that contain probes for 12,600 genes, about twice the number found on the chips in the earlier studies. We also carried out three independent experiments in order to have greater confidence that the changes observed were not due to random events. Indeed, we observed that many genes appeared to be altered in expression in only one of the three experiments. This was true for both the apoptosis-resistant and apoptosis-sensitive clones.

When we compared the list of genes consistently altered after 20 hours of Dex in our CEM clone set with those found at the earlier time points in other CEM or Jurkat cells examined by Tonko et al. (2001) and Obexer et al. (2001), we found some concordances. A few genes observed to be altered by ≥ 2-fold at 3 or 8 hours were also in our 20-hour set. Others found in our set had shown a lesser degree of altered expression in the earlier data from the other CEM-C7 subclones or in the Jurkat cells. The set of genes we found to be distinctively altered in the late preapoptotic stage did not provide much support for the theory of metabolic pathway alterations proposed earlier (Tonko et al., 2001).

Concordance between certain of the genes we detected and those found earlier in single-datapoint experiments provides added confidence that they may be universally important in lymphoid leukemic cell apoptosis. Several possibilities exist regarding the genes that do not match in these “discovery” experiments. The most exciting is based on our hypothesis that the steroid evokes a sequential network of genes: genes induced or repressed directly by activated GR, genes altered as a consequence of those direct effects, and reverberations stemming from these changes, including, for example, genes whose expression is altered as a result of posttranslational effects caused by the protein products of initial gene changes. An example might be an induced protein kinase whose function is to activate or deactivate by phosphorylating some gene-specific transcription factor; another would be proteins that reconfigure chromatin. Many other examples, of course, can be listed. To evaluate this time-dependent sequential gene network theory properly, we are undertaking a full kinetic analysis of our clone set.

Other more-trivial reasons may explain the lack of concordance of some genes in the data from the several experiments in two different laboratories. Some of the genes that seem to be altered in single experiments may not be due to the steroid but only be random variations in expression. Some may be clone specific and not universally important for glucocorticoid-evoked apoptosis. In
some cases, the methods employed may have lacked sufficient sensitivity. Since our later experiments used chips containing twice as many genes, we may have discovered some responding genes not represented on the earlier chips. After all these explanations are considered, the fact remains that the concordance of a number of gene expression alterations due to Dex in several similar clones of all cells originally grown from two different patients is consistent with our sequential gene network hypothesis.

VII. Summary

We have applied microchip gene-array analysis to the problems of phenotypic variability in cultured cells and glucocorticoid-induced apoptosis of lymphoid cells. The system used was closely related clones of human acute lymphoblastic leukemia cells, all derived from the CEM cell line. To study the problem of phenotypic variability, so common in long-term cell culture, we compared subclones derived from our two original CEM clones: CEM-C7 that undergoes apoptosis in glucocorticoid and CEM-C1 that is resistant. The subclones generally maintained the gross initial parental phenotype but showed frequent conversion to polyploidy. Gene arrays showed that the basal gene expression set of the subclones derived from the CEM-C1 (resistant) parent resembled one another more than they did the CEM-C7 (sensitive) parent or its subclones. Within the resistant subclones, the hyperploid subclones (regardless of their Dex resistance or sensitivity) resembled each other more than they did the pseudoploid subclones. Individual clones of the same gross phenotype showed considerable overlap in gene expression pattern but did differ with respect to some genes.

To analyze the difference between cells that respond to glucocorticoid by undergoing apoptosis and those that are resistant, we compared three clones. One clone was resistant to apoptosis in the presence of the glucocorticoid Dex. The other two were sensitive; one was a revertant from the CEM-C1-resistant clone. In the constant presence of a Dex concentration that fully occupied the GR, the two sensitive clones, after a delay of at least 24 hours, increasingly underwent apoptosis. We tested the hypothesis that the sensitive clones would show regulation of a distinctive set of genes, prior to the onset of actual apoptosis. Dex or ethanol vehicle was added to each clone and, 20 hours later (a few hours before apoptosis), the cultures were collected, RNA extracted, and mRNA levels examined on Affymetrix HG.U95Av2 chips. After three such experiments, we asked for genes that were consistently induced or reduced in the two sensitive but not the resistant clone. The results supported our hypothesis. Within our chosen limits, a group of 39 genes was consistently induced and one of 21 consistently deinduced only in the apoptosis-sensitive clones. A different set containing 88 genes was regulated consistently in the resistant clone.
Comparing our results with two earlier reports containing data from single experiments on similar cells, but at earlier time points, discovered some concordances in the genes expressed. Many of our discovered genes were not seen as altered in the earlier experiments. The potential reasons are discussed. Overall, the results are consistent with our hypothesis of a sequential network of altered gene expression (both induced and repressed) being responsible for the ultimate apoptosis in the sensitive cells.

Our data also clearly show that the resistant clone responds to Dex by induction or repression of a significant number of genes. Therefore, the resistance is not due to lack of activated GR but to an alteration in the gene responses to GR. We propose that this is caused by an as-yet-undiscovered switch mechanism that converts sensitive to resistant.

ACKNOWLEDGMENTS

The authors would like to recognize the work of Scott Webb and Bruce Luxon for data analysis and both Scott Webb and Aaron Miller in this laboratory for the preparation of figures and graphs.

REFERENCES

GC-EVOKED APOPTOSIS IN CEM HUMAN LYMPHOID CELLS


Harbour DV, Chambron P, Thompson EB 1990 Steroid mediated lysis of lymphoblasts requires the DNA binding region of the steroid hormone receptor. J Steroid Biochem 35:1–9


Holter J, Wakui E, Tazawa H, Treuter E, Gustafsson JA 2001 Regulation of glucocorticoid receptor activity by 14–3-3-dependent intracellular relocalization of the corepressor RIP140. Mol Endocrinol 15:501–511


Houselsey PR, Sanchez ER, Westphal HM, Beato M, Pratt WB 1985 The molybdate-stabilized L-cell glucocorticoid receptor isolated by affinity chromatography or with a monoclonal antibody is associated with 90–92 kDa nonsteroid-binding phosphoprotein. J Biol Chem 260:13810–817


Thompson EA Jr 1991 Insensitivity to the cytolytic effects of glucocorticoids in vivo is associated with a novel “slow death” phenotype. Cancer Res 51:5544–5550


Overlapping but Distinct Profiles of Gene Expression Elicited by Glucocorticoids and Progestins

YIHONG WAN AND STEVEN K. NORDEEN

Department of Pathology and Program in Molecular Biology, University of Colorado Health Sciences Center, Denver, Colorado 80262

ABSTRACT

Glucocorticoids and progestins bind to receptors that share many structural and functional similarities, including virtually identical DNA recognition specificity. Nonetheless, the two hormones mediate very distinct biological functions. For example, progestins are associated with the incidence and progression of breast cancer, whereas glucocorticoids are growth suppressive in mammary cancer cells. To understand the mechanisms that engender biological specificity, we have employed two systematic approaches to identify genes that are differentially regulated by the two hormones. The first strategy is to utilize Affymetrix oligonucleotide arrays to compare glucocorticoid- and progestin-regulated gene expression in a human breast cancer cell line. This global analysis reveals that the two hormones regulate overlapping but distinct sets of genes, including 31 genes that are differentially regulated. Surprisingly, the set of differentially regulated genes was almost as large as the set of genes regulated by both hormones. Examination of the set of differentially regulated genes suggests mechanisms behind the distinct growth effects of the two hormones in breast cancer. The differential regulation of four genes representing different regulatory patterns was confirmed by reverse transcription-polymerase chain reaction (RT-PCR) and Northern blot analyses. Treatment with cycloheximide or mifepristone (RU486) indicates that the regulation is a primary, receptor-mediated event. The second strategy is to employ a retroviral promoter trap and Cre/loxP-mediated, site-specific recombination to identify genes that are differentially regulated by glucocorticoids and progestins. A mouse fibroblast cell line (4F) stably expressing both glucocorticoid receptor (GR) and progesterone receptor (PR) and containing a single copy of a multifunctional selection plasmid was generated. This line was transduced with a self-inactivating retroviral promoter trap vector carrying coding sequences for Cre-recombinase (Cre) in the U3 region. Integration of the provirus places Cre expression under the control of genomic flanking sequence. Activation of Cre expression from integration into active genes results in a permanent switch between the selectable marker genes that convert the cells from neomycin resistant to hygromycin resistant. Selection for hygromycin resistance after hormone treatment yields recombinants in which Cre sequences in the U3 region are expressed from hormone-inducible, upstream cellular promoters. Because Cre-mediated recombination is a permanent event, the expression of the selectable marker genes is independent of ongoing Cre expression. Thus, this system permits the identification of genes that are transiently or weakly induced by hormone. Detailed analyses of genes identified in these studies will furnish a mechanistic understanding of differential regulation by glucocorticoids and progestins.
I. Introduction

Transcription factors are grouped into families based on their most-conserved feature, usually the DNA binding domain. In many cases, family members share similar, if not indistinguishable, DNA sequence-recognition properties. Thus, a fundamental question in molecular biology concerns the extent to which related factors are functionally redundant in a given cell or tissue and the mechanisms by which factor-specific gene regulation is accomplished.

Glucocorticoid receptor (GR) and progesterone receptor (PR) are closely related members of the steroid receptor family of transcription factors (Thornton, 2001). They share many similar structural and functional characteristics, including DNA sequence recognition specificity (Hynes et al., 1983; Payvar et al., 1983; Scheidereit et al., 1983; Cato et al., 1986; Lieberman et al., 1993). The two receptors associate with a similar complex of molecular chaperones in the absence of hormone (Pratt and Toft, 1997) and with a similar set of coactivators in the presence of hormone (McKenna et al., 1999; Westin et al., 2000). Despite the similarity of the two receptors, the cognate hormones display a very distinct spectrum of physiological actions. Classic actions of glucocorticoids include regulation of metabolism, inhibition of inflammation and the immune system, and suppression of bone formation (Porterfield, 1996). The major physiological role of progestins in the mammal are to establish and maintain pregnancy, to promote lobular-alveolar development in the mammary gland, and to suppress milk protein synthesis before parturition (Graham and Clarke, 1997). Even in tissues that express both GR and PR, these two hormones may exert opposite biological actions. For example, in bone, glucocorticoids stimulate bone resorption (Lane and Lukert, 1998; Rackoff and Rosen, 1998; Ziegler and Kasperk, 1998), while progestins prevent bone loss (Nomura et al., 1989; Prior, 1990). In mammary gland, glucocorticoids promote milk protein synthesis and lactation (Doppler et al., 1989; Groner et al., 1994; Groner and Gouilleux, 1995), while progestins inhibit milk production and secretion (Graham and Clarke, 1997). Furthermore, there is an association of progestins with the incidence and progression of breast cancer (Horwitz, 1992), whereas glucocorticoids are growth suppressive in mammary cancer cells (Lippman et al., 1976; Goya et al., 1993).

How can two receptors with such remarkable similarity mediate such dramatically different biological functions? Only a handful of cellular promoters regulated by GR or PR have been identified. Few cellular promoters have been reported to be differentially regulated by GR and PR. Many of the studies on the mechanisms of GR and PR function have utilized the mouse mammary tumor virus (MMTV) promoter. This promoter is induced by both steroids under most circumstances, although chromatin environment may differentially influence MMTV induction by the two hormones by mechanisms as yet poorly understood.
(Lambert and Nordeen, 1998). Understanding the basis of the distinct physiology of glucocorticoids and progestins is severely limited by the paucity of genes and promoters identified to be differentially regulated by the two receptors. In this review, we summarize two systematic approaches that we have employed to address this void: an Affymetrix microarray analysis and a Cre/loxP-mediated retroviral promoter-trapping strategy. The systematic identification of differentially regulated genes reveals potential avenues of differential regulation of cell growth by the two hormones and opens a new avenue for future studies on the molecular mechanisms underlying hormone-specific gene regulation.

II. Affymetrix Oligonucleotide Array Analysis

A. OVERLAPPING BUT DISTINCT GENE REGULATION PROFILES BY GLUCOCORTICOIDS AND PROGESTINS

To understand the basis of distinct actions of glucocorticoids and progestins in a tissue such as the mammary gland, where both receptors are expressed, and especially to identify genes that are differentially regulated by the two hormones, we performed microarray analysis of gene expression in the human breast cancer cell line T47D/A1-2. T47D/A1-2 cells express comparable levels of GR and PR (Nordeen et al., 1989). Total RNA was isolated from cells treated with vehicle, dexamethasone (Dex, 100 nM), or R5020 (10 nM) for 2 or 6 hours. Probes generated from 10\(\mu\)g of this RNA were hybridized to Affymetrix HuGeneFL Arrays to analyze expression of 5600 full-length human genes. In each single array hybridized with cRNAs from T47D/A1-2 cells, 30–40% of the genes exhibit detectable expression.

Figure 1 depicts the set of genes regulated by glucocorticoids and by progestins and the relationship between the two. Genes were included if the GeneChip software called at least a 3-fold change at one or both time points. In addition, the higher value had to be scored as “present.” Of 5600 genes analyzed, 70 were induced by the glucocorticoid, Dex, at either 2-hour or 6-hour treatment, and 47 were induced by the progestin, R5020. Of these, 25 were induced by both hormones (Table I). The number of genes downregulated 3-fold by glucocorticoids and progestins was remarkably similar, 33 and 34, respectively, but only 12 of these were downregulated by both hormones by more than 3-fold (Table II). Most of the regulated genes identified had not been previously described to be regulated by either glucocorticoids or progestins and therefore represent novel hormone-regulated targets. Novel hormone-regulated targets identified in this study will enhance our understanding of the role of these two hormones as both physiological regulators and pharmacological agents.

Of particular interest for our studies, 31 genes have been identified to be differentially regulated by the two hormones by more than 3-fold (Table III). In
In T47D/A1-2 Breast Cancer Cells (GR+/PR+):

Up-regulated > 3 fold   Down-regulated > 3 fold

Differentially Regulated by > 3 fold: 31 genes

<table>
<thead>
<tr>
<th></th>
<th>Dex</th>
<th>R5020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>21</td>
<td>12</td>
</tr>
</tbody>
</table>

FIG. 1. Diagrammatic representation of the sets of genes regulated by glucocorticoids and by progestins. Total RNA was analyzed by Affymetrix oligonucleotide arrays. Data were analyzed using GeneChip Expression Analysis software. Shown are the numbers of genes up- or downregulated by either hormone by more than 3-fold and the number of genes differentially regulated by the two hormones by more than 3-fold.

light of the paucity of genes described to be differentially regulated by the two hormones, it was surprising that the number of genes differentially regulated by more than 3-fold (31) approached the number regulated by both by more than 3-fold (37). Each set represents about 1.5–2% of the genes whose expression could be detected. These results show that glucocorticoids and progestins regulate overlapping but distinct sets of genes.

B. CONFIRMATION OF DIFFERENTIAL REGULATION WITH REVERSE TRANSCRIPTION-POLYMERASE CHAIN REACTION (RT-PCR) AND NORTHERN BLOT ANALYSES

Four differentially regulated genes identified by the array analysis as representing different patterns of regulation were selected for further analysis. G0S8 (also known as RGS2, regulator of G-protein signaling 2) is specifically induced by Dex but not by R5020 (Table III), whereas the promyelocytic leukemia zinc finger protein gene (PLZF) was preferentially induced by glucocorticoids. In contrast, the gene encoding the βB subunit of inhibins and activins (INHBB) is specifically induced by progestins. The IEX-1/Dif-2 gene represents yet another pattern of differential regulation. The array results suggest that expression of IEX-1/Dif-2 is specifically repressed by glucocorticoids. The possible contribution of differential expression of these genes to differential effects of glucocorticoids and progestins will be discussed later.

To determine whether the array analysis gave a true reflection of the expression pattern of these genes, mRNA levels were quantified by both semi-
TABLE 1

Genes Upregulated by Both Dex and R5020 by More Than 3-Fold

<table>
<thead>
<tr>
<th>Probe set</th>
<th>Description</th>
<th>D2:V</th>
<th>D6:V</th>
<th>R2:V</th>
<th>R6:V</th>
</tr>
</thead>
<tbody>
<tr>
<td>D16227_at</td>
<td>BDP-1, recoverin family, calcium-binding</td>
<td>1.4</td>
<td>8.3</td>
<td>-2.5</td>
<td>5.7</td>
</tr>
<tr>
<td>D38037_at</td>
<td>FK506-binding protein 12kDa homologue</td>
<td>1.6</td>
<td>9.2</td>
<td>4.2</td>
<td>5.6</td>
</tr>
<tr>
<td>D86956_at</td>
<td>KIAA0201 gene</td>
<td>1.8</td>
<td>3.5</td>
<td>2</td>
<td>3.1</td>
</tr>
<tr>
<td>HG4310-HT4580</td>
<td>Cellular retinol binding protein Li</td>
<td>3.6</td>
<td>10.7</td>
<td>1.5</td>
<td>7.3</td>
</tr>
<tr>
<td>M14218_at</td>
<td>Argininosuccinate lyase, arginine metabolism</td>
<td>2.2</td>
<td>3.9</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>M83667_rna1_s</td>
<td>NF-IL6-beta protein, C/EBP family</td>
<td>4.1</td>
<td>8.7</td>
<td>3.6</td>
<td>7.5</td>
</tr>
<tr>
<td>U26726_at</td>
<td>11-beta-hydroxysteroid dehydrogenase type 2</td>
<td>7.4</td>
<td>27.1</td>
<td>6.9</td>
<td>21</td>
</tr>
<tr>
<td>U36922_at</td>
<td>Forkhead domain protein, transcription factor</td>
<td>1.5</td>
<td>4.9</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td>U42031_at</td>
<td>54 kDa Immunophilin FKBP54</td>
<td>1.7</td>
<td>6.3</td>
<td>1.5</td>
<td>3.8</td>
</tr>
<tr>
<td>U54999_at</td>
<td>LGN protein, inhibit G alpha i2</td>
<td>2.1</td>
<td>4.2</td>
<td>1.9</td>
<td>5.1</td>
</tr>
<tr>
<td>U62015_at</td>
<td>Cyr61, growth factor-inducible gene</td>
<td>1.5</td>
<td>3.4</td>
<td>1.9</td>
<td>5.1</td>
</tr>
<tr>
<td>U72342_at</td>
<td>Platelet activating factor acetylhydrolase</td>
<td>3.5</td>
<td>3.5</td>
<td>3.3</td>
<td>2.4</td>
</tr>
<tr>
<td>U73524_at</td>
<td>Putative ATP/GTP-binding protein</td>
<td>5.7</td>
<td>6.4</td>
<td>4.3</td>
<td>4.2</td>
</tr>
<tr>
<td>U77456_at</td>
<td>Nucleosome assembly protein 2</td>
<td>3.8</td>
<td>4.3</td>
<td>4</td>
<td>7.4</td>
</tr>
<tr>
<td>U81554_at</td>
<td>CaM kinase isoform II</td>
<td>6.3</td>
<td>5.4</td>
<td>5.4</td>
<td>4.2</td>
</tr>
<tr>
<td>U81556_at</td>
<td>Hypothetical protein A4</td>
<td>5.1</td>
<td>3.5</td>
<td>4.7</td>
<td>4.8</td>
</tr>
<tr>
<td>U83461_at</td>
<td>Putative copper uptake protein (hCTR2)</td>
<td>2.7</td>
<td>3.2</td>
<td>2.8</td>
<td>3.9</td>
</tr>
<tr>
<td>U85267_at</td>
<td>Down syndrome candidate region 1 (DSCR1)</td>
<td>1.6</td>
<td>4.7</td>
<td>2.4</td>
<td>7.7</td>
</tr>
<tr>
<td>U85611_at</td>
<td>DNA-PK interaction protein (KIP)</td>
<td>3.8</td>
<td>3.9</td>
<td>3.7</td>
<td>3.6</td>
</tr>
<tr>
<td>U90426_at</td>
<td>Nuclear RNA helicase</td>
<td>3.5</td>
<td>3.3</td>
<td>3.3</td>
<td>2.3</td>
</tr>
<tr>
<td>U90919_at</td>
<td>Clones 23667 and 23775 zinc finger protein</td>
<td>2.4</td>
<td>3</td>
<td>2.3</td>
<td>3.1</td>
</tr>
<tr>
<td>U91316_at</td>
<td>Acyl-CoA thioester hydrolase</td>
<td>5.5</td>
<td>5.8</td>
<td>3.3</td>
<td>4.2</td>
</tr>
<tr>
<td>U91327_at</td>
<td>12p15 BAC clone CIT987SK-99D8</td>
<td>5.4</td>
<td>4.3</td>
<td>3.7</td>
<td>3.2</td>
</tr>
<tr>
<td>X60073_rna1</td>
<td>Adenylate kinase 3, nucleotide metabolism</td>
<td>2.9</td>
<td>3.2</td>
<td>3.3</td>
<td>3.2</td>
</tr>
<tr>
<td>Z19002_at</td>
<td>PLZF, transcription repressor</td>
<td>7.3</td>
<td>14.4</td>
<td>2.9</td>
<td>4.5</td>
</tr>
</tbody>
</table>

[Abbreviations: V, vehicle; D2, D6, dexamethasone 2 hr, 6 hr; R2, R6, R5020 2 hr, 6 hr.]
quantitative PCR and Northern blot analyses (Wan and Nordeen, 2002b). Both confirmed the selective induction of G0S8/RGS2 by glucocorticoids (Figure 2A). Quantification of the Northern blot showed that Dex treatment led to a more than 20-fold induction, while R5020 had little effect on transcript levels. As for G0S8/RGS2, the RT-PCR and Northern blot analyses confirmed that PLZF is strongly induced by Dex but, unlike G0S8/RGS2, PLZF is also weakly induced by R5020 (Figure 2B). Additionally, the Northern blots indicated that PLZF is expressed as two transcripts, 11 kb and 9 kb. Both transcripts are differentially induced by Dex. The INHBB gene also is expressed as two transcripts, as had previously been reported for rat INHBB (Feng et al., 1989). Levels of both the 4-kb and 3-kb transcripts were increased by R5020, while Dex had no effect on the larger transcript and downregulated the smaller transcript slightly (Figure 2C). Finally, direct analyses of IEX/Dif-2 gene expression also supported the array data. Dex treatment led to an 80% downregulation by 6 hours, whereas R5020 treatment resulted in a weak, transient downregulation, with transcript levels returning to control levels by 6 hours (Figure 2D). Previous studies have

<table>
<thead>
<tr>
<th>Probe set</th>
<th>Description</th>
<th>D2:V</th>
<th>D6:V</th>
<th>R2:V</th>
<th>R6:V</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB000584_at</td>
<td>Novel TGF-beta superfamily protein</td>
<td>1.8</td>
<td>20.7</td>
<td>1.6</td>
<td>20.6</td>
</tr>
<tr>
<td>D14695_at</td>
<td>KIAA0025 gene</td>
<td>2.9</td>
<td>3.3</td>
<td>1.3</td>
<td>3.5</td>
</tr>
<tr>
<td>HG2724-HT2820</td>
<td>Oncogene Tls/Chop, liposarcomas</td>
<td>1.8</td>
<td>14.8</td>
<td>1.5</td>
<td>21</td>
</tr>
<tr>
<td>L07615_at</td>
<td>Neuropeptide Y receptor Y1, G protein coupled</td>
<td>1.3</td>
<td>3.5</td>
<td>1.9</td>
<td>4.4</td>
</tr>
<tr>
<td>L37347_at</td>
<td>Integral membrane protein, iron uptake</td>
<td>2.9</td>
<td>3</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td>L37882_at</td>
<td>Frizzled gene, receptor for Wnt</td>
<td>2.4</td>
<td>6.7</td>
<td>4.6</td>
<td>6.8</td>
</tr>
<tr>
<td>M88461_s_at</td>
<td>Neuropeptide Y receptor YY, G protein coupled</td>
<td>2.2</td>
<td>3.2</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td>U07225_at</td>
<td>P2U nucleotide receptor</td>
<td>1.7</td>
<td>4.7</td>
<td>1.7</td>
<td>4.3</td>
</tr>
<tr>
<td>U07358_at</td>
<td>Protein kinase (zpk), a leucine zipper domain</td>
<td>1.7</td>
<td>3.7</td>
<td>2.1</td>
<td>3.2</td>
</tr>
<tr>
<td>U22376_cds2_s</td>
<td>c-myb, transcription factor, proto-oncogene</td>
<td>1.5</td>
<td>5</td>
<td>3.3</td>
<td>3</td>
</tr>
<tr>
<td>U39840_at</td>
<td>Hepatocyte nuclear factor-3 alpha</td>
<td>1.2</td>
<td>3.3</td>
<td>2.1</td>
<td>3.3</td>
</tr>
<tr>
<td>X51630_at</td>
<td>Wilms tumor WT1, tumor suppressor</td>
<td>1.2</td>
<td>3.2</td>
<td>1.3</td>
<td>3.2</td>
</tr>
</tbody>
</table>

[Abbreviations: V, vehicle; D2, D6, dexamethasone 2 hr, 6 hr; R2, R6, R5020 2 hr, 6 hr.]
### TABLE III
Genes Differentially Regulated by Dex and R5020 by More Than 3-Fold

<table>
<thead>
<tr>
<th>Probe set</th>
<th>Description</th>
<th>Difference*</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>L13391_at</td>
<td>G0S8, Gqalpha inhibitor</td>
<td>15.4</td>
<td>8.5</td>
</tr>
<tr>
<td>Z19002_at</td>
<td>PLZF, transcription repressor</td>
<td>9.6</td>
<td>7.3</td>
</tr>
<tr>
<td>M38258_at</td>
<td>Retinoic acid receptor gamma 1</td>
<td>6.5</td>
<td>-1.6</td>
</tr>
<tr>
<td>U75272_s_at</td>
<td>Gastricsin, protein degradation</td>
<td>6.3</td>
<td>4</td>
</tr>
<tr>
<td>S73591_at</td>
<td>VDUP1, inhibit thioredoxin &amp; growth</td>
<td>5.1</td>
<td>1.2</td>
</tr>
<tr>
<td>U37546_s_at</td>
<td>IAP homolog C</td>
<td>4.6</td>
<td>1.6</td>
</tr>
<tr>
<td>D80010_at</td>
<td>KIAA0188 gene</td>
<td>4.5</td>
<td>2.1</td>
</tr>
<tr>
<td>HG3494-HT3688</td>
<td>Nuclear factor Nf-II6</td>
<td>4.2</td>
<td>-1.1</td>
</tr>
<tr>
<td>M77140_at</td>
<td>Pro-galanin, a neuropeptide</td>
<td>4.1</td>
<td>1.2</td>
</tr>
<tr>
<td>U15932_at</td>
<td>Dual-specificity protein phosphatase</td>
<td>4</td>
<td>1.2</td>
</tr>
<tr>
<td>X77777_s_at</td>
<td>VIP receptor related protein</td>
<td>4</td>
<td>3.8</td>
</tr>
<tr>
<td>M33317_f_at</td>
<td>Cytochrome P450IA4 (CYP2A4)</td>
<td>3.9</td>
<td>4.2</td>
</tr>
<tr>
<td>HG4310-HT4580</td>
<td>Cellular retinol binding protein Li</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td>U63455_at</td>
<td>Sulfonylurea receptor</td>
<td>3.5</td>
<td>1.5</td>
</tr>
<tr>
<td>X17059_s_at</td>
<td>Arylamine N-acetyltransferase</td>
<td>3.4</td>
<td>3</td>
</tr>
<tr>
<td>X51956_rna1_at</td>
<td>ENO2, neuron specific enolase</td>
<td>3.2</td>
<td>1.4</td>
</tr>
<tr>
<td>M29874_s_at</td>
<td>Cytochrome P450-IIB (hIIB1)</td>
<td>-3</td>
<td>-3.8</td>
</tr>
<tr>
<td>HG110-HT110_s</td>
<td>Nuclear ribonucleoprotein A/B</td>
<td>-3.4</td>
<td>-4.8</td>
</tr>
<tr>
<td>L13210_at</td>
<td>Mac-2 binding protein, tumor antigen</td>
<td>-3.4</td>
<td>1.1</td>
</tr>
<tr>
<td>J03474_at</td>
<td>Serum amyloid A, acute phase</td>
<td>-3.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>

*Difference in fold change compared to control conditions.*
shown that the Affymetrix oligoarray technology produces highly reliable results, many of which have been confirmed by conventional approaches (Fambrough et al., 1999; Harkin et al., 1999; Jelinsky and Samson, 1999). In this study, for all four genes tested individually, there is excellent agreement between the microarray results and the Northern blot analyses. However, we chose some of the more robustly regulated genes for this analysis. Obviously, there is a greater likelihood

### TABLE III

(continued)

<table>
<thead>
<tr>
<th>Probe set</th>
<th>Description</th>
<th>Difference(^a)</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>D:R</td>
<td>D2:V</td>
</tr>
<tr>
<td>M62783_at</td>
<td>alpha-N-Acetylgalactosaminidase</td>
<td>−3.9</td>
<td>−3.6</td>
</tr>
<tr>
<td>U34605_at</td>
<td>Retinoic acid, interferon-inducible</td>
<td>−3.9</td>
<td>−3.7</td>
</tr>
<tr>
<td>U52513_at</td>
<td>Retinoic acid-induced gene G</td>
<td>−4</td>
<td>−1.9</td>
</tr>
<tr>
<td>M31682_at</td>
<td>Testicular inhibin beta-B-subunit</td>
<td>−4.5</td>
<td>2</td>
</tr>
<tr>
<td>M80359_at</td>
<td>Protein p78</td>
<td>−5</td>
<td>−3.4</td>
</tr>
<tr>
<td>U23070_at</td>
<td>Putative transmembrane protein</td>
<td>−5</td>
<td>1</td>
</tr>
<tr>
<td>X51441_at</td>
<td>Serum amyloid A (SAA), acute phase</td>
<td>−5.1</td>
<td>−1.5</td>
</tr>
<tr>
<td>U48807_at</td>
<td>MAP kinase phosphatase 2</td>
<td>−5.7</td>
<td>1.3</td>
</tr>
<tr>
<td>U77735_at</td>
<td>pim-2 proto-oncogene, Ser/Thr kinase</td>
<td>−6.2</td>
<td>1.5</td>
</tr>
<tr>
<td>S81914_at</td>
<td>IEX-1/Dif-2, growth factors inducible</td>
<td>−7.3</td>
<td>−4.8</td>
</tr>
<tr>
<td>M74089_at</td>
<td>TB1 gene, FAP and colorectal cancer</td>
<td>−10.5</td>
<td>−8</td>
</tr>
</tbody>
</table>

**controls**

<table>
<thead>
<tr>
<th>Probe set</th>
<th>Description</th>
<th>Difference(^a)</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>D:R</td>
<td>D2:V</td>
</tr>
<tr>
<td>X00351_f_at</td>
<td>beta-Actin</td>
<td>−1.4</td>
<td>1</td>
</tr>
<tr>
<td>U37689_at</td>
<td>RNA polymerase II subunit</td>
<td>1.2</td>
<td>1.1</td>
</tr>
</tbody>
</table>

[Abbreviations: V, vehicle; D2, D6, dexamethasone 2 hr, 6 hr; R2, R6, R5020 2 hr, 6 hr. *Fold difference was calculated using GeneChip algorithms, not a direct ratio of fold changes. The higher value of the two timepoints is shown.*]
FIG. 2. Expression analyses of selected differentially regulated genes. T47D/A1–2 cells were treated with vehicle (V or Veh), dexamethasone (D, 100 nM), or R5020 (R, 10 nM) for 2 hours (D2 or R2) or 6 hours (D6 or R6). Total RNA was isolated and used for semiquantitative reverse transcription-polymerase chain reaction (RT-PCR) and Northern blot analyses. Quantitation of the Northern blot is shown on the right in each panel. (A) G0S8, a gene specifically induced by glucocorticoids. (B) Promyelocytic leukemia zinc finger (PLZF), a gene preferentially induced by glucocorticoids. (C) INHBB (β-subunit of inhibins and activins), a gene specifically induced by progestins. (D) IEX-1, a gene preferentially downregulated by glucocorticoids. [Reprinted with permission from Wan Y, Nordeen SK 2002 Overlapping but distinct gene regulation profiles by glucocorticoids and progestins in human breast cancer cells. Mol Endocrinol 16:1204–1214. Copyright The Endocrine Society.]
of an incorrect call on genes more weakly regulated. Particularly if one sets a cutoff in the range of 1.8- to 2-fold, it is necessary to perform multiple repeats with cRNAs derived from independent experiments (Richer et al., 2002).

The relatively short induction times for gene identification (i.e., 2 and 6 hours) were chosen to optimize for identification of primary, receptor-mediated regulatory events, rather than secondary induction events. To confirm that the differential regulation of the four genes was direct and mediated by the cognate receptors, T47D/A1-2 cells were treated with the protein synthesis inhibitor cycloheximide or the GR/PR antagonist mifepristone (RU486) along with Dex or R5020. Total RNA was isolated and the expression of each gene was determined by semiquantitative RT-PCR. Although cycloheximide itself can have an effect on mRNA levels, the hormone-regulation pattern is maintained following cycloheximide treatment for all four genes, indicating that the hormone regulation is a direct effect that does not require de novo protein synthesis. In contrast, the hormone regulation is completely abolished following RU486 treatment, indicating that hormonal regulation is mediated through GR and/or PR.

C. DIFFERENTIAL GENE REGULATION BY GLUCOCORTICOIDS AND PROGESTINS MAY MEDIATE HORMONE-SPECIFIC EFFECTS

Progestins are associated with the incidence and progression of breast cancer (Horwitz, 1992). Recent work suggests that progestins may prime mammary cells to respond to growth factors (Lange et al., 1999). In contrast, glucocorticoids are growth suppressive in mammary cancer cells (Lippman et al., 1976; Goya et al., 1993). From the examination of the differentially regulated genes identified, a pattern emerged that gives insight into the differential effects on cell proliferation of the two hormones (Figure 3).

A number of the genes that are preferentially induced by glucocorticoids or suppressed by progestins are growth suppressive (Figure 3A; Table III). G0S8/RGS2 encodes a basic helix-loop-helix phosphoprotein (Siderovski et al., 1994) and specifically inhibits the function of Gqα as a GTPase-activating protein (Heximer et al., 1997). Knockout mice show reduced T-cell proliferation and antiviral immunity, increased anxiety responses, and decreased male aggression (Oliveira-Dos-Santos et al., 2000). G0S8/RGS2 has been shown to be induced in growth-arrested cells and to promote adipocyte differentiation (Nishizuka et al., 2001). The induction of G0S8/RGS2 also may play a role in the growth-suppressive effect of glucocorticoids in mammary carcinoma cells.

PLZF originally was identified as the fusion partner of the retinoic acid receptor alpha (RARα) gene in a variant chromosomal translocation in acute promyelocytic leukemia (APL) (Chen et al., 1993). It is a transcription repressor with a kruppel-like zinc finger domain and a BTB/POZ domain (Chen et al., 1993; Dong et al., 1996). PLZF represses transcription by recruiting a histone
deacetylase through the silencing mediator of retinoic acid and thyroid hormone receptor (SMRT)-mSin3-histone deacetylase (HDAC) co-repressor complex (David et al., 1998). PLZF has been shown to suppress the growth of myeloid cells by inducing G0/G1 arrest and apoptosis, partly through the binding and repression of the cyclin A2 promoter (Shaknovich et al., 1998; Yeyati et al., 1999).

Growth-suppressive genes are also targets of progestin regulation. Of note are genes preferentially suppressed by R5020, including RAR gamma 1 (RARγ1) and vitamin D3 upregulated protein 1 (VDUP1). RARγ inhibits proliferation and activates apoptosis in breast cancer cells (Fanjul et al., 1996; Raffo et al., 2000). VDUP1 suppresses cell proliferation by inhibiting the reducing potential of the disulfide-reducing protein thioredoxin and by down-regulating thioredoxin expression (Nakamura et al., 1992, 1997; Nishiyama et al., 1999). Expression of VDUP1 is downregulated in chemically induced rat mammary tumors (Yang et al., 1998). Thus, in mammary cancer cells, the growth-suppressive effect of glucocorticoids may be mediated through the induction of G0S8/RGS2 and PLZF, while the growth-promoting effect of progestins may be mediated through the downregulation of RARγ and VDUP1.

FIG. 3. Differential gene regulation by glucocorticoids and progestins may mediate hormone-specific effect on mammary cancer cell proliferation. (A) Genes that are preferentially induced by glucocorticoids (G) or suppressed by progestins (P) are growth suppressive. (B) Genes that are preferentially induced by progestins or suppressed by glucocorticoids are potentially proliferation related. RAR = retinoic acid receptor, VDUP = vitamin D3 upregulated protein, MKP = MAP kinase phosphatase, NaGalase = alpha-N-acetylgalactosaminidase.

<table>
<thead>
<tr>
<th>A.</th>
<th>Gene</th>
<th>Regulation</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>G0S8</td>
<td>G↑</td>
<td>Gqa inhibitor, ↑ growth, ↑ differentiation</td>
<td></td>
</tr>
<tr>
<td>PLZF</td>
<td>G↑</td>
<td>transcription repressor, ↓ growth, ↑ apoptosis</td>
<td></td>
</tr>
<tr>
<td>RARγ1</td>
<td>P↓</td>
<td>↓ growth, ↑ apoptosis</td>
<td></td>
</tr>
<tr>
<td>VDUP1</td>
<td>P↓</td>
<td>tumor suppressor, ↓ thioredoxin &amp; growth</td>
<td></td>
</tr>
</tbody>
</table>

B. | Gene | Regulation | Function |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mac-2 BP</td>
<td>P↑</td>
<td>tumor antigen, ↑ in breast cancer</td>
<td></td>
</tr>
<tr>
<td>INHBB</td>
<td>P↑</td>
<td>↑ mammary gland growth &amp; development</td>
<td></td>
</tr>
<tr>
<td>Pim-2</td>
<td>P↑</td>
<td>protooncogene, ↑ lymphoid tumors with c-myc</td>
<td></td>
</tr>
<tr>
<td>NaGalase</td>
<td>G↓</td>
<td>extracellular matrix-degradation, ↑ in cancer</td>
<td></td>
</tr>
<tr>
<td>MKP-2</td>
<td>G↓</td>
<td>↑ by growth factors, ↑ in v-jun or K-ras tumors</td>
<td></td>
</tr>
<tr>
<td>IEX-1</td>
<td>G↓</td>
<td>↑ by serum &amp; growth factors, ↓ differentiation</td>
<td></td>
</tr>
</tbody>
</table>

deacetylase through the silencing mediator of retinoic acid and thyroid hormone receptor (SMRT)-mSin3-histone deacetylase (HDAC) co-repressor complex (David et al., 1998). PLZF has been shown to suppress the growth of myeloid cells by inducing G0/G1 arrest and apoptosis, partly through the binding and repression of the cyclin A2 promoter (Shaknovich et al., 1998; Yeyati et al., 1999).

Growth-suppressive genes are also targets of progestin regulation. Of note are genes preferentially suppressed by R5020, including RAR gamma 1 (RARγ1) and vitamin D3 upregulated protein 1 (VDUP1). RARγ inhibits proliferation and activates apoptosis in breast cancer cells (Fanjul et al., 1996; Raffo et al., 2000). VDUP1 suppresses cell proliferation by inhibiting the reducing potential of the disulfide-reducing protein thioredoxin and by down-regulating thioredoxin expression (Nakamura et al., 1992, 1997; Nishiyama et al., 1999). Expression of VDUP1 is downregulated in chemically induced rat mammary tumors (Yang et al., 1998). Thus, in mammary cancer cells, the growth-suppressive effect of glucocorticoids may be mediated through the induction of G0S8/RGS2 and PLZF, while the growth-promoting effect of progestins may be mediated through the downregulation of RARγ and VDUP1.
In contrast, many of the genes that are preferentially induced by progestins or suppressed by glucocorticoids are potentially proliferation related (Figure 3B; Table III). INHBB is induced by progestins but not glucocorticoids (Table III). Inhibins and activins are members of the transforming growth factor beta (TGFβ) superfamily, which are potent mediators of proliferation or antiproliferation and differentiation in different cell types. A suggestion that INHBB is involved in growth regulation in the mammary gland comes from studies in which female mice deleted for both alleles of INHBB have lactation failure due to retarded ductal elongation and alveolar morphogenesis during puberty, pregnancy, and parturition (Robinson and Hennighausen, 1997). Other preferentially progesterone-induced genes include Pim-2, a proto-oncogene that induces lymphoid tumors synergistically with c-myc in mice (Allen et al., 1997; Baytel et al., 1998), and Mac-2 binding protein/90K, a tumor-derived antigen expressed at elevated levels in the serum of patients with breast and other types of cancer (Fusco et al., 1998).

IEX/Dif-2 represents but one of several genes whose downregulation by glucocorticoids may contribute to the inhibition of growth or tumorigenesis. IEX-1 was identified as a radiation-inducible, immediate-early gene in human squamous carcinoma cells (Kondratyev et al., 1996). Also known as Dif-2, it is downregulated during monocyte differentiation (Pietzsch et al., 1997). The IEX-1/Dif-2 gene is induced by multiple signals, many of which are associated with proliferation (e.g., lipopolysaccharide, C2-ceramide, lysophosphatidylcholine, phorbol esters, serum, growth factors) (Pietzsch et al., 1997; Schafer et al., 1999). The association of IEX-1/Dif-2 with proliferation suggests that the growth-suppressive effect of glucocorticoids in mammary carcinoma cells is mediated, in part, through the downregulation of IEX-1/Dif-2.

Other dexamethasone-downregulated genes include MAP kinase phosphatase (MKP-2) and alpha-N-acetylgalactosaminidase (NaGalase). MKP-2 can be induced by growth factors and is upregulated in cells transformed by v-Jun or mutated K-ras (Fu et al., 2000; Yip-Schneider et al., 2001). NaGalase is an extracellular matrix-degrading enzyme that is produced exclusively by cancer cells (Yamamoto et al., 1996). NaGalase levels in mice bearing squamous cell carcinoma increased with time of tumor growth and were directly proportional to tumor burden (Korbelik et al., 1998). Thus, in mammary cancer cells, growth-suppressive effect of glucocorticoids may be mediated through the downregulation of NaGalase, MKP-2, and IEX-1/Dif-2, while growth-promoting effect of progestins may be mediated through the induction of Mac-2 BP90K, INHBB, and Pim-2.

In addition to the effects on mammary cancer cell growth, glucocorticoids promote milk protein synthesis and lactation (Doppler et al., 1989; Groner et al., 1994; Groner and Gouilleux, 1995), while progestins inhibit milk production and secretion (Graham and Clarke, 1997). Vasoactive intestinal polypeptide (VIP)
receptor-related protein is the long isoform of VIP receptor. VIP has been shown to be a physiological mediator of prolactin release in the rat (Abe et al., 1985). Our data suggest that glucocorticoid-specific stimulation of lactation may be, in part, mediated by glucocorticoid-specific induction of VIP receptors (Table III).

To test the implications of the array data, along with the suggestions of other studies indicating that glucocorticoids and progestins have different effects on the proliferation of the mammary gland and mammary cancer cells, we investigated the effects of hormones on T47D/A1-2 cells. We assessed cell growth by monitoring the total DNA content of hormone-treated cell populations (Figure 4A). The results demonstrated that Dex treatment inhibited cell growth throughout the entire time course, while R5020 treatment initially stimulated cell growth and then became inhibitory. To examine further the effect of glucocorticoids and progestins on the cell-cycle progression of T47D/A1-2 cells within the first 24 to 48 hours after hormone treatment, we also assessed cell cycle by flow cytometry (Figure 4B). Typical of a nonsynchronous population of proliferating cells, approximately 30% of untreated cells are in S or G2/M phases. Dex treatment decreased the fraction of cells in S/G2/M throughout the entire time course. By 48 hours, the fraction of cells in S/G2/M had declined from 30% to 13%.

FIG. 4. Glucocorticoids and progestins mediate different effects on the growth of T47D/A1-2 cells. (A) T47D/A1-2 cells were treated with vehicle, Dex (100 nM), or R5020 (10 nM) for the indicated number of days. Cell proliferation was determined by measuring DNA content in each well using Hoechst DNA assay. The experiment shown is representative of three independent experiments, each done in triplicate ± standard error (SE). (B) T47D/A1-2 cells were treated with vehicle, Dex (100 nM), or R5020 (10 nM) for the indicated number of hours. The fraction of cells in S+G2/M was measured by flow cytometry. The ordinate indicates the difference between the fraction of cells in S+G2/M in the hormone-treated sets and the fraction of cells in S+G2/M in the vehicle-treated controls. In this study, approximately 25–35% of control cells were in S+G2/M. [Reprinted with permission from Wan Y, Nordeen SK 2002 Overlapping but distinct gene regulation profiles by glucocorticoids and progestins in human breast cancer cells. Mol Endocrinol 16:1204–1214. Copyright The Endocrine Society.]
suggesting that Dex treatment led to a cell-cycle arrest in G1. This supports the observation in a previous study that Dex suppresses the growth of Con8 rat mammary tumor cells by inducing a G1/G0 cell-cycle arrest (Goya et al., 1993). In contrast, R5020 treatment initially increased the fraction of cells in S+G2/M. At the peak (i.e., around 18 hours), almost half of the cell population was in S+G2/M, a 63% increase over control. Together with the initial increase in DNA content, this indicates that progestins initially are growth promoting. By 30–36 hours after R5020 treatment, the fraction of cells in S+G2/M dropped below controls, to a level similar to Dex-treated cells, indicating that after the initial stimulation, progestins mediate a growth-suppressive action. This biphasic effect of progestins on the cell-cycle progression of T47D cells has been previously described (Musgrove et al., 1991; Groshong et al., 1997). It has been proposed that progestin treatment initially drives cells to go through the first cell cycle to a decision point at the G1/S boundary. Secondly, it induces cellular changes that permit other factors to influence the ultimate proliferative or differentiative state of the cells (Lange et al., 1999). For example, only progestin-primed T47D cells become highly sensitive to the proliferative effects of epidermal growth factor (EGF) (Groshong et al., 1997). Together, the T47D/A1-2 cell growth and cell-cycle studies clearly demonstrated that, during the first 24 hours of hormone treatment, glucocorticoids are growth suppressive, while progestins are growth stimulating. The early time points (i.e., 2 hours and 6 hours) used in our microarray analysis permitted the identification of candidate genes that may mediate the differential growth effects of the two hormones on T47D/A1-2 cells.

In summary, the genes identified to be differentially regulated by glucocorticoids and progestins provide potential mechanisms through which the two hormones exert different or opposite biological effects. With a better understanding of the functionally uncharacterized genes identified here, more potential mechanisms will emerge. Future studies exploring these mechanisms will assist in developing tumor markers and therapeutic agents for cancer. In addition, genes identified in this study can be used as model systems to investigate the molecular mechanisms underlying differential gene regulation by glucocorticoids and progestins. Our preliminary studies have mapped the glucocorticoid-mediated inhibition of tumor necrosis factor alpha (TNFα)-induced IEX-1/Dif-2 promoter activity to a small proximal promoter region containing a nuclear factor kappa beta (NF-κB) element and juxtaposed SP-1-CCAT/enhancer binding protein (C/EBP) elements (Y. Wan, unpublished data). The mutually inhibitory actions of GR and NF-κB have been well documented (McKay and Cidlowski, 1999). The fruits of these investigations will enhance our appreciation of the larger question of how related transcription factors mediate distinct, even opposing, biological actions.
III. Cre/loxP-mediated Retroviral Promoter-trapping System

A. ESTABLISHMENT OF A CRE/LOXP-MEDIATED RETROVIRAL PROMOTER-TRAPPING SYSTEM

While microarray analysis is of unquestioned utility in many approaches to the study of hormonal regulation, like any methodology, it has specific shortcomings. For example, detection of genes expressed at low levels can be a problem. Also, hormone inductions that are transient or that induce apoptosis can present technical challenges. To address these difficulties and to identify promoters differentially induced by glucocorticoids and progestins, we have adapted a novel retroviral promoter-trapping system developed by von Melchner and colleagues (Russ et al., 1996).

To generate a cell line that expresses both GR and PR, we stably introduced a PR-B expression vector into a glucocorticoid-responsive Ltk\(^{-}\) fibroblast cell line (Thackray et al., 1998). The resulting cell line, 4F, is responsive to both glucocorticoids and progestins, although PR is at least 10-fold more abundant than GR in each 4F cell (Thackray et al., 1998). A multifunctional reporter plasmid was introduced into the 4F cells whose purpose was the sensitive detection of Cre recombinase expression (Figure 5). The ppgklxtkneo/hygro reporter plasmid consists of two, tandemly arrayed selective marker genes that are transcribed from a pgk promoter. The 5’-selective marker gene encodes for a fusion protein between herpes simplex virus 2 thymidine kinase (tk) and the neomycin phosphotransferase (neo) and is flanked by two direct repeats of the loxP-recombination target site (Bartsch et al., 1996). The 3’-selective marker

![Diagram of the multifunctional selection plasmid (ppgklxtkneo/hygro)](image-url)

gene encodes for hygromycin phosphotransferase (hygro) and is located downstream of the second loxP site. In the absence of Cre, the tkneo gene is expressed from the constitutive pgk promoter. The hygro gene does not have a promoter immediately upstream and chimeric transcripts initiated at the pgk promoter are suppressed by two tandem copies of the bovine growth hormone polyadenylation sequence upstream of hygro. Therefore, the cells containing a single copy of ppgkltxtkneo/hygro are G418 resistant but hygromycin sensitive. However, when Cre is expressed, it will catalyze recombination between the two loxP sites, resulting in the deletion of tkneo. This places hygro immediately downstream of the pgk promoter, rendering the cells G418 sensitive but hygromycin resistant (Figure 5). Thus, 4F cells with a stably integrated ppgkltxtkneo/hygro selection plasmid report even transient expression of Cre by undergoing a permanent switch in their drug-resistance phenotype.

A G418-resistant clone containing a single copy of the selection plasmid was selected for further analysis. First, we determined whether the G418-resistant clone was sensitive to hygromycin by placing $5 \times 10^5$ cells under hygromycin (500 μg/ml) selection. Because no colony formed within 12 days, we concluded that neither leaky hygro translation nor spontaneous recombination occurred in these cells at levels that would interfere. Second, we determined whether the 4F clone with a single integrated copy of the selection plasmid would become hygromycin resistant and neomycin sensitive after Cre expression. An expression plasmid (pCMVCre) was transiently transfected into the cells by a calcium phosphate coprecipitation method. Cells were selected for hygromycin (500 μg/ml) resistance. The hygromycin-resistant cells were pooled and placed under G418 (1 mg/ml) selection for 12 days to test neomycin sensitivity. Because all the hygromycin-resistant cells were neomycin sensitive, we concluded that the cells could undergo a clean, drug-resistance phenotype switch after Cre expression. Thus, Cre expression, even transiently, results in a permanent phenotypic change that can be selected.

In order to utilize this property to trap cellular promoters, we employed a self-inactivating retroviral gene-trap vector (U3Cre) (Russ et al., 1996) to deliver a promoterless Cre gene to the 4F target cells with the integrated selection plasmid. The U3Cre plasmid was stably transfected into an amphitropic packaging cell line, PA317 (Miller and Buttimore, 1986). Supernatants from lines producing high titers of recombinant virus were used to infect 4F target cells with a single copy of the ppgkltxtkneo/hygro selection plasmid, resulting in a library consisting of approximately $4 \times 10^6$ independent proviral integration events. The structure of the U3Cre vector, when it is integrated in a proviral form, is shown in Figure 6. Cre coding sequences are positioned near the 5’ end of the U3 region of an enhancer-deleted long terminal repeat (LTR). Viral replication and LTR-mediated duplication place the Cre coding sequence in the 5’ LTR, just 30 nucleotides from the flanking cellular DNA (Figure 6) (von Melchner and Ruley,
Therefore, Cre is not expressed unless the virus integrates into an active cellular promoter. A potential complication arises from the duplication of the Cre coding sequence in the 3′ LTR. In clonal cell lines with a stably integrated U3Cre provirus, there typically are two Cre transcripts: a cellular-proviral fusion transcript initiating from the upstream cellular promoter and terminating at the polyadenylation site of the 5′ LTR, and a “viral genomic” transcript initiating from the 5′ proviral LTR and terminating at the polyadenylation site of the 3′ LTR (Russ et al., 1996). The double Cre transcripts can complicate the analysis of the selected cell lines, since the viral genomic transcript can obscure the Cre transcript from the cellular promoter. However, translation of Cre from the viral genomic transcript within the 3′ LTR is unlikely due to multiple, short, open-reading frames positioned between the LTRs. Furthermore, several previous

FIG. 6. Structure and replication of the U3Cre retroviral vector. Functional components are labeled as follows: Cre, modified bacterial phage P1 Cre recombinase; U3, R, and U5 are segments of the viral long terminal repeat (LTR), “-enh” indicates “enhancer deleted.” Infection of 4Fneo/hygro21 target cells generates provirus in which the 5′ Cre coding sequence is immediately downstream of cellular sequence. NeoR, G418 resistant. [Reproduced by permission of the Society of Endocrinology from Wan Y, Nordeen SK 2002 Identification of genes differentially regulated by glucocorticoids and progestins using a Cre/loxP mediated retroviral promoter-trapping strategy. J Mol Endocrinol 28:177–192.]
studies have shown that, in most cases, activation of the promoter-trap is associated with the translation of Cre from the cellular-proviral fusion transcripts (von Melchner and Ruley, 1989; von Melchner et al., 1990; Reddy et al., 1991; Chang et al., 1993; Russ et al., 1996). Therefore, the 3′ copy of Cre doesn’t contribute to Cre protein expression and interfere with the assessment of promoter trapping. In this work, we employ Cre expression coupled to drug selection phenotypes to identify hormone-regulated promoters.

B. ISOLATION OF CELL LINES REPRESENTING TRAPPED, HORMONE-INDUCIBLE, AND DIFFERENTIALLY HORMONE-INDUCIBLE PROMOTERS

Clonal cell lines were isolated following sequential drug selections, as outlined in Figure 7. Clones in which U3Cre had integrated near constitutive promoters, resulting in constitutive Cre expression, were eliminated because Cre-mediated recombination eliminates the neomycin phosphotransferase gene. This renders these cells sensitive to continued selection in G418. The selection scheme diagrammed for cell clones of group A enriches for promoters induced by progestins, whereas group B is enriched for promoters induced by glucocorticoids. Trapped promoters preferentially induced by glucocorticoids vs. progestins are enriched in cell clones of group C; conversely, promoters preferentially induced by progestins are trapped in the scheme giving group D cell clones. Forty cell lines from each of group A and group B, and an additional 20 cell lines from each of group C and group D, were isolated for further analysis.

To determine the expression pattern of the gene represented in each clonal cell line isolated, we developed a reporter assay to quantitate the Cre protein level, based on Cre recombinase activity (Figure 8). A reporter plasmid (ppgkltkneo/luc) (Figure 8A) was constructed in which the hygro gene of the multifunctional selection plasmid was replaced by luciferase (luc). Luciferase expression from this plasmid was dependent on Cre expression. A titration of Cre expression vector demonstrated that the higher the input of Cre, the higher the output of luciferase expression (Figure 8B). Therefore, this sensitive and quantitative luciferase reporter strategy can determine the exact regulation pattern of the trapped promoters in each isolated cell clone and allow the identification of the lines that represent genes that are differentially regulated by the two hormones.

Each clonal cell line with potentially hormone-regulated, trapped promoters was transfected with ppgkltkneo/luc and treated with Dex or R5020 24 hours later. The induction of Cre activity was assessed by the deletion of sequences flanked by loxP sites, as measured by the appearance of luciferase activity. In 58 cell lines, luciferase expression was induced by either glucocorticoids or progestins by more than 2-fold. In 19 cell lines, luciferase expression was differentially regulated by the two hormones by more than 1.5-fold (Figure 8C).
C. HORMONE REGULATION OF SELECTED PROMOTERS

In order to determine whether the endogenous cellular genes trapped in the clonal cell lines are, indeed, hormone regulated, genomic DNA sequences representing a trapped promoter that is induced or differentially induced by glucocorticoids and progestins. Retrovirus-transduced cells were placed first under G418 (1 mg/ml) selection for 5 days. Those cells in which the Cre gene is integrated downstream of a constitutively active cellular promoter will be eliminated due to the expression of Cre and thus the deletion of the tkneo gene. The survivors were split into two groups in media without G418. One group was treated with R5020 (10 nM) and the other with Dex (100 nM) for 24 hours. The cells then were subjected to hygromycin (500 µg/ml) selection, to select for cells in which Cre expression has been induced by hormone. Thus, this selection will enrich for cells in which the Cre gene was integrated downstream of a progestin-inducible promoter (group A) or a glucocorticoid-inducible promoter (group B). To maximize the identification of genes preferentially induced by one of the two hormones, an additional step was added to the selection scheme. After the initial G418 selection to eliminate constitutively expressed promoters, another selection was added to eliminate promoters induced by progestins (group C) or glucocorticoids (group D). This was done by adding the appropriate hormone for 24 hours and continuing G418 selection for 5 days. Survivors in the progestin-treated group C were removed from G418 and treated with Dex for 24 hours. Similarly, the glucocorticoid-treated group D was removed from G418 and treated with R5020 for 24 hours. The two groups were then placed under hygromycin selection to enrich for cells expressing Cre under the control of the appropriate hormone. GC, glucocorticoids; PG, progestins; Dex, dexamethasone, a synthetic glucocorticoid; R5020; a synthetic progestin. [Reproduced by permission of the Society of Endocrinology from Wan Y, Nordeen SK 2002 Identification of genes differentially regulated by glucocorticoids and progestins using a Cre/loxP mediated retroviral promoter-trapping strategy. J Mol Endocrinol 28:177–192.]
upstream of the proviral integration site of several clonal isolates were retrieved by inverse PCR and sequenced. Recovered sequences ranged from 154 to 874 bp. Blast searches using different databases revealed that the 441-bp sequence upstream of the proviral integration site in clone 43 is 98% identical with the expressed sequence tag (EST) ic84f11.x1 in the cDNA library “Melton Normal-
ized Mixed Mouse Pancreas 1 N1-MMS1’ (dbEST ID 9280663, GenBank accession number BI438077). However, the sequences upstream of the proviral integration sites in clone 32, 42, and 66 showed no significant homology to known genes or ESTs at the time of analysis. This was not unexpected, since this trapping system is very sensitive to even low levels of Cre expression, so even promoters with weak basal expression are eliminated in the initial selection. Because of their very low expression, identification of these transcription units might be easily missed by conventional gene-discovery approaches. Another reason that cDNAs for these transcription units may not be found in the databases stems from the design of the U3Cre retroviral construct that selects for integration sites in 5′ nontranslated regions (von Melchner et al., 1990; see Wan and Nordeen, 2002a, for discussion). The 5′-most part of the transcription unit is most likely to be absent from all but full-length cDNA clones in the databases. The sequences of the novel transcripts have been deposited in GenBank; accession numbers are AF465703 (for clone 32 gene), AF465704 (for clone 42 gene), and AF465705 (for clone 66 gene).

Three of the four proximal flanking sequences (clone 32, 42, and 43) were able to be amplified by RT-PCR from the RNA isolated from the parental cell line (4F), suggesting that the upstream sequences are within an exon of the gene. The primers designed for the upstream sequences of clone 66 did not give RT-PCR product, even though they can amplify the genomic DNA. The integration site in this clone may not be in exonic sequence or, more likely, the 5′ primer is upstream of the transcription start site.

The mRNA level of the gene represented by clone 32 (clone 32 gene) in the parental cell line (4F) was measured by semiquantitative RT-PCR. Serially diluted cDNAs were amplified by PCR to select the cycle number that gave the best linear relationship between the cDNA input (x value) and the PCR output (y value) for each specific primer set (Figures 9A and B). Then, the cDNAs in each hormone-treated sample were amplified at this cycle number, with the linear standard run at the same time. The PCR output in each sample was quantitated and the amount of cDNA input was calculated using the linear relationship derived from the standard (Figure 9C). The result demonstrated that clone 32 gene was preferentially induced by glucocorticoids (5-fold), compared to progestins (2.5-fold). Although this is only a 2-fold difference in induction, it is more impressive when one considers that progesterone receptors are 10-fold more abundant than glucocorticoid receptors in these cells. Expression of the genes represented by the other two clones was quantitated in the same way. Clone 43 gene also was preferentially induced by glucocorticoids. Glucocorticoids induced gene 43 by two-fold, while progestins gave no induction or a slight inhibition. Clone 42 gene was induced by both hormones to a similar level. For all three clones, the mRNA-regulation patterns of the endogenous genes agree
with the Cre protein-regulation patterns shown in Figure 8C, suggesting that the genes identified using this strategy are, indeed, regulated by hormones.

For two clones (32 and 43), we further examined the expression of the cellular-Cre fusion transcript to ascertain whether it is regulated by hormones in the same fashion as the endogenous gene. Clonal cell lines 32 and 43 were treated with hormone for 2 hours and total RNA was isolated. Semiquantitative RT-PCR analyses were performed with a 5′ primer from the endogenous gene and a 3′ primer from Cre sequence. A PCR product of the predicted size was amplified from cDNA. Furthermore, both fusion transcripts were preferentially induced by glucocorticoids compared to progestins, as was observed for the endogenous genes.

The hormone regulation of clone 32 gene was examined further to ascertain whether the induction is a direct hormone effect and whether the differential

---

![Graphs](image.png)

**FIG. 9.** Quantitation of the mRNA level of clone 32 gene in 4F cells by semiquantitative RT-PCR. (A) Parental cells (4F) were treated with vehicle (V), dexamethasone (D, 100 nM), or R5020 (R, 10 nM) for 2 hours or 6 hours. Total RNA was isolated and used for semiquantitative RT-PCR. The left four lanes show the standard. (B) The linear relationship between the cDNA input and the PCR output for clone 32 gene (left) and the β-actin control (right) at the indicated PCR cycle. (C) Quantitation of clone 32 gene-specific cDNA normalized by β-actin-specific cDNA in each hormone-treated sample. The result was shown as fold of induction, compared to vehicle-treated sample. The figure depicts the results of three independent experiments. Bars represent ± SE. [Reproduced by permission of the Society of Endocrinology from Wan Y, Nordeen SK 2002 Identification of genes differentially regulated by glucocorticoids and progestins using a Cre/loxP mediated retroviral promoter-trapping strategy. J Mol Endocrinol 28:177–192.]
induction is mediated by the appropriate receptor. Treatment with the protein synthesis inhibitor cycloheximide did not abolish hormone induction, suggesting that the regulation is a direct event independent of de novo protein synthesis. Treatment with the GR/PR antagonist RU486 blocked the glucocorticoid induction but ZK112993, which preferentially blocks progesterone receptor at the dose used, did not, demonstrating that the induction of clone 32 gene by glucocorticoid is mediated by GR. The glucocorticoid induction of the clone 32 gene peaked by 30 minutes, again indicating that this is a direct response. The hormone induction lasted for at least 22 hours. Assessment of the hormone regulation of clone 32 gene in two other cell lines indicated that the glucocorticoid induction of clone 32 gene is cell-type specific. Clone 32 gene is induced by glucocorticoid in mouse fibroblast cell lines (L929 and 4F) but not the GR-expressing human breast cancer cell line T47D/A1-2. In summary, the Cre/loxP-mediated retroviral promoter-trapping system described here has successfully identified novel genes that are differentially induced by glucocorticoids and progestins.

IV. Summary

In this review, we describe two different strategies that we have employed to identify genes that are differentially regulated by glucocorticoids and progestins. The first strategy, Affymetrix microarray analysis, is a high-throughput approach that allows quick examination of the expression of thousands of known genes at the same time. The design of the Affymetrix oligoarrays and the supporting statistical analysis algorithm permit the accurate determination of the relative level of each mRNA species. This global analysis reveals that glucocorticoids and progestins regulate overlapping, but distinct, sets of genes, including 31 genes that are differentially regulated. Surprisingly, the set of differentially regulated genes was almost as large as the set of genes regulated by both hormones. Examination of the set of differentially regulated genes suggests mechanisms behind the distinct growth effects of the two hormones in breast cancer.

The second strategy combines a retrovirally delivered promoter trap and Cre/loxP-mediated, site-specific recombination to identify differentially hormone-regulated genes. This strategy has several advantages, compared to microarray analysis. It is proficient at the detection of transiently hormone-activated promoters. Since the Cre-mediated recombination at loxP sites is an irreversible event, even transiently activated promoters can produce enough Cre recombinase to create a permanent switch of drug-resistance phenotype. Thus, screening can be accomplished even when a brief induction period is used to avoid cytotoxicity or induction of apoptosis that might result from a longer treatment (Russ et al., 1996). In addition, the promoter trap is not biased toward highly expressed genes. Little Cre expression
is required to catalyze recombination at the \textit{loxP} sites. Therefore, unlike many conventional strategies (e.g., cDNA library screening, RNA differential display, microarray analysis), this Cre/\textit{lox} selection strategy can detect weakly expressed genes. The promoter trap allows discovery of unknown genes. Finally, it allows a near-saturation screening of the whole genome. If we infect the target cell line with the U3Cre retrovirus at a multiplicity of infection (MOI) of 0.5 to ensure single integration events and perform selections on $10^7$ integrants, this will yield one integration per 300 base pairs, assuming random integration. Nonetheless, the advantage of the exquisite sensitivity of this system also can be a drawback. Because the selection against any basal expression is strong, regulated promoters exhibiting some basal level of expression will be missed. Future application of this system might benefit from damping the sensitivity through the use of a less-active Cre or suboptimal \textit{loxP} sites. This might permit some level of Cre expression without promoting recombination of the selection vector sequences. Using this system, a group of 19 clonal cell lines representing genes differentially regulated by glucocorticoids and progestins have been isolated. Two novel genes (clone 32 and 43 genes) have been confirmed to be preferentially induced by glucocorticoids, compared to progestins.

Despite the disparate biological activities of the two receptors, there is little information on what genes may be differentially regulated by the two receptors and how this may be accomplished. Our studies document novel approaches toward addressing this void. Further studies of the mechanisms underlying the differential regulation will enhance our understanding of how glucocorticoids and progestins function as specific physiological regulators or therapeutic agents.

\section*{ACKNOWLEDGMENT}

This work has been supported by grant award DK37061 from the National Institutes of Health to S.K.N. and predoctoral fellowship award DAMD17-99-1-9445 from the Department of Defense Breast Cancer Research Program to Y.W.

\section*{REFERENCES}


Differential Gene Expression by Steroid Hormones


Feng ZM, Bardin CW, Chen CL 1989 Characterization and regulation of testicular inhibin beta-subunit mRNA. Mol Endocrinol 3:939–948


Goya L, Maiyar AC, Ge Y, Firestone GL 1993 Glucocorticoids induce a G1/G0 cell cycle arrest of Con8 rat mammary tumor cells that is synchronously reversed by steroid withdrawal or addition of transforming growth factor-alpha. Mol Endocrinol 7:1121–1132


Miller AD, Buttimore C 1986 Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol Cell Biol 6:2895–2902

Musgrove EA, Lee CS, Sutherland RL 1991 Progestins both stimulate and inhibit breast cancer cell cycle progression while increasing expression of transforming growth factor alpha, epidermal growth factor receptor, c-fos, and c-myc genes. Mol Cell Biol 11:5032–5043


DIFFERENTIAL GENE EXPRESSION BY STEROID HORMONES


Porterfield SP 1996 Endocrine Physiology. St. Louis: Mosby-Year Book; 139–146


Thornton JW 2001 Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc Natl Acad Sci USA 98:5671–5676


Analysis of Gene Expression in the Normal and Malignant Cerebellum

ROBERT J. WECHSLER-REYA

Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710

ABSTRACT

The developing nervous system consists of a small number of multipotent precursors that undergo extensive proliferation to generate the neurons and glia that make up the adult brain. Elucidating the mechanisms that control the growth and differentiation of these cells is important not only for understanding normal neural development but also for understanding the etiology of central nervous system tumors. A particularly striking example of this is in the cerebellum. Recent studies have suggested that the Sonic hedgehog-Patched signaling pathway plays a critical role in regulating the proliferation of cerebellar granule cell precursors and is also a major target of mutation in the cerebellar tumor medulloblastoma. In light of these observations, identification of additional genes that control cerebellar growth and differentiation is likely to provide important insight into the basis of cerebellar tumors. Similarly, analysis of gene expression in medulloblastoma will no doubt shed light on previously unknown signaling pathways that regulate normal cerebellar development. The advent of high-throughput gene expression analysis techniques — such as adapter-tagged competitive polymerase chain reaction (ATAC-PCR), serial analysis of gene expression (SAGE), and DNA microarrays — makes identification of such genes faster and easier than ever before. This review summarizes recent studies of gene expression in the cerebellum and discusses the value of such approaches for understanding development and tumorigenesis in this tissue.

I. Development of Cerebellar Granule Cells

The cerebellum is required for motor coordination and has been implicated in a variety of cognitive and affective functions as well (Leiner et al., 1993; Altman and Bayer, 1997). These functions depend on precise interactions among at least five types of neurons (Altman and Bayer, 1997). The most abundant of these are the Purkinje cells, which carry signals from the cerebellum to other parts of the brain, and the granule cells, which regulate the activity of Purkinje cells. The murine cerebellum contains about $10^8$ granule cells, more than the total number of neurons in the rest of the brain. The critical importance of these cells is evident from mutant mice, in which loss of granule cells leads to severe ataxia (Mullen et al., 1997), and from patients with congenital granule cell degenera-
tion, who have severe deficits in motor coordination, language use, and cognitive function (Pascual-Castroviejo et al., 1994).

Granule cell development has a number of unique features that distinguish it from other kinds of neurogenesis (Hatten and Heintz, 1995; Altman and Bayer, 1997). Whereas most neurons are born around the ventricles and then migrate outward toward the surface of the brain, granule cells are generated on the outside of the cerebellum and migrate inward (Figure 1). Granule cell precursors (GCPs) initially arise from a dorsal hindbrain structure called the rhombic lip (Alder et al., 1996; Alcantara et al., 2000; Wingate, 2001). During embryonic life, these cells leave the rhombic lip and stream across the outer surface of the cerebellum to form a region called the external germinal layer (EGL). After birth, cells in the EGL undergo extensive proliferation to generate a large pool of GCPs (Fujita et al., 1966; Mares et al., 1970). As new GCPs are generated, older cells move inward, then exit the cell cycle and differentiate (Fishell and Hatten, 1991; Komuro and Rakic, 1998). The differentiating cells extend axons that form synapses with Purkinje cells, then continue to migrate inward past the Purkinje cell bodies to their final destination, the internal granule layer (IGL). The waves of GCP proliferation and differentiation continue until about 3 weeks of age, at which time the EGL disappears and all GCPs complete their migration and differentiation into mature granule cells (Nicholson and Altman, 1972).

FIG. 1. Granule cell development. Granule cells are generated in the external germinal layer (EGL). After birth (P0), they proliferate to generate a large pool of precursors (white circles, P0–P14). These cells move inward, exit the cell cycle, and differentiate (gray circles). They extend axons that synapse with Purkinje cells (gray ovals). Finally, cells migrate through the molecular layer (ML) and Purkinje cell layer (PL) to the internal granule layer (IGL). By adulthood, all precursors have become mature granule cells (black circles) and no EGL remains.
II. Control of Granule Cell Precursor Proliferation

The observation that mutations in the Sonic hedgehog (Shh) signaling pathway result in cerebellar tumors in both mice and humans (Johnson et al., 1996; Goodrich et al., 1997; Raffel et al., 1997) suggested that this pathway might play a role in normal cerebellar growth and differentiation. Our own studies (Wechsler-Reya and Scott, 1999) and others (Dahmane and Ruiz-i-Altaba, 1999; Wallace, 1999) demonstrated that Purkinje cells make Shh and that developing GCPs express all of the elements of the signaling pathway necessary to respond to it. Addition of recombinant Shh protein to GCPs in culture induces a 75- to 100-fold increase in thymidine incorporation within 3 days. Treatment of cerebellar slice cultures with Shh also causes a dramatic proliferative response and inhibits granule cell differentiation and migration. Finally, intracranial injection of Shh-blocking antibodies during early postnatal development leads to decreased proliferation of GCPs and a significant reduction in the thickness of the EGL. Together, these studies suggest that Shh is a critical regulator of GCP proliferation.

A number of factors present in the outer EGL enhance Shh-induced proliferation of GCPs. One of these is the extracellular matrix molecule, laminin (Wechsler-Reya and Scott, 1999; Pons et al., 2001). GCPs cultured on laminin show a 1.5- to 2-fold increase in Shh-induced proliferation, compared to cells cultured on other substrates. The chemokine stromal cell-derived factor-1 alpha (SDF-1α), produced by the pia mater that surrounds the EGL, induces GCP chemotaxis and also can synergize with Shh to enhance GCP proliferation in culture (Klein et al., 2001). Finally, heparan sulfate proteoglycans (HSPGs), which have been shown to be required for maximal hedgehog signaling in Drosophila, are present in the postnatal EGL and can increase Shh-induced proliferation of GCPs (Rubin et al., 2002). While none of these factors induces significant proliferation on its own, all may contribute to the proliferation of cells in the EGL.

Growth factors that act through receptor tyrosine kinases (RTKs) – epidermal growth factor (EGF), insulin-like growth factor (IGF-1), and basic fibroblast growth factor (bFGF) – have been reported to induce GCP proliferation (Gao et al., 1991; Tao et al., 1996; Ye et al., 1996; Lin and Bulleit, 1997). On their own, these factors cause a 2- to 4-fold increase in thymidine incorporation in cultured GCPs. Interestingly, however, they show no evidence of synergy with Shh; rather, EGF and IGF-1 cause a modest suppression of Shh-induced proliferation and bFGF reduces the Shh response by 90% (Wechsler-Reya and Scott, 1999). This suggests that RTK-binding growth factors control proliferation through mechanisms distinct from the one used by Shh or that they act on distinct subsets of cerebellar cells. The fact that EGF, IGF-1, and bFGF also have potent effects on granule cell survival and differentiation (Hatten et al., 1988; Dudek et al.,
III. Granule Cell-cycle Exit and Differentiation

GCPs undergo extensive proliferation during the first 2–3 weeks after birth. But even as proliferation in the EGL reaches its peak, some GCPs are beginning to exit the cell cycle and differentiate into neurons. One possible explanation for this cell-cycle exit could be movement of cells away from Shh protein. For example, Shh might be present only in the outer EGL and, as cells move inward and away from the mitogenic signal, they might exit the cell cycle. Although this model has not been completely ruled out, the available evidence argues against it. Shh is made by Purkinje cells (Wechsler-Reya and Scott, 1999) and the protein has been detected throughout the EGL (Gritli-Linde et al., 2001; Pons et al., 2001). Moreover, Shh expression begins during embryonic life and persists into adulthood (Traiffort et al., 1999). Thus, termination of the proliferative response probably is not due to reduced exposure to Shh. Rather, it is likely to result from reduced responsiveness to Shh or conversion of the proliferative response into a differentiative one.

To date, three signals have been shown to overcome Shh-induced proliferation: the extracellular matrix molecule vitronectin (VN), the protein kinase A (PKA) activator forskolin, and bFGF. In contrast to laminin, VN is found primarily in the inner EGL and granule cells in this region express alpha-v integrins, which can function as VN receptors (Pons et al., 2001; Wechsler-Reya, 2001). GCPs grown on VN show a 20% reduction in the proliferative response to Shh and increased β-tubulin expression and neurite outgrowth, compared to cells grown on other substrates. Growth on VN also causes increased phosphorylation of the transcription factor cAMP response element binding protein (CREB). CREB function appears to be required for VN’s effects on neurite outgrowth (Pons et al., 2001). These observations suggest that VN may contribute to granule cell-cycle exit and differentiation.

Forskolin also can inhibit Shh-induced proliferation of GCPs (Wechsler-Reya and Scott, 1999; Kenney and Rowitch, 2000). This is not surprising, since PKA activators can inhibit Shh responses in most tissues and in most species (Li et al., 1995; Hammerschmidt et al., 1996; Ungar and Moon, 1996). Because forskolin is not a physiologic signal, there must be an endogenous factor that activates PKA for this to play a role in granule cell development. One good candidate for a PKA activator is pituitary adenylate cyclase activating polypeptide (PACAP) (Sherwood et al., 2000). PACAP is made by Purkinje cells (Nielsen et al., 1998; Skoglosa et al., 1999) and its receptors are expressed by GCPs in the outer EGL (Basille et al., 1993). To date, most studies of PACAP
in the cerebellum have indicated that it increases proliferation of GCPs and promotes survival in the presence of various death-inducing stimuli (Vaudry et al., 1999, 2002; Tabuchi et al., 2001). On the other hand, recent studies of neurons in the embryonic cortex and hindbrain indicate that PACAP can inhibit proliferation of those cells (Suh et al., 2001; Lelievre et al., 2002). Thus, PACAP might contribute to cell-cycle exit in the cerebellum as well.

Finally, bFGF is a potent inhibitor of proliferation, reducing the response to Shh by 90% (Wechsler-Reya and Scott, 1999). bFGF is made by cerebellar astrocytes and Purkinje cells (Hatten et al., 1988; Matsuda et al., 1994). At least two receptors capable of binding it (i.e., fibroblast growth factor (FGF)-1 and -4) are expressed in GCPs in the EGL (el-Husseini et al., 1994; Miyake et al., 1995; Ozawa et al., 1996; Meiri et al., 1998; Sleptsova-Friedrich et al., 2001). The observation that bFGF can prevent proliferation is consistent with a variety of studies demonstrating that it promotes granule cell differentiation and neurite outgrowth in vitro and in vivo (Williams et al., 1994; Hatten and Heintz, 1995; Saffell et al., 1997; Liu and Kaczmarek, 1998). However, as mentioned earlier, bFGF has been reported to induce proliferation of cerebellar cells in vitro and following subcutaneous injection (Tao et al., 1996; Cheng et al., 2001). In fact, we have observed that bFGF induces a small increase in thymidine incorporation in our cultures as well (Wechsler-Reya and Scott, 1999). One possible explanation for these observations is that the proliferative response to bFGF occurs in a subset of GCPs or in a distinct class of progenitor cells that are present in the postnatal cerebellum.

As granule cells differentiate, they extend axons and migrate inward, past the Purkinje cells, to the IGL. Bergmann glia are thought to provide a substrate for the migration (Edmondson and Hatten, 1987; Hatten, 1990; Komuro and Rakic, 1998) and therefore are a likely source of signals and guidance cues. Autocrine signaling by brain-derived neurotrophic factor (BDNF) is critical for initiation of migration from the EGL (Borghesani et al., 2002). Astrotactin, a granule cell surface molecule with EGF and fibronectin repeats, has been shown to be important for guiding granule cell migration (Edmondson et al., 1988; Zheng et al., 1996). Migration also has been shown to be regulated by N-methyl-D-aspartate (NMDA) receptor activity and intracellular calcium concentration (Komuro and Rakic, 1993, 1996).

IV. Gene Expression Analysis of the Developing Cerebellum

The studies described herein have revealed a number of important factors that control proliferation and differentiation of granule cells. However, many important questions remain. For example, the molecular mechanisms by which Shh and other factors regulate growth of GCPs and the identity of the signals that control granule cell differentiation and migration in vivo are poorly understood.
Moreover, the molecules that regulate growth and differentiation of other cell types in the developing cerebellum have not been identified. One important tool for identifying such molecules is gene expression analysis. By analyzing changes in gene expression during the course of cerebellar development, or by isolating specific cell types and comparing their gene expression under conditions of growth and differentiation, it may be possible to identify molecules that are important in normal growth and differentiation.

One of the first efforts at broad-based gene expression analysis in the cerebellum was carried out by Kuhar et al. (1993). These investigators raised polyclonal antisera against immature GCP and adsorbed them against PC12 cells (a cell line resembling peripheral neurons) and adult cerebellum to deplete non-GCP-specific antibodies. They then used the antisera to screen a cDNA expression library derived from GCPs, to identify GCP-specific genes. Using this approach, they cloned 39 unique cDNAs, 28 of which represented novel genes and showed tissue- and stage-specific expression. The majority were expressed at high levels early in development (i.e., postnatal days 0–10), then downregulated by adulthood. Among these genes were at least four distinct expression patterns:

1) those that were localized in the outermost part of the EGL, where GCPs are proliferating; 2) those expressed in the inner EGL, where cells have exited the cell cycle and begun to differentiate; 3) genes expressed in the upper IGL, in postmitotic cells that had not yet completed migration; and 4) genes expressed in fully differentiated, postmigratory granule cells. These genes (some of which have now been cloned) (Miwa et al., 1999) not only provide valuable molecular markers for the stages of granule cell development but also represent a set of molecules that may play important functional roles in this process.

More-recent studies of gene expression in the cerebellum have taken advantage of high-throughput approaches using polymerase chain reaction (PCR) or microarrays. For example, Matoba and colleagues (2000a,b,c) generated cDNA libraries from cerebella of 4-day-old (P4), 12-day-old (P12), and 6-week-old (6W) mice and sequenced several thousand clones from each library. They then chose approximately 400 of these genes (i.e., those with the highest expression levels) and quantitated their expression levels using a technique known as adapter-tagged competitive PCR (ATAC-PCR) (Figure 2A). This involves isolating RNA from each sample (i.e., stage of development), converting it to cDNA, and then “tagging” one end of that cDNA with an adapter. For each sample, the cDNA is tagged with a different-sized adapter. The samples from each stage are mixed into a single tube and amplified by PCR using an adapter-specific primer and a gene-specific primer. After separation by gel electrophoresis, products from each sample can be discriminated based on their size (determined by the unique adapter). The amount of each fragment reflects the amount of original template. Relative expression levels in each sample can be deduced from their signal intensities.
Using this type of analysis, Matoba et al. (2000a,b) determined the expression of genes in the P4, P12, and 6W samples and divided genes into groups based on expression profiles. Although a number of different cell types were represented, the differences in gene expression profiles correlated best with the development of granule cells, which are by far the most abundant cell type. Early in development (P4), when GCPs are undergoing rapid proliferation, the most abundant genes are those encoding ribosomal proteins, cytoskeletal proteins (e.g., tubulin, actin, thymosin β4), and genes associated with proliferation (e.g., cdc2 kinase). Later in development (P12), when granule cells are engaged in axon extension and synapse formation, genes for mitochondrial activities (cytochrome c oxidase, mitochondrial proton/phosphate symporter), ion channels (Na/K-ATPase), and markers of differentiated neurons (NeuroD, glutamate transporter, trkC) are elevated. Expression of many of these genes continues to increase during development. Finally, at 6 weeks, when differentiation of cerebellar neurons and glia is complete, elevated expression of synaptic proteins (e.g., soluble NSF (N-ethylmaleimide-sensitive factor) attachment protein SNAP25), neurotransmitter receptors (gammaaminobutyric acid receptors, GABA-Rs), and components of myelin (myelin basic protein and myelin proteolipid protein) are detected.

FIG. 2. Methods of genomic analysis. Procedures for adapter-tagged competitive polymerase chain reaction (ATAC-PCR), serial analysis of gene expression (SAGE), and Affymetrix GeneChip (Microarray) analysis are illustrated. [See text for details.]
In a subsequent study, Matoba et al. (2000c) expanded their analysis to include 1800 genes expressed over six different time points in cerebellar development: 2, 4, 8, and 12 days and 3 and 6 weeks. Their results again were clustered into 1) genes elevated early in development (e.g., 2, 4, and 8 days); 2) genes elevated later in development (e.g., 12 days, 3 weeks, 6 weeks); and 3) genes with complex expression patterns. About 80% of these could be classified based on function. The investigators looked for correlations between temporal expression profiles and functional categories and found that genes expressed early in development were enriched for those encoding cancer-related proteins (e.g., adenomatous polyposis coli (APC), c-erbA, abl, Krox24/Egr1) and ribosomal proteins. Other genes in this group were involved in RNA processing, cell adhesion/migration (e.g., VN, tenascin, neural cell adhesion molecule (NCAM), reelin), signaling, and transcription. In contrast, the genes expressed primarily in late postnatal development included those involved in protein transport, carbohydrate metabolism, “brain-specific functions,” ion channels and transporters (ryanodine receptor 2, potassium channels, sodium channels, Na/K ATPases), lipid metabolism and neurotransmitter receptors (GABA-A receptors, NMDA receptors, AMPA (alpha-amino-3-hydroxy-5-methylisoxasole propionic acid) receptors) and synapse components, and genes associated with mature oligodendrocytes (myelin basic protein, myelin proteolipid protein). Having found a number of known genes with documented functions in cerebellar development, the authors suggest that this approach may be used to identify novel genes that regulate granule cell growth and differentiation.

Gene expression in the postnatal cerebellum has also been studied by Kaltschmidt and Kaltschmidt (2001). These investigators initially were interested in the role of the transcription factor nuclear factor kappa B (NF-κB) in neuronal survival. They found that NF-κB is expressed at high levels in the EGL before P7 and declines between P7 and P12. To identify signals that might account for this change in expression, the researchers used cDNA arrays to compare gene expression in P4 vs. P12 cerebellum. The arrays they used were commercial nylon membranes (macroarrays) spotted with 588 known cDNAs. mRNA from each developmental stage was labeled with 32P and hybridized to a filter. Then, the level of hybridization was quantitated and compared using a phosphorimager. While most of the genes did not change, a small group (~40/588) was developmentally regulated; most of these increased between P4 and P12 but some decreased. Changes in expression were found among genes associated with proliferation and differentiation (c-myc, cyclin D3, cdc-like kinase 2, Id3), apoptosis and DNA repair (growth arrest and DNA damage (GADD45), topoisomerase II, TNF receptor-associated death domain (TRADD)), cell-signaling molecules (extracellular signal-regulated kinase (ERK)1, ERK3, jun kinase (JNK)3, mitogen-activated protein kinase (MAPK)2), cell adhesion (contactin 1, integrin β4, N-cadherin, lymphocyte function-associated antigen 1 (LFA-1)), and
growth factors and their receptors (insulin receptor, fibroblast growth factor receptor (FGFR)1, transforming growth factor (TGF)-β2, macrophage colony-stimulating factor (MCSF), fms-like tyrosine kinase (FLT-3) ligand).

The most highly induced gene during this period of development was the cytokine, TGF-β2. The authors confirmed expression of TGF-β2 in the EGL by in situ hybridization (ISH) and antibody staining and showed that expression levels increased significantly from P4 to P12. To test the functional effects of TGF-β2, they added the factor to granule cells in culture and showed that it causes a rapid decrease in expression of the transcription factor, NF-κB. Since TGF-β2 can promote granule cell apoptosis (de Luca et al., 1996) and NF-κB can promote survival (Koulich et al., 2001; Piccioli et al., 2001), the authors suggest that increased expression of TGF-β2 may play a role in the apoptosis that occurs during normal granule cell development in vivo (Wood et al., 1993).

Among the other genes whose expression increased from P4 to P12 were a number of elements of the FGF signaling pathway, including FGF receptor 1, MAPK2, ERK1, ERK3, and CREB2. As previously discussed, FGF signaling can promote granule cell-cycle exit and differentiation in vitro. The fact that multiple components of the FGF signaling pathway are coordinately regulated during a period of granule cell differentiation is consistent with the possibility that it plays this role in vivo as well. However, further studies of the effects of FGF signaling on granule cell development are necessary to validate this notion.

A more-comprehensive study of gene expression in the developing cerebellum was carried out by Zhao et al. (2002). These investigators initially examined gene induction in GCPs following Shh treatment, to identify target genes induced by this stimulus. They isolated cells from P4–5 cerebellum and treated them with Shh protein for 3 or 24 hours, then harvested RNA for microarray analysis using Affymetrix GeneChips (Figure 2B). These arrays consist of thousands of short oligonucleotides (25-mers) conjugated to a solid matrix. Each gene is represented by 10–15 oligonucleotides that match the coding sequence and a corresponding set of oligonucleotides that contain a one-base mismatch; the latter serve as controls for the specificity of hybridization. To analyze gene expression in a sample, RNA is isolated, converted to cDNA, and transcribed to generate biotin-labeled cRNA. This cRNA is hybridized to the Chip, incubated with a streptavidin-conjugated fluorescent dye, and scanned to detect the amount of fluorescence for each oligonucleotide. Analytical software is used to determine the expression levels of each gene and to compare expression between samples.

In their experiments, Zhao and colleagues found that ~4% of the 13,000 genes on their microarrays were upregulated by Shh. Among these genes, ~90% were involved in cell-cycle regulation (e.g., cyclins D1, B1, B2, A2, proliferating cell nuclear antigen (PCNA), E2F1, Ki67, cdc20). Other genes overexpressed in Shh-treated cells included transcription factors (Pax2, Math1, Gli2, N-myc) and genes involved in DNA replication and ribosome assembly. For comparison, the
researchers also examined cells that were growth arrested and then treated with Shh, to identify genes that are induced by Shh in postmitotic granule cells. Although Shh did induce gene expression in these cells, the targets were largely distinct from those induced in GCPs.

To test whether the targets of Shh in proliferating GCPs were expressed in these cells in vivo, Zhao et al. examined expression of a number of genes by ISH. These studies revealed that 10/12 of the genes analyzed were expressed in EGL at postnatal day 1–7 and were downregulated by postnatal day 15. Based on these findings, the authors proposed the concept of a temporal gene regulation profile (TRP) for the EGL: a set of genes that is temporally regulated in a pattern consistent with expression and function in proliferating GCPs in the EGL. In other words, genes whose expression was high from P1–P7 and low after P15 were likely to represent genes expressed in the EGL. If this concept could be validated, EGL-specific genes could be identified based solely on their expression profile during postnatal cerebellar development.

To test this hypothesis, the authors performed microarray analysis on whole cerebellum from P1–P30. In an initial test, they found that 10/12 genes that were known to be expressed in the EGL were expressed at high levels in P3–7 and downregulated from P15–30. To further validate this, they used microarray data to generate a list of genes whose developmental expression pattern fit that of the TRP-EGL, then examined their localization by ISH. The TRP-EGL profile correctly predicted EGL expression for more than 80% of the genes. These studies suggest that one can prospectively identify genes expressed in GCPs based on temporal changes in expression within the whole cerebellum.

These studies demonstrate that gene expression analysis can be a powerful tool for studying development and can provide valuable information about the genes that regulate cell growth and differentiation in a tissue. But when studying changes in gene expression during development – especially within an intact tissue – it is important to keep several issues in mind. First, changes in gene expression in the tissue may reflect changes in expression levels within a given cell type or changes in abundance of that cell type within the tissue. In fact, for much of cerebellar development, the abundance of granule cells tends to swamp out gene expression from most other cell types, including Purkinje cells, astrocytes, oligodendrocytes, and other interneurons. In order to learn about the genes that control development of these cells, it may be necessary to isolate them using laser capture microdissection (LCM) (Vincent et al., 2002) or antibody-based purification methods such as panning or fluorescence-activated cell sorting (Trotter and Schachner, 1989; Baptista et al., 1994; Catapano et al., 2001; Sawamoto et al., 2001). Alternatively, studies of mutant mice that have defects in particular cell types (Mullen et al., 1997) may shed light on the gene expression patterns important for particular aspects of cerebellar development.
Regardless of the cell type or stage being studied, it is clear that gene expression analysis is only a first step towards understanding the molecular mechanisms of development. Once a profile of gene expression is identified, additional criteria must be used to narrow down the list of genes to a manageable number for further study. For example, one might focus on transcription factors or secreted proteins or genes that are likely to control cell adhesion or movement. Alternatively, it might be useful to perform additional screens to determine which of the expressed genes is dysregulated in a particular mutant mouse. Given a short list of genes, functional analysis — both \textit{in vitro} and \textit{in vivo} — will be necessary to understand how these genes contribute to a particular developmental process.

In addition to their importance in understanding cerebellar development, studies of gene expression in the normal cerebellum may have important implications for understanding the etiology of cerebellar tumors. For example, a number of genes expressed in the EGL and in GCPs stimulated with Shh (e.g., cyclin D1, N-myc, gli2) also have been shown to be expressed at high levels in some types of medulloblastoma (Pomeroy et al., 2002). This raises the possibility that other genes involved in regulating cerebellar growth, differentiation, and apoptosis might play a role in tumorigenesis.

\section*{V. Medulloblastoma}

Medulloblastoma is a highly malignant tumor of the cerebellum. It occurs most commonly in children between 5 and 10 years of age and accounts for 20–30\% of all pediatric brain tumors (Schiffer, 1997; Zakhary et al., 2001). The disease usually is treated with a combination of surgery, radiation, and chemotherapy. While these approaches often are effective at shrinking the primary tumor, recurrence and metastasis are common and only 50\% of patients survive for 5 years after diagnosis. Moreover, aggressive treatment of children with radiation and chemotherapy has been found to impair intellectual and physical development (Zakhary et al., 2001). New approaches to the diagnosis and treatment of medulloblastoma clearly are necessary and are most likely to come from a deeper understanding of the cellular and molecular basis of this disease.

Histologically, medulloblastoma often is divided into two major subtypes: classic and desmoplastic (Zakhary et al., 2001; Pomeroy et al., 2002). Classic medulloblastoma consists of small, round, densely packed cells that show little evidence of morphologic differentiation. Desmoplastic (or nodular) medulloblastomas contain regions of densely packed cells with extensive reticulin fibers surrounding “pale islands” of cells that are much less dense. The majority of medulloblastomas have the classic morphology; only 20–25\% of tumors are considered desmoplastic. The cell of origin is not clear for either subtype but some studies suggest that desmoplastic tumors may arise from GCPs in the EGL,
whereas classic tumors may derive from multipotent precursors that surround the ventricles and normally give rise to Purkinje cells, cerebellar interneurons, and glia (Katsetos and Burger, 1994; Buhren et al., 2000).

An important step in understanding the molecular basis of medulloblastoma came from studies of the Sonic hedgehog-Patched signaling pathway (Ingham and McMahon, 2001). As discussed, Shh is a secreted molecule that plays a critical role in embryonic development and is a major regulator of proliferation in the developing cerebellum (Dahmane and Ruiz-i-Altaba, 1999; Wechsler-Reya and Scott, 1999). Patched (Ptc) is a transmembrane protein that functions as both a Shh receptor and an antagonist of Shh signaling. Several lines of evidence indicate that Shh-Ptc signaling is involved in medulloblastoma. First, the human patched gene has been identified as the locus mutated in Gorlin’s syndrome, a disease characterized by skin tumors, craniofacial defects, and increased incidence of medulloblastoma (Hahn et al., 1996; Johnson et al., 1996). Second, many sporadic medulloblastomas (especially those of the desmoplastic type) have been found to harbor mutations in patched and other elements of the Shh pathway (Pietsch et al., 1997; Raffel et al., 1997; Lam et al., 1999; Taylor et al., 2002). Finally, mice in which the patched gene has been disrupted develop tumors that resemble medulloblastoma (Goodrich et al., 1997; Hahn et al., 2000).

Among the other genes associated with human medulloblastoma are components of the Wnt signaling pathway. Patients with Turcot’s syndrome— which results from mutations in the APC gene— have a high incidence of colorectal cancers and brain tumors, especially medulloblastoma (Hamilton et al., 1995). About 4% of sporadic medulloblastomas have been shown to contain APC mutations (Huang et al., 2000). In addition, 8–15% have been reported to harbor activating mutations in β-catenin and 12% have mutations in Axin, a negative regulator of Wnt signaling (Zurawel et al., 1998; Eberhart et al., 2000; Huang et al., 2000; Dahmen et al., 2001). These findings suggest that a subset of medulloblastomas may result from activation of the Wnt pathway. Medulloblastomas also have been found to exhibit overexpression of the transcription factors N-myc, c-myc, pax5, and zic and of the receptor tyrosine kinase ErbB2 (Garson et al., 1989; Bigner et al., 1990; Gilbertson et al., 1995; Kozmik et al., 1995; Yokota et al., 1996). Whether these genes contribute to the development or progression of medulloblastoma, or whether they simply represent markers of the transformed cell type, remains to be determined. Animals in which these genes are misexpressed will provide valuable insight into their role in the etiology of medulloblastoma.

One of the most important genes involved in medulloblastoma has yet to be identified. Of human medulloblastomas, 30–50% have a deletion or rearrangement of part of chromosome 17 (Cogen and McDonald, 1996; Bigner et al., 1997; Burnett et al., 1997). In most cases, the short arm (17p) is lost and head-to-head apposition of the long arms (17q) occurs, which is referred to as
isochromosome 17q \([i(17q)]\). This rearrangement frequently is detected in leukemias, lymphomas, and cancers of the stomach, colon, and cervix. The loss of 17p in a number of types of cancer suggests that at least one potent tumor suppressor gene is located there. Fine mapping of deletions from different tumors has narrowed the region of interest considerably. Most investigators now believe that the putative tumor suppressor is located at 17p13.3, a region of \(\approx\) 20 known genes, including those encoding the lissencephaly-associated protein Lis1, the breakpoint cluster region (BCR)-related gene \(ABR\), the Max-binding protein Mnt, and the transcription factor hypermethylated in cancer-1 (Hic1) (Koch et al., 1996; Steichen-Gersdorf et al., 1997; Sommer et al., 1999; Hoff et al., 2000; Rood et al., 2002). Although none of these genes has been linked definitively to the etiology of medulloblastoma, there is no question that when the chromosome 17 tumor suppressor is identified, it will provide important insight into the basis of cerebellar tumors as well as into normal cerebellar development.

VI. Gene Expression Profiling of Medulloblastoma

In an effort to identify important genes involved in medulloblastoma, a number of investigators have begun to carry out gene expression analysis on primary tumor samples. One of the first efforts was performed by Michiels and colleagues (1999), who analyzed gene expression in a human medulloblastoma sample using serial analysis of gene expression (SAGE) (Figure 2C). SAGE is based on the principle that a sequence of 9–10 nucleotides can be used to identify a transcript, if the position of these nucleotides within the transcript is known. A biotinylated oligo(dT) primer is used to synthesize cDNA from mRNA and, after digestion with a restriction enzyme, fragments are isolated from the 3′ end of each transcript. Fragments are ligated to linkers, then cleaved with a restriction enzyme to release a short sequence (i.e., 9–10 bp) from a defined region of the original cDNA (a “tag”). Tags are ligated together to form long concatamers, which are cloned and sequenced. Using this approach, one sequencing reaction can yield information about the number and distribution of many different tags. With the appropriate software, the sequence and position of the tag can be used to identify the gene from which each tag was derived. Moreover, the number of times each tag appears in the sequencing reaction is proportional to the abundance of that gene in the original sample. In this manner, SAGE gives a quantitative representation of gene expression in a sample.

For their study, Michiels and colleagues used SAGE to compare genes expressed in medulloblastoma to those expressed in 24.5-week fetal brain. They sequenced 10,000 tags from each sample and found about 6000 unique genes in each case. Among the most highly expressed genes in both samples were those encoding ribosomal proteins, consistent with the idea that both of these tissues contain highly proliferative (and protein-synthesizing) cells. Comparing the
medulloblastoma and fetal brain samples, the investigators found 138 genes whose expression differed significantly. About half of these were known genes, including the transcription factors Zic1, Otx2, and Sox4; the secretory protein secretogranin; vascular endothelial growth factor (VEGF); and PCNA. Expression of Zic1 and Otx2 was examined by Northern blotting and found to be elevated in a number of independent medulloblastoma samples. The fact that both of these genes are expressed in the EGL lend support to the notion that at least some cases of medulloblastoma arise from GCPs.

A particularly elegant study of medulloblastoma gene expression was carried out by Pomeroy and colleagues (2002), who used Affymetrix GeneChips to compare medulloblastoma to other types of brain tumors. Since pathologists commonly have grouped medulloblastoma with primitive neuroectodermal tumors (PNETs) from other parts of the brain, the investigators were particularly interested in determining how similar these tumor types were. Studying the gene expression profiles of 42 brain tumors — including medulloblastoma, malignant glioma, atypical teratoid/rhabdoid tumors, and PNETs — they found that each of these tumor types was molecularly distinct. For example, gliomas expressed markers of astrocytic and oligodendrocytic cells (phosphoprotein-enriched in astrocytes-15 kDa (PEA15), SRY-related box2 (SOX2), peripheral myelin protein 2 (PMP2), Olig-2, S100, glial fibrillary acidic protein (GFAP)), whereas medulloblastomas expressed genes characteristic of cerebellar granule cells (e.g., Zic, neurological stem cell leukemia transcription factor (NSCL1)). Interestingly, although medulloblastoma and PNET look similar morphologically, they are molecularly distinct, with PNETs lacking Zic and NSCL1 and expressing high levels of nicotinic cholinergic receptor subunits and DNA polymerase delta1 (POLD1).

Pomeroy et al. next addressed whether the two major subtypes of medulloblastoma described by pathologists — classic and desmoplastic — had distinct gene expression profiles. Consistent with the observation that desmoplastic tumors have a high incidence of Shh pathway mutations (Pietsch et al., 1997; Taylor et al., 2002), the researchers showed that desmoplastic tumors express high levels of Patched, Gli1, IGF-2, and N-myc, which have been identified as transcriptional targets of the Shh pathway (Goodrich et al., 1996; Hahn et al., 2000; Wetmore et al., 2000; Zhao et al., 2002). In addition, desmoplastic tumors express high levels of the antiapoptotic protein Bcl2 and ribosomal proteins, similar to those seen in normal EGL GCPs. In contrast, classic tumors show elevated levels of TGFβ3 and elements of its signaling pathway (Smad2, Smad5) and increased expression of the transcription factors distal-less homeobox gene 7 (DLX7), LIM-homeobox domain protein-2 (LH-2), and NeuroD3. These data clearly indicate that desmoplastic and classic medulloblastoma are not only histologically distinct but also have different gene expression profiles.
In interpreting these differences, it is important to consider several possibilities. First, they may reflect differences in the cell type of origin or cellular composition of the tumors. For example, desmoplastic tumors may arise from GCPs, while classic tumors may arise from multipotent precursors in the ventricular zone. On the other hand, these tumors could arise from the same cell type. Differentially expressed genes might reflect activation of distinct signaling pathways that contribute to the etiology of tumor formation. Determining the functional significance of each of these genes is an important task for future study.

Independent of the role of these genes in the etiology of tumor formation, they may be significant as indicators of tumor prognosis and responsiveness to therapy. This is particularly important in medulloblastoma, since a significant percentage of patients may be cured by chemotherapy and radiation, while others are resistant and succumb to the disease. To address this possibility, Pomeroy et al. analyzed tumors from patients who had been treated and followed to determine the outcome of treatment. The data then were analyzed using a “supervised learning” scheme to determine which genes were best correlated with long-term survival. Among the best predictors of survival were genes associated with cerebellar differentiation (the vesicle coat protein β-NAP, the transcription factor NSCL1, the neurotrophin receptor TrkC, and sodium channels), and genes encoding extracellular matrix proteins (procollagen lysyl hydroxylase (PLOD), collagen type V, elastin). On the other hand, poor prognosis was associated with increased expression of ribosomal proteins, the proliferation-associated genes B-myb and E2F5, and metabolic genes such as lactate dehydrogenase and cytochrome C oxidase. The multidrug resistance gene sorcin also was associated with poor outcome. Exactly how these genes might contribute to tumor progression (or resistance to therapy) remains to be determined. However, as a diagnostic tool, the ability to predict tumor outcome based on gene expression profile is likely to be extremely valuable.

A study of gene expression in metastatic medulloblastoma by MacDonald et al. (2001) may offer similar advantages. About one third of patients with medulloblastoma have metastatic disease at time of diagnosis and these patients are known to have a particularly poor outcome. In fact, disseminated disease is among the most powerful predictors of poor survival in medulloblastoma patients. To find molecular markers of metastatic medulloblastoma, these researchers used Affymetrix GeneChips to analyze gene expression in 10 metastatic (M+) and 13 nonmetastatic (M0) tumors. Of the 1992 genes on these arrays, 59 (3%) showed significantly increased expression in M+ tumors and 26 (1%) showed significantly decreased expression. One prominent category of regulated genes included cell-surface and secreted proteins involved in adhesion and angiogenesis: α-catenin; α- and β-integrins; secreted protein, acidic and rich in cysteine (SPARC); tissue inhibitor of matrix metalloproteinases (TIMP1); and
TIE (a receptor for angiopoietin). Genes involved in growth factor receptor-ras-MAP kinase signaling (platelet-derived growth factor receptor alpha (PDGFRA), FGF receptor 2, the adapter protein Src homology/collagen-related-I (SHC1), the guanine nucleotide exchange factor (GEF2), rac-kinase B, protein kinase C-IIβ, JNK1) also were differentially expressed in metastatic vs. nonmetastatic tumors. Finally, metastatic tumors showed altered expression of transcription factors, including Hox A4 and A7 (which were overexpressed) and Nur77 and c-myb (which were expressed at lower levels). No differences were seen in expression of N-myc, patched, or ErbB2, which had previously been shown to be associated with medulloblastoma.

MacDonald and colleagues went on to develop an algorithm to predict whether a tumor was metastatic or nonmetastatic, based on its gene expression profile. This algorithm was able to predict tumor class with 72% accuracy and correctly categorized four new tumors. It predicted nonmetastatic tumors more accurately than metastatic ones, possibly because metastatic tumors are more heterogeneous with respect to cell type or gene expression profile. In addition to primary tumors, the authors analyzed the gene expression profile of several medulloblastoma cell lines. Although some of these lines could not be categorized, Daoy cells, a commonly used medulloblastoma cell line, clearly were categorized as metastatic.

The authors were particularly interested in elevated expression of PDGFRA and elements of the Ras-MAP kinase pathway in metastatic tumors, since PDGF has been shown to regulate angiogenesis, adhesion, and metastasis in other systems. They confirmed expression of PDGFRA protein in a panel of independent metastatic tumors as well as in Daoy cells. They then showed that soluble PDGFA causes activation of the Ras-MAPK pathway (including phosphorylation of MEK1, MEK2, and p42/p44 MAPK) in Daoy cells and enhances migration of these cells in culture. These effects could be prevented by PDGFRA-blocking antibodies and by MEK inhibitors. These findings suggested that inhibitors of the PDGFR or the Ras pathway could be used to treat metastatic medulloblastoma.

Together, these studies clearly demonstrate that gene expression profiles can be used to categorize tumors and to learn about their etiology. They also can be employed as diagnostic tools to predict prognosis and to choose appropriate treatment strategies. Finally, by highlighting signaling pathways that are dysregulated in tumors, gene expression information may yield new molecular targets or new approaches to treating tumors.

As with studies of normal development, there are a number of caveats to studying gene expression in intact tumors. First, tumors are heterogeneous and may contain a variety of cell types, including proliferating tumor cells, tumor cells that have undergone differentiation or apoptosis, blood and endothelial cells, and reactive astrocytes. Again, isolation of these cell types may yield important information about the molecular mechanisms of tumorigenesis. In
addition, it is important to note that differences in gene expression between normal and tumor cells (or between different types of tumors) may reflect differences in the cell types represented, in genes that are important for the etiology of the tumor, or genes that are expressed as a consequence of tumor growth. Distinguishing between these possibilities may not be important if the goal is to identify prognostic or diagnostic markers but will be critical for understanding the underlying mechanisms of tumorigenesis.

ACKNOWLEDGMENTS

The authors thank Audra Carroll for designing the figures and Trudy Oliver for critical review of the manuscript.

REFERENCES

Alcantara S, Ruiz M, De Castro F, Soriano E, Sotelo C 2000 Netrin 1 acts as an attractive or as a repulsive cue for distinct migrating neurons during the development of the cerebellar system. Development 127:1359–1372

Alder J, Cho NK, Hatten ME 1996 Embryonic precursor cells from the rhombic lip are specified to a cerebellar granule neuron identity. Neuron 17:389–399


Borghesani PR, Peyrin JM, Klein R, Rubin J, Carter AR, et al. 2002 BDNF stimulates migration of cerebellar granule cells. Development 129:1435–1442


Catapano LA, Arnold MW, Perez FA, Macklis JD 2001 Specific neurotrophic factors support the survival of cortical projection neurons at distinct stages of development. J Neurosci 21:8863–8872

Cogen PH, McDonald JD 1996 Tumor suppressor genes and medulloblastoma. J Neurooncol 29:103–112


el-Husseini Ae-D, Paterson JA, Shiu RP 1994 Basic fibroblast growth factor (bFGF) and two of its receptors, FGFR1 and FGFR2: gene expression in the rat brain during postnatal development as determined by quantitative RT-PCR. Mol Cell Endocrinol 104:191–200


Gunn-Moore PJ, Tavare JM 1998 Apoptosis of cerebellar granule cells induced by serum withdrawal, glutamate or beta-amyloid, is independent of Jun kinase or p38 mitogen activated protein kinase activation. Neurosci Lett 250:53–56


Hatten ME 1990 Riding the glial monorail: a common mechanism for glial-guided neuronal migration in different regions of the developing mammalian brain. Trends Neurosci 13:179–184
Kenney AM, Rowitch DH 2000 Sonic hedgehog promotes G(1) cyclin expression and sustained cell cycle progression in mammalian neuronal precursors. Mol Cell Biol 20:9055–9067
Komuro H, Rakic P 1993 Modulation of neuronal migration by NMDA receptors. Science 260:95–97

Li W, Ohlmeyer JT, Lane ME, Kalderon D 1995 Function of protein kinase A in hedgehog signal transduction and Drosophila imaginal disc development. Cell 80:553–562

Lin X, Bulleit RF 1997 Insulin-like growth factor I (IGF-I) is a critical trophic factor for developing cerebellar granule cells. Brain Res Dev Brain Res 99:234–242


Trotter J, Schachner M. 1989 Cells positive for the O4 surface antigen isolated by cell sorting are able to differentiate into astrocytes or oligodendrocytes. Brain Res Dev Brain Res 46:115–122
Ungar AR, Moon RT 1996 Inhibition of protein kinase A phenocopies ectopic expression of hedgehog in the CNS of wild-type and cyclops mutant embryos. Dev Biol 178:186–191


Wechsler-Reya RJ, Scott MP 1999 Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog [see comments]. Neuron 22:103–114


Wood KA, Dipasquale B, Youle RJ 1993 In situ labeling of granule cells for apoptosis-associated DNA fragmentation reveals different mechanisms of cell loss in developing cerebellum. Neuron 11:621–632


Paired-like Repression/Activation in Pituitary Development

LORIN E. OLSON,* † JEREMY S. DASEN,* BONG GUN JU,* JESSICA TOLKUHN,† AND MICHAEL G. ROSENFELD* ‡
*Howard Hughes Medical Institute; † Biomedical Sciences Graduate Program; ‡ Department of Medicine, University of California, San Diego, La Jolla, California 92093-0648

ABSTRACT

Pituitary gland development is controlled by signals that guide expression of specific combinations of transcription factors that dictate serial determination and differentiation events. One class of factors is paired-like homeodomain factors. Two that have been investigated are the repressor Hex1/Rpx and activator prophet of Pit-1 (Prop-1), which exert selective roles during pituitary development. The opposing actions of these factors provide one aspect of pituitary organogenesis.

I. Prophet of Pit-1 (Prop-1)

The anterior pituitary gland develops from a midline structure contiguous with the primordium of the ventral diencephalon (Watanabe, 1982; Couly and Le Douarin, 1988). It integrates signals from the peripheral and central nervous systems, regulating production and secretion of critical regulatory hormones, including growth hormone (GH), prolactin, thyroid-stimulating hormone (TSH), gonadotrophins, and adrenocorticotropic hormones by specific cell types (somatotropes, lactotropes, thyrotropes, gonadotropes, and corticotropes, respectively) (Voss and Rosenfeld, 1992). The initial organ determination involves interaction of the primordial stomadeal ectoderm and the neuroepithelium during head folding at embryonic day (e) 8.5 in the mouse (Simmons et al., 1990). The resulting development of Rathke’s pouch (RP) is characterized by expression of multiple homeodomain factors, including the LIM homeodomain factors P-Lim/Lhx3 and Lhx4 (Seidah et al., 1994; Bach et al., 1995; Sheng et al., 1996, 1997) and the OTX-related homeodomain factors P-OTX/Pitx1 (Lamonerie et al., 1996; Szeto et al., 1996) and Pitx2 (Gage et al., 1999; Lin et al., 1999; Lu et al., 1999). Restriction of expression of the paired-like homeodomain factor Hesx1/Rpx (Hermesz et al., 1996) to RP also occurs. Following proliferation and early organ expansion, different cell phenotypes arise in a distinct spatial and temporal fashion.
The Snell and Jackson allelic murine dwarf (\textit{dw,dw'}) mutations established that the product of the POU domain gene, Pit-1 — expressed in thyrotropes, somatotropes, and lactotropes (Camper et al., 1990; Li et al., 1990) — was required for initial expression of genes encoding secreted hormonal products and receptors for releasing factors. Pit-1 is required later for the continued expression of the Pit-1 gene itself and the proliferation and survival of these three cell types (Godfrey et al., 1993; Li et al., 1993; Rhodes et al., 1993). A mouse genetic defect, referred to as the Ames dwarf (\textit{df}), was mapped to chromosome 11 (Bartke, 1965; Buckwalter et al., 1991) and proved to be cell-autonomous (Buckwalter et al., 1991), resulting in a hypoplastic anterior pituitary similar to that of the Snell and Jackson dwarfs. In contrast to the complete absence of somatotropes, lactotropes, and thyrotropes in the Snell mouse, the Ames mouse pituitary gland contains between $\approx 1\%$ (Gage et al., 1996b) and 0.001\% (Andersen et al., 1995) of the normal complement of somatotropes as well as a few lactotropes and thyrotropes (Gage et al., 1996b). Failure to detect Pit-1 transcripts later in ontogeny (Andersen et al., 1995) is consistent with the hypothesis that the Ames dwarf mutation is epistatic to Pit-1.

The region surrounding the \textit{df} locus was generally mapped using a CAST/Ei $\approx$ C57BL/6J intercross and a \textit{Mus spretus} $\approx$ C57BL/6J backcross to related CA-repeat markers (Dietrich et al., 1996) to the previously identified locations of the markers Pad-1 and the interleukin (IL) cluster (IL-3, IL-4, and IL-5) (Buckwalter et al., 1991). Four markers provided two proximal and two distal loci that were used to genotype progeny of a large intercross between Cast/Ei and DF/B \textit{df/df}. The mice were genotyped and phenotypically characterized by size at 4–7 weeks. To obtain more closely linked markers, genetically directed representational difference analysis (GDRDA) (Lisitsyn et al., 1993,1994) was performed using two subsets of F$_2$ and F$_3$ \textit{df/df} mice with local recombination events, targeting 0.5-centimorgan (cM) and 0.1-cM intervals surrounding \textit{df}. This technique permitted documentation of a contig. Complementary DNA selection was used to identify the candidate gene, referred to as Prop-1 (Sornson et al., 1996). This gene encodes a \textit{paired}-like homeodomain factor and a single functional point mutation S83P that decreased DNA binding to the cognate palindromic DNA sites, decreasing the ability to regulate DNA site-dependent genes.

Expression of Prop-1 commences after initial establishment of pituitary structure on e10.5. The expression initially is observed dorsally but subsequently involves most cells in the anterior pituitary gland. The \textit{df} mutation caused dysmorphogenesis of RP at e12–e12.5, with convolution of the lumen and a failure of expression of the caudomedial Pit-1 lineage. A delayed appearance of gonadotropes was noted but corticotropes appeared as expected. Temporally extended expression of the \textit{paired}-like homeodomain repressor Hesx1/Rpx1 also was noted. More recently, subtractive hybridization techniques have suggested
targets, perhaps including components of the Wnt pathway, which may be under control of Prop-1 (Douglas et al., 2001).

II. Rpx/Hesx1 in Pituitary Development

In the anterior neural ridge (ANR), from which the pituitary is derived, a series of transcription factors initially induced in the ANR remain present following the invagination of oral ectoderm to form RP, which subsequently gives rise to the pituitary gland (reviewed in Sheng and Westphal, 1999, and Dasen and Rosenfeld, 2001). Following RP formation, opposing dorsal fibroblast growth factor/bone morphogenetic protein 4 (FGF/BMP4) and ventral Sonic hedgehog (Shh)/BMP2 gradients impart positional and proliferative signals to the pituitary progenitor field, acting to induce combinatorial patterns of transcription factor gene expression. As many mediators of the early signaling events appear capable of acting as either transcriptional activators or repressors, a critical issue has been to define the mechanisms by which these transcriptional codes positively and/or negatively influence downstream gene programs. We examined this with respect to Prop-1 and Rpx/Hesx1.

Multiple transcription factors in pituitary development act in sequential fashion to mediate the appearance of six hormone-producing cell types, which include the homeodomain factors expressed in the anterior neural plate as well as Pitx1/2, Pax6, and Hesx1/Rpx. Upon invagination of the oral ectoderm, the LIM homeodomain factor Lhx3 is induced on e9.5 in the nascent RP and is required for initial organ commitment and growth (Sheng et al., 1996). Subsequently, a second paired-like homeodomain factor, Prop-1, appears on e10.5 and is required for determination of four ventral cell types, including the Pit-1-dependent lineages (somatotropes, lactotropes, and thyrotropes) and gonadotropes (Sornson et al., 1996; Wu et al., 1998). Additional factors — including Pit-1, steroidogenic factor-1 (SF-1), and GATA-2 — are required for cell-type specification within these four lineages (reviewed in Dasen and Rosenfeld, 2001).

Based on the analysis of Ames dwarf (df) mice, which bear a hypomorphic mutation in the homeodomain of the Prop-1 gene, and more severe human mutations (Cogan et al., 1998; Fluck et al., 1998; Wu et al., 1998; Deladoey et al., 1999; Pernasetti et al., 2000), Prop-1 is necessary to activate gene programs required for ventral proliferation and determination of four cell lineages. Hesx1, in contrast, is typical of critical transcriptional repressors that regulate the development of multiple placodally derived anterior structures (Dattani et al., 1998; Martinez-Barbera et al., 2000; Thomas et al., 2001). Attenuation of Hesx1 expression in the developing pituitary coincides with the Prop-1-dependent progression of the pituitary, suggesting that temporal regulation of Hesx1 expression is essential for deployment of the Prop-1-dependent gene activation program (Gage et al., 1996b; Hermesz et al., 1996; Sornson et al., 1996).
III. Functional Antagonism Between Hesx1 and Prop-1 in Pituitary Organogenesis

The reciprocal expression and actions of Prop-1 and Hesx1 suggested that their functional interactions might regulate pituitary organogenesis. Hesx1 expression initially marks the oral ectoderm and invaginating RP but then becomes restricted to RP (excluding the ventral rostral tip), where it is maintained until e13.5. In contrast, Prop-1 expression is undetectable until e10.5–e11, becomes maximal during the ventral migration of pituitary lineage precursors beginning at e12.5, and remains detectable between e14.5–e15.5. Following e15.5, expression is maintained only at low levels (Figure 1). Prop-1 and Hesx1 can each bind effectively to a well-described palindromic site (Wilson et al., 1993) as cooperative homodimers or heterodimers, with Prop-1 acting as an activator but not as a repressor (Sornson et al., 1996). Hesx1 acts only as a repressor that can inhibit Prop-1 activation function. Both the N-terminal and homeodomain (HD) regions of Hesx1 can independently act as repressors.

Based on the requirement for Hesx1 for development of anterior organs (Dattani et al., 1998; Martinez-Barbera et al., 2000; Thomas et al., 2001), we investigated defects in pituitary development in Hesx1 gene-deleted mice. A striking, although infrequent (~5%), phenotype (Hesx1−/− mice) was a complete lack of the pituitary gland (Figure 2). Initial thickening of oral ectoderm and minimal activation of Lhx3 initially was observed at e12.5 but the pituitary gland was absent by e18.5. Thus, we tested whether Prop-1 would be capable of phenocopying the pituitary arrest observed in Hesx1−/− mice. When Prop-1 expression was targeted in transgenic mice under control of a promoter expressed only in oral ectoderm, RP, and first branchial arch (Treier et al., 1998, 2001), the mice exhibited absent anterior pituitary glands with no initial induction of Lhx3 expression but no abnormalities of developing ventral diencephalon (Figure 3).

FIG. 1. In situ hybridization showing reciprocal temporal expression of Hesx1 and Prop-1 during the initial phases of pituitary organogenesis. RP indicates Rathke’s pouch.
Thus, either absence of *Hesx1* or the premature expression of *Prop-1* can block pituitary organogenesis.

### IV. Multiple Pituitary Patterning and Growth Defects in *Hesx1* Mutant Mice

Unexpectedly, however, the highest percentage of pituitaries with *Hesx1* gene deletion resulted in the formation of multiple oral ectoderm invaginations, generating supernumerary pituitary glands with dramatic cellular overproliferation of all the hormone-producing cell types.

*Hesx1* mutants exhibited increased domains of Lhx3 and Prop-1 expression (Figure 4). The expression domains of *FGF8* and *FGF10* in the infundibulum were expanded rostrally, consistent with previous studies demonstrating that

---

**FIG. 2.** In the most-severe class of *Hesx1* mutants, the anterior pituitary (AP) is absent by embryonic day (e) 18.5. Hematoxylin and eosin (H&E) staining. Arrowhead indicates where the gland should be found.

**FIG. 3.** Analysis of *Pitx1/Prop-1* transgenic founder embryos shows the absence of the AP when *Prop-1* is temporally misexpressed. H&E staining. Arrowhead indicates where the developing organ should be found.
Lhx3 expression can be regulated by FGF signaling (Ericson et al., 1998; Treier et al., 1998), indicating that Hesx1 is required for maintaining the proper domains of FGF expression.

V. Hesx1 Recruits the Groucho-related Corepressor Transducin-like Enhancer of Split

The coregulatory apparatus that mediates Hesx1-dependent repression is unknown, although the nuclear receptor corepressor (N-CoR) has been linked to repression mediated by the homeodomain (Laherty et al., 1998; Xu et al., 1998). Another class of corepressors that has been linked to many homeodomain factors is the mammalian orthologues of the *Drosophila* protein Groucho (reviewed in Chen and Courey, 2000). Groucho is distantly related to the yeast corepressor Tup1, which binds several components of the core transcriptional apparatus — including Srb7, Srb10, Srb11, and Med6 — as well as histone deacetylases (Kuchin and Carlson, 1998; Gromoller and Lehming, 2000; Watson et al., 2000; Wu et al., 2001). Both Groucho and Tup1 contain WD40 repeats, a well-characterized protein-protein interaction domain that has been linked to repression mediated by associations with DNA binding proteins, including transcription factors and histones.

The phenotypes of Hesx1 mutants are consistent with Hesx1 acting as an *in vivo* transcriptional repressor. Domain mapping revealed that both the N-terminus and homeodomain regions exhibited repression activity. Comparing Hesx1 sequences from several vertebrate species revealed the presence of two conserved motifs outside the homeodomain. The N-terminus contains a motif...
similar to the eh1 motif originally characterized in the *Drosophila* repressor engrailed (Smith and Jaynes, 1996) and one similar to the WRPW motif found in several bovine helix-loop-helix (bHLH) proteins (Paroush et al., 1994), both of which are genetically and biochemically linked to the Groucho class of corepressors (Jimenez et al., 1997; Tolkunova et al., 1998). Affinity chromatography revealed a strong and specific interaction between Hesx1 and transducin-like enhancer of split (TLE) corepressors in the mammalian orthologs of Groucho.

Both Hesx1 and TLE1 proved to be coexpressed within RP between e9.5 and e12, with expression becoming rapidly extinguished beginning at e13.5, coincident with the appearance of anterior lobe pituitary cell types. Both the N-terminal tetramerization (Q and GP) domain and WD40 repeat region of TLE were required for interaction with Hesx1/Rpx. Full transcriptional repression required the GP and WD40 domains. Thus, as in the case of the yeast repressor Tup1, the repression and interaction domains appear to be separable, although the WD40 repeat proved to be required for both functions. Finally, repression by the N-terminal domain of Hesx1 was inhibited by αTLE-1 immunoglobulin G (IgG), indicating that it alone could serve as a component required for Hesx1-dependent repression.

In an attempt to address the functional significance of TLE1 and Hesx1 actions in pituitary development, we generated transgenic mice expressing Hesx1 under the control of a promoter (Pitx1) that targeted misexpression to ventral pituitary cell types, or a promoter alpha glycoprotein hormone subunit (αGSU) to maintain Hesx1 expression in specific cell lineages through later stages of development. In these transgenic animals, we observed only minimal phenotypes, characterized by a modest reduction in some cell lineages. In similar experiments, TLE1 expression alone was found to have minimal effects. However, when we generated transgenic animals that maintained expression of both Hesx1 and TLE1 in pituitary development, these mice were characterized by the near-complete absence of ventral anterior pituitary cell types, with an ontogeny and dysmorphogenesis quite similar to that observed in Prop-1-defective Ames dwarf mice (Gage et al., 1996a; Sornson et al., 1996). Thus, there was an absence of Prop-1-dependent cell lineages as determined by loss of expression of the Pit-1, GH, TSHβ, and αGSU genes, while expression of Prop-1 and proopiomelanocortin (POMC) were maintained (Figure 5A and B). Indeed, the ventral aspect of RP expressed a uniform field of POMC-expressing cells.

However, when specific interaction between transgenically expressed TLE1 and Hesx1 was abolished through mutation of the eh1 domain, no defects occurred in the appearance of pituitary cell types, nor was organ morphology altered (Figure 5B). Thus, specific interactions between TLE1 and Hesx1 are required for Hesx1 repression in vivo; in this case, to prevent the appearance of the Pit-1 and gonadotrope lineages.
VI. Conclusions

The sequential actions of transcriptional repressors and activators and their required coregulatory machinery on overlapping sets of gene targets serve as a gene-regulatory strategy in mammalian organogenesis, exemplified here in anterior pituitary development. Two highly related paired-like homeodomain factors — Hesx1/Rpx and Prop-1 — with distinct but overlapping patterns of expression over the entire period of pituitary organogenesis (including initial commitment, patterning, and cell-type determination) exert temporally specific roles (Figure 6). Hesx1, influenced by the actions of the linked modifier genes

FIG. 5. Analysis of Pitx1/Hesx1::Pitx1/TLE1 double-transgenic founder embryos. (A) Loss of anterior pituitary and morphological similarity to pituitary defects in Ames df/df mice, except when the Pitx1/Hesx1 transgene encodes a mutated eh1 domain, which cannot interact with TLE1. H&E staining. (B) In situ hybridization showing loss of growth hormone (GH), Pit-1, and ventral Prop-1, while expression of proopiomelanocortin (POMC) appears unaffected and Lhx3 and Prop-1 continue to be expressed in the ectoderm of RP. Asterisk indicates the dorsal pouch ectoderm.
based on the genetic background, is required for early organ commitment and cell-determination events. Hesx1, with TLE1, serves to prevent Prop-1 from initiating the program required for asymmetric division and proliferation of the Pit-1 and gonadotrope lineages. Conversely, premature expression of Prop-1 can block pituitary organogenesis, phenocopying the effects of Hesx1-gene deletion. These data suggest that the switch in binding of a paired-like homeodomain repressor for a paired-like homeodomain activator results in alteration in the expression of key target genes and prevents organogenesis. Maintained expression of Hesx1, with the TLE1 corepressor, subsequently can block the activation of Prop-1-dependent genes required for appearance of four anterior pituitary cell types. Thus, opposing actions of related repressors and activators, potentially binding to overlapping sets of gene targets, provide critical temporal control of organ development. Interestingly, later persistent expression of Prop-1 under control of the eGSU promoter caused decreased gonadotrope differentiation and increased adenomatous hyperplasia (Cushman et al., 2001). This indicates that properly extinguishing Prop-1 also may be an important later step in paired-like homeodomain-mediated organogenesis.

Hesx1 appears to contain two repressor domains, located in the N-terminal and homeodomain regions of the protein, each recruiting a distinct corepressor complex. The corepressors recruited by the N-terminus includes one mammalian Groucho orthologue, TLE1, that appears to be required for Hesx1-dependent repression. In vivo, only coexpression of both Hesx1 and TLE1 are sufficient to phenocopy the effects of Prop-1 gene deletion. The strong association between
TLE1 and Hesx1 is mediated by a highly conserved helical motif (FXLXXIL) present in the Hesx1 N-terminus, analogous to interactions of Nkx, Six, and certain Pax homeodomain factors with other TLE family members (Eberhard et al., 2000; Muhr et al., 2001).

The domains required for effective repression function appears to depend upon cellular and promoter context. Under specific circumstances, the Hesx1/Rpx homeodomain region is alone capable of mediating repression, utilizing a corepressor complex that includes mSin3A/B, histone deacetylases (HDACs) 1 and 2, but not of the high-affinity N-CoR/TBL1/HDAC3 complex (Guenther et al., 2000, Underhill et al., 2000) to the Hesx1 homeodomain. The Hesx1 N-terminal domain binding TLE permits cooperative binding of N-CoR/HDAC1/Sin3A/B to the homeodomain, a strategy that might be quite common with respect to many homeodomain repressors, and provides an additional linkage between TLEs and N-CoR/mSin3 complexes in the actions of strong homeodomain repressors.

Thus, coordinated regulation of a repressor (Hesx1) and corepressor (TLE1) serves as a determinant of organogenesis and the temporal control in pituitary cell lineage generation. The sequential repression and activation of a common set of regulatory genes may prove to be an underlying strategy in the temporal code of pituitary organ development, with initial repression required for organ commitment and proliferation, and subsequent activation for commitment of specific cell lineages.

REFERENCES


Bach I, Rhodes SJ, Pearse RV II, Heinzel T, Gloss B, Scully KM, Sawchenko PE, Rosenfeld MG 1995 P-Lim, a LIM homeodomain factor, is expressed during pituitary organ and cell commitment and synergizes with Pit-1. Proc Natl Acad Sci USA 92:2720–2724

Bartke A 1965 The response of two types of dwarf mice to growth hormone, thyrotropin, and thyroxine. Gen Comp Endocrinol 5:418–426


Couly G, Le Douarin NM 1988 The fate map of the cephalic neural primordium at the presomitic to the 3-somite stage in the avian embryo. Development 103:101–113


Gromoller A, Lehning N 2000 Sb7p is a physical and physiological target of Tup1p. EMBO J 19:6845–6852


Laherty CD, Billin AN, Lavinsky RM, Yochum GS, Bush AC, Sun JM, Mullen TM, Dave JR, Rose DW, Glass CK, Rosenfeld MG, Ayer DE, Eisenman RN 1998 SAP30, a component of the mSin3 corepressor complex involved in N-CoR-mediated repression by specific transcription factors. Mol Cell 2:33–42


Li S, Crenshaw EB III, Simmons DM, Swanson LW, Rosenfeld MG 1990 Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. Nature 347:528–533


Lisitsyn N, Lisitsyn N, Wigler M 1993 Cloning the differences between two complex genomes. Science 259:946–951


Seidah NG, Barade JC, Marcinkiewicz M, Mattei MG, Day R, Chretien M 1994 The mouse homeoprotein mLIM-3 is expressed early in cells derived from the neuroepithelium and persists in adult pituitary. DNA Cell Biol 13:1163–1180


Tolkunova EN, Fujioka M, Kobayashi M, Deka D, Jaynes JB 1998 Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. Mol Cell Biol 18:2804–2814


Dynamic Changes in Gene Expression During Human Trophoblast Differentiation

STUART HANDWERGER AND BRUCE ARONOW

Departments of Endocrinology and Molecular and Developmental Biology, Children’s Hospital Research Foundation, and Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio 45229

ABSTRACT

The genetic program that directs human placental differentiation is poorly understood. In a recent study, we used DNA microarray analyses to determine genes that are dynamically regulated during human placental development in an in vitro model system in which highly purified cytotrophoblast cells aggregate spontaneously and fuse to form a multinucleated syncytiotrophoblast that expresses placental lactogen, human chorionic gonadotropin, and other proteins normally expressed by fully differentiated syncytiotrophoblasts. Of the 6918 genes present on the Incyte Human GEM V microarray that we analyzed over a 9-day period, 141 were induced and 256 were downregulated by more than 2-fold. The dynamically regulated genes fell into nine distinct kinetic patterns of induction or repression, as detected by the K-means algorithm. Classifying the genes according to functional characteristics, the regulated genes could be divided into six overall categories: cell and tissue structural dynamics, cell cycle and apoptosis, intercellular communication, metabolism, regulation of gene expression, and expressed sequence tags and function unknown. Gene expression changes within key functional categories were tightly coupled to the morphological changes that occurred during trophoblast differentiation. Within several key gene categories (e.g., cell and tissue structure), many genes were strongly activated, while others with related function were strongly repressed. These findings suggest that trophoblast differentiation is augmented by “categorical reprogramming” in which the ability of induced genes to function is enhanced by diminished synthesis of other genes within the same category. We also observed categorical reprogramming in human decidual fibroblasts decidualized in vitro in response to progesterone, estradiol, and cyclic AMP. While there was little overlap between genes that are dynamically regulated during trophoblast differentiation versus decidualization, many of the categories in which genes were strongly activated also contained genes whose expression was strongly diminished. Taken together, these findings point to a fundamental role for simultaneous induction and repression of mRNAs that encode functionally related proteins during the differentiation process.

I. Introduction

The placenta performs many different functions, including 1) exchange of substrates, gases, and other factors between the maternal and fetal circulations; and 2) synthesis and secretion of protein and steroid hormones, growth factors, and other substances vital for regulation of maternal and fetal metabolism and
growth (Benirschke and Kaufmann, 1995). Most of these biologic actions occur at the trophoblast layer of the placental villous that is composed of two cell types: syncytiotrophoblasts and cytotrophoblasts. As shown in Figure 1, syncytiotrophoblast cells form the continuous, uninterrupted, multinucleated, epithelium-like surface of the placental villous that separates maternal blood from the villous interior. The mononuclear cytotrophoblast cells (Langhans' cells), which are located between the syncytiotrophoblast layer and its basement membrane, proliferate and fuse during trophoblast differentiation to form the overlying multinucleated syncytium (for review, see Benirschke and Kaufmann, 1995).

This chapter will present a brief overview of the differentiation of human cytotrophoblast cells to syncytiotrophoblast cells, emphasizing genes that are regulated during the differentiation process and factors that are known to regulate placental development. We then will present recent DNA microarray studies from our laboratory that have identified many previously unrecognized genes that

are dynamically regulated during cytotrophoblast differentiation and their patterns of expression during the differentiation process. The data indicate that cytotrophoblast differentiation results from a dynamic genetic program in which some genes within specific functional groups are induced, while others within the same groups are repressed. The data provide insight into molecular mechanisms critical for the induction of this differentiation process.

II. Dynamics of Placental Differentiation

A. DIFFERENTIATION SCHEME

During mammalian embryogenesis, the first differentiation event in the blastocyst leads to the formation of trophoblast cells. Prior to this developmental stage, blastomeres are totipotent and may form either trophoblast cells or cells of the inner cell mass, which develop into the embryo. The placenta stem cell divides into the villous cytotrophoblast cell, which is the precursor of the syncytiotrophoblast cell, and the invasive extravillous trophoblast cell, which is the precursor for the column of trophoblast cells that invades the myometrium of the pregnant uterus and anchors the placenta. A schematic representation of the differentiation scheme and many of the factors known to induce or inhibit cytotrophoblast differentiation is depicted in Figure 2. Several excellent reviews that summarize the differentiation process in detail have been published (Morrish et al., 1998; Cross, 2000; Knofler et al., 2001).

B. IN VITRO MODEL OF TROPHOBLAST DIFFERENTIATION

The dynamics of cytotrophoblast differentiation have been studied using primary cultures of human trophoblast cells as a model system (Kliman et al., 1986; Ringler and Strauss, 1990, Richards et al., 1994). Highly purified preparations of mononucleated cytotrophoblast cells can be isolated from preterm and term placental tissue by enzymatic dispersion. The isolated cells from term placental tissue aggregate spontaneously in culture and fuse to form a multinucleated syncytiotrophoblast that synthesizes and secretes placental lactogen (hPL), chorionic gonadotropin (hCG), and other syncytiotrophoblast-specific protein and steroid hormones (Figure 3). These in vitro changes, which recapitulate important activities accomplished by normal cytotrophoblast cells during in vivo maturation, implicate a critical relationship between the differentiation of cytotrophoblast cells into syncytiotrophoblast cells and the induction of hPL, hCG, and several other placental hormones (Hoshina et al., 1984; Boime, 1991). Isolated cells from first-trimester placentas enter the invasive pathway.
C. REGULATION OF TROPHOBLAST DIFFERENTIATION

Villus cytotrophoblast cells can be induced to differentiate in vitro by epidermal growth factor (EGF) (Maruo et al., 1995a), hCG (Shi et al., 1993), leukemia inhibitory factor (LIF) (Bischof et al., 1995), colony-stimulating factor-1 (CSF-1) (Pollard et al., 1987), granulocyte/macrophage (GM)-CSF (Garcia-Lloret et al., 1994), insulin-like growth factor-I (IGF-I) (Maruo et al., 1995b), and cyclic AMP (cAMP) (Wice et al., 1990). Transforming growth factor beta 1 (TGFβ1) has been shown to inhibit cytotrophoblast differentiation in vitro (Morrish et al., 1991) and redirect the pathway of trophoblast differentiation from a villous syncytiotrophoblast phenotype to an anchoring phenotype (Feinberg et al., 1994; Nachtigall et al., 1996). Tumor necrosis factor-alpha (TNF-α) and interferon-gamma (IFN-γ) induce trophoblast cell apoptosis in vitro, while EGF inhibits apoptosis (Morrish et al., 1991; Ho et al., 1999). The genetic program for trophoblast differentiation in the mouse has been shown to be regulated by several transcription factors, including HOXB6, HOXC5, HOXC6, HOX3E, HB24, GAX, MSX2, DLX4,
Pit-1, TF-1, TEF5, and c-Ets1 (Scott et al., 2000). However, the genetic program that controls trophoblast lineage determination and villous cytotrophoblast cell differentiation into syncytiotrophoblast cells is poorly understood. In addition, little is known about the genes that are induced and repressed during the differentiation process. Studies from our laboratory suggest that human cytotrophoblast differentiation is induced in vitro by retinoic acid receptor alpha (RARα) (Stephanou and Handwerger, 1995a), thyroid hormone receptor beta (TRβ) (Stephanou and Handwerger, 1995a), nuclear factor-interleukin-6 (NF-IL6) (Stephanou and Handwerger, 1995b), and activator protein-2 alpha (AP-2α) (Richardson et al., 2001).

FIG. 3. Expression of placental lactogen (hPL), hCGα, and hCGβ mRNAs during *in vitro* differentiation of human cytotrophoblast cells. By day 1, the isolated cytotrophoblast cells are beginning to aggregate. By day 3, many of the cells have fused to form a syncytium; by day 5, syncytialization is nearly complete. HPL, hCGα, and hCGβ mRNA levels were determined by Northern blot analysis. [Adapted with permission from Richards RG, Hartman SM, Handwerger S 1994 Human cytotrophoblast cells cultured in maternal serum progress to a differentiated syncytial phenotype expressing both human chorionic gonadotropin and human placental lactogen. Endocrinology 135:321–329. Copyright The Endocrine Society.]
D. GENE EXPRESSION DURING TROPHOBLAST DIFFERENTIATION

Several studies have begun to define groups of genes that are induced during placental differentiation. Morrish and coworkers (1996), using a subtraction cDNA library between undifferentiated and differentiating cytotrophoblast cells, identified six novel genes and four known syncytial products that are induced during differentiation (hCGα, pregnancy-specific β1 glycoprotein, 3β-hydroxysteroid dehydrogenase, plasminogen activator inhibitor (PAI) type I). Ten other genes were identified that increased during differentiation, five of which (keratin 19, calrectiulin, heat shock protein (HSP) 27, serum and glucocorticoid-regulated kinase, adrenomedullin) were not known previously to be expressed in placenta. The other induced genes included keratin 8, fibronectin, mitochondrial adenosine triphosphate (ATP) synthase, and superoxide dismutase-1. Dizon-Townson and coworkers (2000) recently found 17 of 186 random clones of a cDNA library from first-trimester placenta represented potentially novel placental genes that have not been characterized. Using differential display analysis, Xu and coworkers (1999) identified seven genes induced in BeWo choriocarcinoma cells during in vitro differentiation in response to cAMP: cytochrome p450 IIC, inosine monophosphate dehydrogenase type II, reducing agent and tunicamycin-responsive protein, and four unknown genes.

III. DNA Microarray Study of Villous Cytotrophoblast Differentiation

A. METHODS

In order to identify genes that are dynamically regulated during placental development and their expression profiles during the differentiation process, we studied gene expression in primary cultures of highly purified trophoblast cells undergoing spontaneous differentiation (Aronow et al., 2001). The cytotrophoblast cells were prepared by enzymatic dispersion and purified to > 95% homogeneity by negative selection using a monoclonal antibody to CD9. The cells were cultured in medium containing human maternal serum, since earlier studies from our laboratory demonstrated that cytotrophoblast cells cultured in medium containing human maternal serum express greater amounts of hPL, hCG, and other proteins than cells cultured in other media (Richards et al., 1994). RNA was isolated from the trophoblast cells at 12 hours after plating (time 0), when the cells were adherent to the culture dish, and at 1, 2, 3, 4, and 6 days of culture. Cy3- and Cy5-labeled probes were prepared and DNA microarrays were performed using the Incyte Human GemV microarray, which contains 6918 genes. Primary data were examined using Incyte Gemtools software and Gene-Spring software (Silicon Genetics, Redwood City, CA). Each microarray contained 192 control genes present as nonmammalian, single-gene “spikes” or
complex targets” that consisted of probe-sets that contain a pool of cellular
genes expressed in most cell types. In addition, each experimental mRNA sample
was augmented with incremental amounts of nonmammalian gene RNA, to
permit assessment of the dynamic range attained within each microarray. The
reliability of microarray quantitative data was corroborated independently
through reverse transcription-polymerase chain reaction (RT-PCR) or Northern
blot analysis of the mRNAs used in the microarray experiments as well as by
replicate analyses using additional cell and mRNA preparations.

B. IDENTIFICATION OF REGULATED GENES

DNA microarray analyses indicated that 397 of the nearly 7000 genes
exhibited robust changes during differentiation, using the criteria of ≥ 2-fold
induction or repression relative to the day 0 sample. Of these, 141 were induced
by ≥ 2-fold and 256 were repressed by ≥ 50%. The number of regulated genes
increased progressively during in vitro differentiation. Overall, 93% of the
induced genes and 73% of the repressed genes exhibited changes at two or more
time points. The 25 most-induced and repressed genes are shown in Table I. As
will be discussed, many of the most-induced genes are involved in cell adhesion
and extracellular matrix formation.

C. KINETICS OF TROPHOBLAST DIFFERENTIATION

Several different inductive and repressive kinetic patterns were associ-
ated with villous cytotrophoblast differentiation. Figure 4 shows the patterns
of gene expression that were determined by mathematical clustering of the
log-transformed normalized ratio values using hierarchical tree and K-means
algorithms. The hierarchical tree structure revealed a major division between
induced and repressed genes, with the principal variations within each major
division attributable to the delay period prior to induction or repression. Using a K-means cluster analysis, we divided the expression patterns into
nine distinct kinetic patterns (Figure 5). The patterns consisted of variable
delays prior to the induction or repression of different groups of genes, with
most gene groups exhibiting rapid initiation of their transcriptional pattern.
Pattern 1 genes were induced strongly at day 1, then either slowed in their
rate of accumulation or declined. Pattern 2 genes reached peak induction at
day 2, then leveled or declined at later days. Genes in pattern groups 3 and
4 exhibited further time delays in their induction, increasing after the
induction of the hPL and hCG genes. Patterns 5–9 were composed of
repressed genes that exhibited varying delays prior to their decline. Only
groups 4 and 6 exhibited a significant delay prior to initiation of induction or
repression. Only three genes were repressed and subsequently induced; only
two were initially activated and subsequently repressed. The distribution of
<table>
<thead>
<tr>
<th>Accession number</th>
<th>Gene name</th>
<th>Fold change</th>
<th>Accession number</th>
<th>Gene name</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>M20881</td>
<td>Pregnancy specific beta-1-glycoprotein 1</td>
<td>70.43</td>
<td>M64571</td>
<td>Microtubule-associated protein 4</td>
<td>0.32</td>
</tr>
<tr>
<td>AA216685</td>
<td>Prostate differentiation factor</td>
<td>18.88</td>
<td>M11296</td>
<td>Colony-stimulating factor 1 (macrophage)</td>
<td>0.31</td>
</tr>
<tr>
<td>D87258</td>
<td>Protease, serine, 11 (insulin-like growth factor binding)</td>
<td>17.27</td>
<td>X07696</td>
<td>Keratin 15</td>
<td>0.31</td>
</tr>
<tr>
<td>X52378</td>
<td>Carcinoembryonic antigen gene family member 6</td>
<td>16.80</td>
<td>A304657</td>
<td>Peptidylprolyl isomerase A (cyclophilin A)</td>
<td>0.31</td>
</tr>
<tr>
<td>D58798</td>
<td>Syndecan 1</td>
<td>10.23</td>
<td>D45917</td>
<td>Tissue inhibitor of metalloproteinase 3</td>
<td>0.31</td>
</tr>
<tr>
<td>AF070612</td>
<td>Estrogen sulfotransferase??</td>
<td>8.53</td>
<td>U03877</td>
<td>Epidermal growth factor-containing fibulin-like extracellular matrix protein 1</td>
<td>0.30</td>
</tr>
<tr>
<td>U39050</td>
<td>Disabled (Drosophila) homolog 2</td>
<td>7.89</td>
<td>L037652</td>
<td>Ribosomal protein S20</td>
<td>0.30</td>
</tr>
<tr>
<td>X02761</td>
<td>Fibronectin 1</td>
<td>7.87</td>
<td>A131550</td>
<td>Homo sapiens ataxia-telangiectasia group D-associated protein mRNA, complete cds</td>
<td>0.30</td>
</tr>
<tr>
<td>D00169</td>
<td>Cytochrome P450, subfamily XIA</td>
<td>7.81</td>
<td>U29953</td>
<td>Pigment epithelium-derived factor</td>
<td>0.29</td>
</tr>
<tr>
<td>AI004656</td>
<td>Placental growth factor vascular endothelial growth factor-related protein</td>
<td>7.33</td>
<td>A876532</td>
<td>Annexin A3</td>
<td>0.28</td>
</tr>
<tr>
<td>J05401</td>
<td>Creatine kinase, mitochondrial 2 (sarcomeric)</td>
<td>6.00</td>
<td>AI878886</td>
<td>Heat shock 70kD protein 5</td>
<td>0.27</td>
</tr>
<tr>
<td>AI014497</td>
<td>Leukemia inhibitory factor receptor</td>
<td>5.91</td>
<td>M57730</td>
<td>Ephrin-A1</td>
<td>0.27</td>
</tr>
<tr>
<td>J03060</td>
<td>Glucosidase, beta; acid (includes glucosylceramidase)</td>
<td>5.36</td>
<td>A307373</td>
<td>Keratin 7</td>
<td>0.27</td>
</tr>
</tbody>
</table>
the regulated genes into the nine patterns of gene expression is shown in Figure 5. Most of the induced genes exhibited patterns 1 and 2, while most of the repressed genes exhibited patterns 6 and 7.
The observation that most genes that are dynamically regulated during differentiation exhibit rapid initiation of their transcriptional pattern strongly suggests that cytotrophoblast cells are poised to enter directly the differentiation process. Activation of the cytotrophoblast-to-syncytiotrophoblast gene program results from release of one or more sequential regulatory triggers. The mechanisms for gene activation or mRNA repression during trophoblast differentiation remain to be identified. Mechanisms could include both transcriptional and posttranscriptional activation of gene expression systems already in place within the cytotrophoblast cell or through the synthesis of new gene products that impact on gene expression. The different kinetic patterns may represent the occurrence of multiple regulatory mechanisms. Repression is particularly intriguing, since specific machinery for selective mRNA decay or accelerated turnover has not been described in the trophoblast cell.

D. FUNCTIONAL CLASSIFICATION OF REGULATED GENES

The dynamically regulated genes that were annotated or partially annotated were divided into six functional categories: cell and tissue structural dynamics (110 genes), cell cycle and apoptosis (21 genes), intercellular communication (45 genes), metabolism (79 genes), regulation of gene expression (85 genes), and unknown function (two genes). The distribution of the genes in the functional categories into different patterns of expression, as determined by K-means clustering, is shown in the lower panel of Figure 5.

FIG. 4. Hierarchical tree cluster analysis of the 397 genes dynamically regulated during trophoblast differentiation. The code for the signal strength in the classification scheme is shown in the box at the bottom. [Adapted from Aronow BJ, Richardson BD, Handwerger S 2001 Microarray analysis of trophoblast differentiation: gene expression reprogramming in key gene function categories. Physiol Genomics 6:105–116.]
Each of the functional groups consisted of genes that were strongly induced and strongly repressed. For example, the category of cell and tissue structural dynamics contained six cell-adhesion genes that were induced prior to the aggregation and fusion of the cytotrophoblast cells and four adhesion genes that were induced at later times (Table II). In contrast, 24 of the 34 adhesion genes were repressed, with 10 of these belonging to pattern 6. Among the genes involved in cytoskeletal organization, 17 of 21 were repressed, with seven exhibiting pattern 6. Some categories were composed of genes that were more uniformly induced or repressed. For example, most of the genes relating to intercellular communication were induced, with nine of the polypeptide hormone genes following a pattern identical to hPL, hCGα, and hCGβ (pattern 2) (Table III). Many of the genes related to gene expression were repressed, including 27 of the 28 translation-related genes, 11 of which

FIG. 5. K-means analysis of the 397 genes dynamically regulated during trophoblast differentiation. The K-means algorithm was applied to the log2 values for the ratio of each gene’s expression. Nine patterns of gene expression were selected, with four patterns of induction and five patterns of repression. The total number of genes in each of the K-means patterns is indicated in the individual graphs. The lower panel indicates the distribution of the different functional categories of the genes within the different patterns of gene expression. [Adapted from Aronow BJ, Richardson BD, Handwerger S 2001 Microarray analysis of trophoblast differentiation: gene expression reprogramming in key gene function categories. Physiol Genomics 6:105–116.]

Each of the functional groups consisted of genes that were strongly induced and strongly repressed. For example, the category of cell and tissue structural dynamics contained six cell-adhesion genes that were induced prior to the aggregation and fusion of the cytotrophoblast cells and four adhesion genes that were induced at later times (Table II). In contrast, 24 of the 34 adhesion genes were repressed, with 10 of these belonging to pattern 6. Among the genes involved in cytoskeletal organization, 17 of 21 were repressed, with seven exhibiting pattern 6. Some categories were composed of genes that were more uniformly induced or repressed. For example, most of the genes relating to intercellular communication were induced, with nine of the polypeptide hormone genes following a pattern identical to hPL, hCGα, and hCGβ (pattern 2) (Table III). Many of the genes related to gene expression were repressed, including 27 of the 28 translation-related genes, 11 of which
<table>
<thead>
<tr>
<th></th>
<th>Induced genes</th>
<th>Repressed genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1</td>
<td></td>
<td>Pattern 5</td>
</tr>
<tr>
<td>X52378</td>
<td>Carcinoembryonic antigen</td>
<td>X63629 Cadherin 3, P-cadherin</td>
</tr>
<tr>
<td></td>
<td>gene family member 6</td>
<td></td>
</tr>
<tr>
<td>Z2555S</td>
<td>CD36 antigen (collagen type I</td>
<td>U03877 Epidermal growth factor-containing fibulin-like extracellular matrix</td>
</tr>
<tr>
<td></td>
<td>receptor)</td>
<td>protein 1</td>
</tr>
<tr>
<td>Y00636</td>
<td>CD58 antigen (lymphocyte</td>
<td>AF047433 Integrin beta 4 binding protein</td>
</tr>
<tr>
<td></td>
<td>function-associated antigen 3)</td>
<td></td>
</tr>
<tr>
<td>X02761</td>
<td>Fibronectin 1</td>
<td>U14391 Myosin IC</td>
</tr>
<tr>
<td>X51841</td>
<td>Integrin, beta 4</td>
<td></td>
</tr>
<tr>
<td>M20881</td>
<td>Pregnancy specific beta-1-</td>
<td>L36720 Bystin-like</td>
</tr>
<tr>
<td></td>
<td>glycoprotein 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pattern 2</td>
<td>X79981 Cadherin 5, VE-cadherin</td>
</tr>
<tr>
<td>AF023476</td>
<td>A disintegrin and metalloc</td>
<td>AI394286 Desmocollin 2</td>
</tr>
<tr>
<td></td>
<td>proteinase domain 12 (meltrin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>alpha)</td>
<td></td>
</tr>
<tr>
<td>AI683760</td>
<td>Semaphorin 3B</td>
<td>Z26317 Desmoglein 2</td>
</tr>
<tr>
<td>D58798</td>
<td>Syndecan 1</td>
<td>AF012023 Integrin cytoplasmic domain-associated protein 1</td>
</tr>
<tr>
<td></td>
<td>Pattern 3</td>
<td>AI521645 Integrin, alpha 2 (CD49B)</td>
</tr>
<tr>
<td>L11370</td>
<td>Protocadherin 1 (cadherin-like</td>
<td>M61199 Sperm-specific antigen 2</td>
</tr>
<tr>
<td></td>
<td>1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AL043034 Sperm surface protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L12350 Thrombospondin 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pattern 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AI310309 CD24 antigen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AL031058 Human desmoplakin mRNA, 3’ end</td>
</tr>
</tbody>
</table>
followed pattern 7. However, genes for transcription and signal transduction molecules were split more evenly, with nine of 31 and eight of 26 exhibiting upregulated behavior, respectively.

E. CATEGORICAL REPROGRAMMING

The occurrence of strongly divergent behaviors within tightly related categories strongly suggests that functional reprogramming is necessary to accomplish differentiation. The efficient execution of some biologic processes (e.g., adhesion, tissue remodeling) is best accomplished through both the induction and repression of individual genes within the specific functional category. Our recent finding that human decidualization is also characterized by strong repression of many genes and strongly divergent behaviors within related categories further suggests that categorical reprogramming may be a fundamental process in the differentiation of many cell types (Figure 6) (Brar et al., 2001). Interestingly, comparison of genes activated and repressed in the differentiation of cytotrophoblast cells and decidual fibroblasts demonstrated limited overlap in their identities, with only 81 of 569 in common. Of these overlapping genes, many were reciprocally regulated. This indicates the unique identity of each cell type.

<table>
<thead>
<tr>
<th>Induced genes</th>
<th>Repressed genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF013711 Transgelin</td>
<td></td>
</tr>
<tr>
<td>N91557 Cadherin 1, E-cadherin (epithelial)</td>
<td></td>
</tr>
<tr>
<td>X53586 Integrin, alpha 6</td>
<td></td>
</tr>
<tr>
<td>X16662 Annexin A8</td>
<td></td>
</tr>
<tr>
<td>AA876532 Annexin A3</td>
<td></td>
</tr>
<tr>
<td>M24283 Intercellular adhesion molecule 1 (CD54), human rhinovirus receptor</td>
<td></td>
</tr>
<tr>
<td>AF000652 Syndecan binding protein (syntenin)</td>
<td></td>
</tr>
<tr>
<td>X14787 Thrombospondin 1</td>
<td></td>
</tr>
</tbody>
</table>

[Patterns of gene expression were determined by K-means cluster analysis (see Figure 5). The accession number is given for each gene.]
TABLE III

Patterns of Expression of Protein Hormone Gene

<table>
<thead>
<tr>
<th>Induced genes</th>
<th>Repressed genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1</td>
<td>Pattern 5</td>
</tr>
<tr>
<td>AA033953</td>
<td>M11296</td>
</tr>
<tr>
<td>Prolactin receptor</td>
<td>Colony-stimulating factor 1 (macrophage)</td>
</tr>
<tr>
<td></td>
<td>M31159</td>
</tr>
<tr>
<td></td>
<td>Insulin-like growth factor binding protein 3</td>
</tr>
<tr>
<td>Pattern 2</td>
<td>M12783</td>
</tr>
<tr>
<td></td>
<td>Platelet-derived growth factor beta polypeptide</td>
</tr>
<tr>
<td>AF055008</td>
<td>AF055008</td>
</tr>
<tr>
<td>AF055008 granulin [Incyte PD:812141]</td>
<td></td>
</tr>
<tr>
<td>U50330</td>
<td>M54963</td>
</tr>
<tr>
<td>Bone morphogenetic protein 1</td>
<td>Human chorionic gonadotropin beta</td>
</tr>
<tr>
<td>X51801</td>
<td>Pattern 6</td>
</tr>
<tr>
<td>Bone morphogenetic protein 7 (osteogenic protein 1)</td>
<td>Y12084</td>
</tr>
<tr>
<td></td>
<td>Insulin-like growth factor binding protein 10</td>
</tr>
<tr>
<td>R68866</td>
<td>M73238</td>
</tr>
<tr>
<td>Human chorionic gonadotropin alpha</td>
<td>Ciliary neurotrophic factor receptor</td>
</tr>
<tr>
<td>M54963</td>
<td>M73238</td>
</tr>
<tr>
<td>Human chorionic gonadotropin beta</td>
<td>Ciliary neurotrophic factor receptor</td>
</tr>
<tr>
<td>Mn_001317</td>
<td>M73238</td>
</tr>
<tr>
<td>Placental lactogen</td>
<td>Human chorionic gonadotropin beta</td>
</tr>
<tr>
<td>T29543</td>
<td>Pattern 7</td>
</tr>
<tr>
<td>Luteinizing hormone beta polypeptide</td>
<td></td>
</tr>
<tr>
<td>A1004656</td>
<td>Pattern 8</td>
</tr>
<tr>
<td>Placental growth factor, vascular endothelial growth factor-related protein</td>
<td></td>
</tr>
<tr>
<td>AA216685</td>
<td>J05008</td>
</tr>
<tr>
<td>Prostate differentiation factor</td>
<td>Endothelin 1</td>
</tr>
<tr>
<td>Pattern 3</td>
<td></td>
</tr>
<tr>
<td>A1014497</td>
<td></td>
</tr>
<tr>
<td>Leukemia inhibitory factor receptor</td>
<td></td>
</tr>
<tr>
<td>Pattern 4</td>
<td></td>
</tr>
<tr>
<td>V00571</td>
<td></td>
</tr>
<tr>
<td>Corticotropin-releasing hormone</td>
<td></td>
</tr>
<tr>
<td>A1220927</td>
<td></td>
</tr>
<tr>
<td>Inhibin, alpha</td>
<td></td>
</tr>
<tr>
<td>S83157</td>
<td></td>
</tr>
<tr>
<td>Natriuretic peptide receptor B</td>
<td></td>
</tr>
</tbody>
</table>

[Patterns of gene expression were determined by K-means cluster analysis (see Figure 5). The accession number is given for each gene.]
as well as the specificity of the differentiation program that each follows. A large number of the genes in both models are related to cell and tissue structure and cell-cell interaction.

F. PROPOSED SCHEME OF VILLOUS CYTOTROPHOBLAST DIFFERENTIATION CORRELATING PATTERNS OF GENE FUNCTION TO MORPHOLOGICAL CHANGES

The occurrence of distinct temporal behaviors for gene activation and repression within different functional categories correlated well with the progressive morphological changes that underlie trophoblast differentiation (Figure 7). Thirty annotated genes were induced and 75 annotated genes were repressed during the time that the cells were aggregating and fusing. However, prior to the greatest induction of hPL and hCG, 32 annotated genes were induced with a delayed pattern similar to that of hPL, hCGα, and hCGβ. During this early stage of differentiation, many specialized adhesion genes were induced (pregnancy-specific glycoprotein 1, carinoembryonic antigen family member 6, fibronectin 1, integrin B4, C58, and C36) as well as ADAM 12. ADAM 12 is a member of the ADAM (a disintegrin and a metalloproteinase domain) family of tissue-specific fusogenic proteins that play a key role in cell-cell fusion, especially in
skeletal muscle and bone. In addition, genes for the long form of the prolactin receptor, hydroxy-delta-5-steroid dehydrogenase and endoglin, a component of the TGFβ receptor complex that binds β1 and β3 isoforms, were induced in the cytotrophoblast cells early in differentiation. Absent in melanoma 1 (AIM1), a novel, nonlens member of the βγ-crystallin superfamily associated with the
control of tumorigenicity in human malignant melanoma, also was induced prior
to cell aggregation and fusion.

Following syncytial formation, there was marked induction of genes in-
volved in intercellular communication. These included genes for nine polypep-
tide hormone genes: hPL, hCGα, hCGβ, luteinizing hormone (LH) β, granulin,
vascular endothelial growth factor (VEGF)-related protein, and three members of
the TGFβ superfamily (prostate differentiation factor, bone morphogenetic
protein-1 (BMP-1), and BMP-7, also known as osteogenic protein 1). BMP-7 and
BMP-1 play strong roles in the development and differentiative transformation of
many organ systems. Corticotropin-releasing hormone (CRH), which is known to
be a specific marker for terminally differentiated syncytiotrophoblast cells, was
not significantly induced until day 4, following inductions of hPL and hCG.
Expression of genes for insulin-like growth factor binding protein-3 (IGFBP-3)
and IGFBP-10 were repressed. Surprisingly, CSF-1, which is known to induce
trophoblast differentiation in vitro, also was repressed.

IV. Summary and Conclusions

In summary, we have identified a temporal event sequence that underlies
cytotrophoblast differentiation, based on the induction and repression of a
series of genes not recognized previously to play a role in placental de-
velopment. We have shown that cytotrophoblast-to-syncytiotrophoblast cell
differentiation is comprised of a highly dynamic gene program that signifi-
cantly affects the mRNA levels of nearly 400 of 7000 individual genes
queried. Several distinct kinetic patterns of gene induction and repression
were observed. Repression was a highly significant phenomenon but its
mechanism is unclear. Two possible mechanisms are selective degradation of
a subgroup of mRNAs or strong transcriptional repression coupled with
relatively high constitutive mRNA turnover. We have termed the phenomena
of simultaneous induction and repression of genes with similar function as
categorical reprogramming. We hypothesize that trophoblast differentiation
requires both activation and repression of a substantial number of genes. We
further postulate the existence of two classes of gene regulatory processes
that are necessary to accomplish cellular differentiation. The first is the
induction of gene products that are responsible for cell functions that were not
necessary prior to differentiation but subsequently are required for differenti-
tated cell functions. Examples would include hormone production, unique
metabolic processes, or mediators of differentiation per se. The other class
represents induction of genes that replace existing gene products with those
that cause the cell to switch structure and function. To accomplish this, we
envision that the cell must eliminate mRNAs of gene products that could
compete or interfere with the induced gene set. Successful accomplishment of
cellular differentiation may require both the induction of the effectors of the differentiated cell and dynamic reprogramming of genes within functional pathways that are critical for precursor and product cell lineages.

ACKNOWLEDGMENTS

We thank Brian Richardson, Michael Hubert, and Sarah Williams for technical assistance. Our research was supported by National Institutes of Health grant HD-07447.

REFERENCES


Boime I 1991 Human placental lactogen production is linked to the stage of trophoblast differentiation. In: Miller RK, Thiede HA, eds. Trophoblast Research. Rochester, NY: Plenum Medical; 57–60


Maruo T, Matsuo H, Otani T, Mochizuki M 1995a Role of epidermal growth factor (EGF) and its receptor in the development of the human placenta. Reprod Fertil Dev 7:1465–1470


Stephanou A, Handwerger S 1995a Retinoic acid and thyroid hormone regulate placental lactogen expression in human trophoblast cells. Endocrinology 136:933–938


Regulation of Hematopoietic Stem Cell Self-Renewal

TANISHTHA REYA

Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710

ABSTRACT

Every day, billions of new blood cells are produced in the body, each one derived from a hematopoietic stem cell (HSC). Because most mature blood stem cells have a limited life span, the ability of HSCs to perpetuate themselves through self-renewal and generate new blood cells for the lifetime of an organism is critical to sustaining life. A key problem in hematopoietic stem cell biology is how HSC self-renewal is regulated. Recent evidence suggests that signaling pathways classically involved in embryonic development — such as the Wnt signaling pathway — play an important role in regulating stem cell self-renewal. The Wnt signaling pathway has been shown to regulate stem cell fate choice in a variety of organs, including the skin, the nervous system, and the hematopoietic system. In the hematopoietic system, stimulation of hematopoietic progenitors and stem cells with soluble Wnt proteins or downstream activators of the Wnt signaling pathway leads to their expansion. Future studies focusing on the mechanism of action of the Wnt signaling pathway and its interaction with other pathways are needed to gain further insight into the regulation of stem cell self-renewal, not only in the hematopoietic system but also in a variety of other tissues.

I. Isolation and Characterization of Hematopoietic Stem Cells

Hematopoietic stem cells (HSCs) originally were identified functionally by Till and McCulloch (1961). A variety of methods have been used to isolate HSCs from the bone marrow, the primary site of hematopoiesis in the adult (Visser et al., 1984; Spangrude and Johnson, 1990; Goodell et al., 1996; Osawa et al., 1996; Adolfsson et al., 2001; Christensen and Weissman, 2001). One approach has been to utilize antibodies to a variety of HSC-associated cell surface molecules (e.g., c-kit, Sca-1, Thy1.1) and hematopoietic lineage markers such as B220, CD3, and Mac-1. Using this approach, all the HSC activity in the bone marrow has been found to reside within a small population of cells, characterized by their expression of low or undetectable levels of hematopoietic lineage markers (lin^lo\textsuperscript{-}), high levels of c-kit and Sca-1, and low levels of Thy1.1 (lin^lo\textsuperscript{-} c-kit\textsuperscript{+} Sca-1\textsuperscript{+} Thy1.1\textsuperscript{lo}) (Spangrude et al., 1988; Uchida and Weissman, 1992). Upon injection into lethally irradiated mice, these cells migrate to the appropriate microenvironments and undergo rapid expansion and differentiation. In the short term, they provide radioprotection, generating the erythroid and...
myeloid cells that are necessary for survival. Subsequently, they begin to
self-renew and generate other hematopoietic cells (Figure 1) and can maintain
steady-state hematopoiesis for the lifetime of the animal (Uchida and Weissman,
1992; Morrison and Weissman, 1994). The HSCs identified by such methods
make up 0.05% of the total bone marrow. They have been separated further into
two populations: one with long-term repopulating ability (Lin"ckit"Sca"Thy1.1"o
cells, present at a frequency of 1:10,000 in the marrow) and the other with short-term
repopulating ability (Lin"ckit"Sca"Thy1.1"o cells, present at a frequency of 1:2000
in the marrow). While both these populations can confer radioprotection, only
the long-term-HSCs can provide reconstitution beyond 10 weeks (Morrison and
Weissman, 1994).

FIG. 1. Hematopoietic stem cell development. Hematopoietic stem cells can be classified into
long-term, short-term, and multipotent progenitors, based on the extent of their self-renewal abilities.
These cells go through a number of proliferative and commitment steps to give rise to all the myeloid
and lymphoid lineages of the blood. Abbreviations: common lymphocyte progenitor (CLP); common
myeloid progenitor (CMP); granulocyte monocyte progenitor (GMP); megakaryocyte erythrocyte
progenitor (MEP). [Adapted from Reya T, Morrison SJ, Clarke MF, Weissman IL 2001 Stem cells,
II. Hematopoietic Stem Cell Self-Renewal

While the phenotypic and functional properties of HSCs have been extensively characterized (for reviews, see Morrison et al., 1995; Weissman, 2000), a fundamental question that remains is how self-renewal is regulated. In most cases, combinations of growth factors that can induce extensive proliferation are unable to prevent differentiation of HSCs in long-term cultures. Although progress has been made in identifying conditions that maintain HSC activity in culture for a brief period of time (Miller and Eaves, 1997), it has proven exceedingly difficult to identify combinations of growth factors that cause significant expansion in culture in the number of progenitors with transplantable HSC activity.

Recent in vitro and in vivo studies investigating intracellular factors that mediate self-renewal have significantly advanced our understanding of HSC development. Ectopic expression of the transcription factor HoxB4 has been shown to be able to increase the numbers of transplantable hematopoietic stem cells both in vitro and in vivo (Sauvageau et al., 1995; Antonchuk et al., 2002). Alterations in HSC growth in transgenic and knockout mice also have identified potential mediators of HSC homeostasis in vivo. Mice that overexpress the antiapoptotic gene bcl-2 (Domen et al., 2000) have increased numbers of HSCs, suggesting that such antiapoptotic signals may contribute to regulating stem cell numbers. Mice lacking the G1 checkpoint regulator p21 display a higher rate of HSC proliferation and differentiation and a lower self-renewal capacity, suggesting that p21 is required for maintaining HSC quiescence and that, in its absence, HSCs rapidly proliferate and differentiate to more-committed lineages (Cheng et al., 2000). While these are clearly important mediators of HSC development, the upstream signal that may control their activity remains unclear.

Signaling pathways classically studied in context of embryonic development — such as the Notch, Sonic hedgehog, and Wnt signaling pathways — have emerged as candidates for regulating self-renewal. The expression of constitutively active Notch1 in hematopoietic progenitors has been shown to lead to the establishment of at least some immortalized, cytokine-dependent cell lines that retain the potential to generate both lymphoid and myeloid cells in vitro and in long-term mouse reconstitution assays (Varnum-Finney et al., 2000). More recently, it was demonstrated that HSCs infected with Notch expand in vivo (Stier et al., 2002). While in many cases Notch1 expression in HSCs can lead to the rapid formation of T-cell leukemias in vivo, using RAG1−/− mice to prevent lymphocyte development allowed the effects on HSC expansion to be revealed in this study (Stier et al., 2002). Sonic hedgehog, another pathway studied classically as a regulator of embryonic development, also has emerged as a potential mediator of HSC development. Human cells highly enriched for hematopoietic progenitors (CD34+Lin−CD38−) exhibited increased self-renewal in response to
Sonic hedgehog signaling in vitro, albeit in combination with a cocktail of six other growth factors (Bhardwaj et al., 2001). The fact that both Notch and Sonic hedgehog activation has been observed to inhibit the differentiation of progenitors in many different systems by acting to maintain progenitors in an undifferentiated state (Artavanis-Tsakonas et al., 1995; Harris et al., 1997; Rowitch et al., 1999; Wechsler-Reya and Scott, 1999) certainly supports the idea that these signals may promote self-renewal in a variety of tissues.

In an effort to identify novel signaling pathways that regulate stem cell self-renewal, we analyzed the transcription factor profile of mouse bone marrow HSCs. HSCs express high levels of GATA-2 and SCL, PU.1 (Akashi et al., 2000), and lymphoid enhancer factor-1 (LEF-1) (S.H. Cheshier, K. Li, L.L. Weissman, unpublished results). The expression of LEF-1, the transcriptional mediator of Wnt signaling, in HSCs and the expression of Wnts in the bone marrow (Reya et al., 2000), together with our previous findings that Wnt signaling regulates early B-cell growth (Reya et al., 2000), suggested that Wnts may control hematopoietic stem cell self-renewal.

### III. Wnt Signaling

Wnt proteins represent a growing family of secreted signaling molecules that are expressed in diverse tissues and have been shown to influence multiple processes in vertebrate and invertebrate development (reviewed in Cadigan and Nusse, 1997). The founding member of the family, Int-1 (now called Wnt1), was identified as a protooncogene (Nusse and Varmus, 1982). Wnt proteins have been shown to regulate segment polarity in Drosophila (Siegfried and Perrimon, 1994) and axis specification in Xenopus (Moon et al., 1997). In the mouse, Wnt proteins are widely expressed. Mutations in Wnt genes result in defects in limb, somite, and axis formation and abnormal development of brain, kidney, and reproductive tract (Parr and McMahon, 1994; Monkley et al., 1996; Yoshikawa et al., 1997; Miller and Sassoon, 1998; Liu et al., 1999). In addition to its importance in normal development, dysregulation of the Wnt pathway can have potent oncogenic effects in tissues such as colon, breast, prostate, and skin (Tsukamoto et al., 1988; Korinek et al., 1997; Morin et al., 1997; Polakis, 2000).

Wnts act by binding to two types of receptor molecules at the surface of a cell (Figure 2). One is the Frizzled family of seven-pass transmembrane proteins, which contain a cysteine-rich extracellular domain that binds to Wnt proteins (Wodarz and Nusse, 1998). The second is a subset of the low-density lipoprotein receptor-related protein (or the LRP) family, specifically, LRP-5 and LRP-6 (called arrow in Drosophila) (Pinson et al., 2000; Tamai et al., 2000; Wehrli et al., 2000), a single-pass transmembrane protein. Experiments have provided evidence that both LRP-5/6 and Frizzled are needed to functionally activate the
downstream components of the canonical pathway (reviewed in Pandur and Kuhl, 2001).

In the absence of a Wnt signal, beta-catenin is associated with a large, multiprotein complex (the “destruction complex”) that includes the scaffold protein Axin and the serine/threonine kinase, glycogen synthase kinase-3 beta (GSK-3β). In this complex, beta-catenin is phosphorylated at its NH₂ terminus by GSK-3β and thereby targeted for ubiquitination and degradation by proteosomes (Cadigan and Nusse, 1997). Axin is a key component of this complex, since it acts as a scaffold to significantly enhance the ability of GSK-3β to phosphorylate beta-catenin. Binding of Wnt proteins to their receptors inhibits phosphorylation of beta-catenin by GSK-3β. This prevents beta-catenin’s degradation and results in
stabilization and accumulation of β-catenin in the cytosol (Willert and Nusse, 1998). β-catenin then translocates to the nucleus, where it binds to members of the LEF/T-cell factor (TCF) family of transcription factors. LEF/TCF proteins normally are associated with the transcriptional repressor, Groucho (Cavallo et al., 1998; Roose et al., 1998), which suppresses their activity and represses gene transcription. Binding of β-catenin relieves this repression and allows LEF/TCF factors to induce expression of the appropriate target genes (reviewed in Eastman and Grosschedl, 1999).

IV. Wnt Signaling in Stem Cell and Progenitor Cell Development

The possibility that the Wnt signaling pathway might contribute to the regulation of stem cell self-renewal in the hematopoietic system is supported by studies of nonhematopoietic progenitors and stem cells. For example, cultured human keratinocytes with higher proliferative potential have been shown to have increased levels of β-catenin, compared to those with lower proliferative capacity. Moreover, retroviral transduction of a constitutively activated form of β-catenin (mutated such that it cannot be degraded) results in increased self-renewal of epidermal progenitors/stem cells (Zhu and Watt, 1999). Transgenic mice overexpressing activated β-catenin in epidermal stem cells develop hair follicle tumors (Gat et al., 1998). Moreover, the Wnt signaling pathway may be required in the maintenance or self-renewal of stem cells of the gut, based on the finding that TCF-4-deficient mice progressively exhaust the undifferentiated progenitors in the crypts of the gut epithelium during fetal development (Korinek et al., 1998). Finally, transgenic mice that overexpress activated β-catenin on a nestin promoter display increased cell-cycle entry of neural precursors and, consequently, a larger brain with characteristics of higher mammals (Chenn and Walsh, 2002).

Wnt signaling also has been implicated in the maintenance and renewal of committed precursors of a variety of lineages. Wnt10B has been shown to maintain preadipocytes in an undifferentiated state, while inhibition of Wnt signaling allows their differentiation into adipocytes (Ross et al., 2000). Moreover, we have shown that Wnt3A can induce proliferation of B-cell precursors in a LEF-1-dependent manner and that these precursors are depleted in LEF-1-deficient mice (Reya et al., 2000). Cumulatively, these findings suggest that the Wnt signaling pathway may regulate self-renewal of stem and progenitor cells in a variety of tissues (summarized in Table I). However, Wnt signaling also has been implicated in lineage commitment of stem cells, most notably, in the skin (Huelsken et al., 2001). Thus, its influence on self-renewal is likely to be context specific and does not preclude an ability to regulate lineage commitment as well.
V. Wnt Signaling in Hematopoietic Stem Cell and Progenitor Cell Development

While the Wnt pathway has been shown to play a critical role in the development of a variety of organs and lineages, relatively little is known about its function in the hematopoietic system. Recent evidence strongly suggests that Wnt signaling has an important regulatory role in hematopoietic progenitors/stem cells during both fetal and adult development.

During fetal hematopoiesis, Wnt proteins – specifically, Wnt5A and Wnt10B — are expressed in the yolk sac and the fetal liver, both sites of hematopoiesis in the embryo (Austin et al., 1997). Moreover, hematopoietic progenitor populations were found to express Wnt10B, suggesting that Wnts may be utilized in an autocrine manner. Conditioned media containing Wnt1, Wnt5A, or Wnt10B stimulated an 11-fold expansion of fetal liver progenitors in synergy with the stem cell growth factor, SLF (Austin et al., 1997). To determine whether the cells that proliferated as a consequence of Wnt stimulation in vitro maintained immature characteristics, cells were tested in semisolid colony-forming assays and found to have increased colony-forming ability, especially of blast cell-containing colonies. This result suggested that treatment with Wnt proteins

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Cell type</th>
<th>Effect</th>
<th>Assay</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activated β-catenin</td>
<td>Keratinocytes</td>
<td>Inhibition of differentiation</td>
<td>In vitro</td>
<td>Zhu et al. (1999) Development 126:2285</td>
</tr>
<tr>
<td>Wnt10B</td>
<td>Preadipocytes</td>
<td>Inhibition of differentiation</td>
<td>In vitro</td>
<td>Ross et al. (2000) Science 289:950</td>
</tr>
<tr>
<td>Wnt3A</td>
<td>B-cell progenitors</td>
<td>Induction of proliferation</td>
<td>In vitro</td>
<td>Reya et al. (2000) Immunity 13:15</td>
</tr>
<tr>
<td>Lef-1 knockout</td>
<td>B-cell progenitors</td>
<td>Depletion of progenitors</td>
<td>In vivo</td>
<td>Reya et al. (2000) Immunity 13:15</td>
</tr>
<tr>
<td>Tcf-4 knockout</td>
<td>Intestinal stem cells</td>
<td>Depletion of stem cells</td>
<td>In vivo</td>
<td>Korinek et al. (1998) Nat Genet 4:379</td>
</tr>
<tr>
<td>Activated β-catenin</td>
<td>Follicular stem cells</td>
<td>Hair follicle and tumor formation</td>
<td>In vivo</td>
<td>Gat et al. (1998) Cell 95:605</td>
</tr>
<tr>
<td>β-catenin knockout</td>
<td>Follicular stem cells</td>
<td>Cell fate alteration</td>
<td>In vivo</td>
<td>Huelsken et al. (2001) Cell 105:533</td>
</tr>
</tbody>
</table>
retained immature functional characteristics, at least as determined by *in vitro* assays. Consistent with the effects of soluble Wnt proteins in mouse hematopoiesis are their effects in human hematopoiesis (Van Den Berg et al., 1998). The exposure of CD34<sup>+</sup>Lin<sup>−</sup> human hematopoietic progenitors to Wnt5A, which was found to be expressed in these precursors, promoted the expansion of undifferentiated progenitors in the presence of stromal cells. The presence of undifferentiated progenitors was determined by *in vitro* colony-formation assays, which revealed a 10- to 20-fold higher number of colony-forming unit (CFU)-Mix colonies, which reflect the presence of an immature population. These *in vitro* experiments clearly point to an increase in immature progenitors in response to Wnt protein stimulation; however, further experiments with *in vivo* transplants will enable a better distinction between the expansion of hematopoietic progenitors and stem cells.

In contrast to these studies, some groups have reported that soluble Wnt proteins can inhibit self-renewal of hematopoietic progenitors. In an attempt to determine the role of Wnts in the context of the stromal cell microenvironment, one group (Yamane et al., 2001) utilized Wnt3A-conditioned media in Dexter cultures and noted a decrease in the numbers of B lineage and myeloid lineage cells. These data are difficult to interpret due to the complex nature of exposing both stromal elements and hematopoietic cells to Wnt3A and the inability to distinguish direct and indirect effects. In a separate system using the quail mesodermal cell line QCE6 and avian whole-bone marrow, Wnt11 was shown to preferentially specify erythroid lineage fate in avian whole-bone marrow cultures (Brandon et al., 2000). While both these studies seem to indicate that Wnt proteins can influence lineage decisions, purified populations of stem cells will need to be tested in the context of purified Wnt proteins to make this conclusion definitively.

These studies (summarized in Table II) are difficult to compare directly due to the wide variation in the species used, the hematopoietic populations studied, the Wnt stimulus utilized, and the assays performed. The use of a highly purified, defined population of primary cells is critical to understanding the true nature of Wnt effects on HSCs, since an indirect effect through non-HSCs in a mixed population could confound interpretation of results. Moreover, due to the complexity of the Wnt pathway, it will be important to analyze the role of different components of the pathway to determine whether canonical Wnt signaling regulates HSC development. Finally, the use of a system in which the effects on HSCs can be assayed through both *in vitro* and *in vivo* approaches is critical to establishing the true effects of modulating Wnt signaling in HSCs. Many *in vitro* assays cannot distinguish between the activity of a fully uncommitted stem cell and an early committed progenitor cell (e.g., a common myeloid progenitor).

We have undertaken a set of experiments to identify methodically the role of Wnts in HSCs. We retrovirally transduced downstream components of the Wnt
pathway into highly purified mouse bone marrow HSCs and analyzed their effects by in vitro and in vivo assays. Through this approach, we found that overexpression of activated β-catenin in long-term cultures of HSCs expands the pool of HSCs, as determined by both phenotype in vitro and ability to reconstitute the hematopoietic system in vivo (T. Reya, J. Domen, D. Scherer, A.W. Duncan, K. Willert, R. Nusse, H. Weissman, unpublished results). Activated β-catenin induced cells to enter the cell cycle and grow in long-term cultures for 1 to 4 weeks, while control HSCs did not survive beyond 48 hours (Table III). These expanded HSCs also retained the functional characteristics of HSCs, following transplant into allelically distinct irradiated mice. Moreover, ectopic expression of Axin, an inhibitor that acts by degrading β-catenin, led to inhibition of HSC proliferation, increased cell death of HSCs in vitro, and reduced reconstitution in vivo. These studies demonstrated the ability of downstream components of the Wnt pathway to regulate bone marrow HSC function. Moreover, using Wnt3A protein, we found that HSCs expand dramatically to soluble Wnt protein (up to 100-fold), while maintaining the phenotypic characteristics of HSCs (K. Willert, J. Brown, E. Danenberg, R. Nusse, unpublished results). Cumulatively, our studies, together with those of others described here, indicate that Wnt signaling is important in the maintenance and self-renewal of hematopoietic stem and progenitor cells.

VI. Perspectives

The studies that have examined the role of Wnt signaling in the hematopoietic stem and progenitor cell growth and cell-fate decisions indicate that the Wnt
signaling pathway plays an important role in the hematopoietic system. These studies have raised many key questions about the mechanisms of HSC regulation. For example, in order to understand the mechanisms by which Wnt and β-catenin promote self-renewal, molecular analysis of the targets that are regulated in HSCs is needed. Moreover, given that multiple signaling pathways (e.g., Notch, Sonic hedgehog) and transcription factors (e.g., HoxB4) are emerging as regulators of HSC self-renewal, it will be important to determine if and how these signals are integrated to regulate HSC development. It is exciting to speculate that the Wnt, Notch, and Sonic hedgehog pathways may regulate similar targets genes in HSCs or may act in a hierarchical manner to regulate self-renewal. Answering these questions will enable us to understand the signals that regulate HSCs specifically and stem-cell growth in general. This will, in the long term, lead to the ability to create methods to maintain the undifferentiated state of hematopoietic stem cells in culture and have numerous practical ramifications for transplantation and regenerative therapies.

ACKNOWLEDGMENTS

The author would like to thank Andrew Duncan for help with the figures and tables and Jos Domen for helpful discussions. Some of the research mentioned in this review was supported in part by funds from the Cancer Research Institute and the Leukemia Research Foundation (awarded to TR).

REFERENCES


Christensen JL, Weissman IL 2001 Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. Proc Natl Acad Sci USA 98:14541–14546
Miller C, Sassoon DA 1998 Wnt-7a maintains appropriate uterine patterning during the development of the mouse female reproductive tract. Development 125:3201–3211


Morrison SJ, Weissman IL 1994 The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity 1:661–673


Nusse R, Varmus HE 1982 Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell 31:99–109


Spangrude GJ, Johnson GR 1990 Resting and activated subsets of mouse multipotent hematopoietic stem cells. Proc Natl Acad Sci USA 87:7433–7437


Uchida N, Weissman IL 1992 Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin-Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. J Exp Med 175:175–184


Weissman IL 2000 Stem cells: units of development, units of regeneration, and units in evolution. Cell 100:157–168


Investigation of the Transcriptional Changes Underlying Functional Defects in the Mammary Glands of Prolactin Receptor Knockout Mice

CHRISTOPHER J. ORMANDY,* MATTHEW NAYLOR,* JESSICA HARRIS,* FIONA ROBERTSON,* NELSON D. HORSEMAN,† GEOFFREY J. LINDEMAN,‡ JANE VISVADER,‡ AND PAUL A. KELLY¶

*Cancer Research Program, Garvan Institute of Medical Research, Darlinghurst 2010 Australia; †Department of Molecular and Cellular Physiology, University of Cincinnati, Cincinnati, Ohio 45267-0576; ‡The Walter and Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Melbourne, Victoria 3050 Australia; ¶Institut National de la Santé et de la Recherche Médicale (INSERM) Unité 344, Faculté de Médecine Hôpital Necker-Enfants Malades, Paris, France

ABSTRACT

Knockout (KO) mice have been created that carry null mutations of genes encoding molecules essential for prolactin (PRL) release, PRL, the receptor for prolactin (PRLR), and various members of the receptor’s signaling pathway. This allowed an in vivo genetic analysis of the role of PRL in target organ function. In PRLKO and PRLRKO mice, mammary ductal side branching was absent, terminal end bud (TEB)-like structures persisted at the ductal termini well into maturity, and no alveolar buds formed along the ductal tree. Transplants of recombined mammary glands formed from stromal and epithelial elements with and without PRLR showed normal development, while supplementation of progesterone levels in PRLKO animals restored ductal side branching. During pregnancy, PRLR heterozygous animals initially showed normal ductal and alveolar development. However, alveolar development stalled during late pregnancy, preventing successful lactation. This defect could be rescued by the loss of a single allele of the suppressor of cytokine signaling (SOCS) 1 gene. Transplants of recombined glands containing PRLRKO epithelium and wild-type (WT) stroma formed alveolar buds during pregnancy but showed no lobuloalveolar development. Recombinations of WT epithelium and PRLRKO stroma showed normal development, demonstrating that a direct action of the lactogenic hormones is confined to the epithelium, to promote lobuloalveolar development. Transcript profiling of epithelial transplants expressing or not expressing PRLR was used during early pregnancy to investigate the transcriptional response to lactogens underlying this defect. Such profiling has identified a number of genes with well-characterized roles in mammary development, in addition to a number of novel transcripts.

I. Background

The mouse mammary gland develops in four discrete stages: 1) in utero, where a rudimentary ductal structure is first produced; 2) during puberty, when
ducts elongate and bifurcate to fill the mammary fat pad; 3) during each estrus cycle, where in a strain-dependent manner, the density of ductal side branches and alveolar buds increases with each cycle; and 4) during pregnancy, where the alveolar buds that formed on the ductal tree give rise to large, lobuloalveolar structures capable of milk production. Following weaning and each estrus, the gland undergoes involution, losing most of the epithelial component gained during the preceding event. In humans, the gland involutes further, with declining ovarian function in later life.

A number of hormonal factors controlling these developmental stages have been described. Embryonic mammary epithelium appears to develop independently of ovarian and pituitary influence, although responsive to hormonal stimuli (Ceriani, 1970). Hormonal replacement in hypophysectomized, ovariectomized, and adrenalectomized mice showed that development of the mammary ducts (resembling pubertal development) was produced by a combination of estrogen and growth hormone, while further alveolar development (resembling pregnancy) required additional progesterone and prolactin (PRL) (Nandi, 1958). These hormonal combinations were shown to produce similar results in serum-free, in vitro culture of whole mammary glands, although mammary development did not achieve the extent seen in normal animals (Ichinose and Nandi, 1964; Vonderhaar, 1998).

In rodents, lobuloalveolar development during pregnancy initially depends upon increased PRL production by the pituitary, maintained by the medial preoptic area of the brain in response to cervical stimulation during copulation (Jakubowski and Terkel, 1986). Development becomes independent of the pituitary from midgestation (Collip et al., 1933), due to lactogen production by the placenta’s trophoblast cells (Thordarson and Talamantes, 1987). These lactogenic hormones may act indirectly via the modulation of endocrine organs capable of producing mammotrophic factors, such as the ovary, where lactogenic hormones provide trophic support of the corpora luteum, maintaining estrogen and progesterone production (Galosy and Talamantes, 1995), or the liver, where PRL increases output of insulin-like growth factor-1 (IGF-1) (Wennbo et al., 1997). PRL also may direct mammary development via interaction with prolactin receptors (PRLRs) in the mammary gland, which are expressed by the epithelium and the stroma. The mammary PRLR signals via the Jak2 kinase, to activate the Stat5a, mitogen-activated protein (MAP) kinase, and other signaling pathways.

Knockout (KO) mice now allow these hormonal pathways to be dissected by hormone replacement and tissue recombination in intact animals (Hennighausen and Robinson, 1998). These techniques have shown, for example, that a mammary stromal estrogen receptor (ER) is essential for ductal development (Korach et al., 1996; Cunha et al., 1997) and that a mammary epithelial progesterone receptor is required for alveolar development (Lyndon et al., 1995; Humphreys et al., 1997; Brisken et al., 1998). Genetic analysis of PRL action has been
undertaken through creation of KO mice carrying null mutations of genes that act at multiple points in the PRL pathway. For example, genes have been knocked out that encode regulators of PRL secretion (Wynick et al., 1998), PRL (Horseman et al., 1997; Vomachka et al., 2000), PRLR (Ormandy et al., 1997b; Brisken et al., 1999), and PRL signaling pathway members such as Stat5a (Liu et al., 1997). These models have provided new insight into PRL’s action in the mammary gland, which is the subject of this review.

II. Mammary Development in Utero: Formation of the Ductal Rudiment

At birth, PRLR−/− animals, both male and female, possessed a rudiment of mammary ductal architecture identical to wild-type (WT) animals (Ormandy et al., 1997a). In males, however, the nipples of PRLR−/− animals were destroyed as normal following testosterone production by the fetal testis, leaving a rudimentary ductal system embedded in the mammary fat pad in two thirds of both PRLR−/− and WT males. Both male and female ductal systems underwent normal allometric growth prior to puberty. These observations show that PRL plays no essential role during mammary organogenesis.

III. Development During Puberty and with Each Estrous Cycle: Ductal Branching and Alveolar Bud Formation

A. FAILURE OF DUCTAL SIDE BRANCHING

At the onset of puberty, terminal end buds (TEBs) formed in females of both PRLR genotypes and ductal elongation and bifurcation commenced. Examination of the mammary gland at sexual maturity (Figure 1A and B) showed that the major ducts appear at the same density in mammary glands of all genotypes but that ductal side branching failed in the PRLR−/− animals (Figure 1B). Ductal side-branch density increased with age in WT animals but the complexity achieved by 14 weeks in PRLR−/− females remained virtually unchanged for the life of the animal. A similar effect was seen in the PRLKO (PRL−/−) (Horseman et al., 1997), Stat5a KO (Teglund et al., 1998), and the galanin KO (Wynick et al., 1998). Thus, a distinction needs to be drawn between the processes that cause bifurcation and those that cause side branching.

B. PERSISTENCE OF TEB-LIKE STRUCTURES IN PRLR−/− ANIMALS

By 14 weeks of age, the TEBs of the major ducts and side branches in PRLR+/+ animals had differentiated to alveolar buds (Figure 1A and C). However, in PRLR−/− animals, a TEB-like structure persisted at the termini of most ducts (Figure 1B and D). In 20-week-old animals, these TEB-like structures
were present at most ductal termini, despite having ceased ductal elongation at the edge of the mammary fat pad, and still could be seen at the ends of some minor ducts at 32 weeks of age. With increasing age, most of the major ducts lost the TEB-like structures and the ducts ended without an apparent terminal

FIG. 1. Prolactin (PRL) indirectly influences mammary development during and following puberty. Development at age 14 weeks of the fourth inguinal gland from virgin animals (A–F). Whole mounts at low magnification (A,B) show failed ductal side branching in PRL receptor (PRLR) knockout (KO) animals (PRLR−/−), compared to wild type (WT) (PRLR+/+). Failure of alveolar bud differentiation in PRLR−/− animals is revealed by examination of the terminal ductal structures, visualized by high-power whole mounts (C,D) and hematoxylin/eosin (H&E) histology (E,F). Recombination of stroma and epithelium of differing PRLR genotypes, combined with transplantation (Tx) to a PRLR+/+ endocrine environment (G,H), reveals the side branching and alveolar bud defects to be indirect effects of PRLR loss. Supplementation for 18 days with a 25-mg progesterone pellet rescues the side-branching defect but not the failure of alveolar bud differentiation (I).
structure. Microdissection of these structures allowed viewing of whole-mount preparations at 200 × original magnification, in addition to hematoxylin/eosin (H&E)-stained sections. The persistent TEB-like structures (Figure 1D and F) showed no resemblance to the alveolar buds seen at the ductal termini of PRLR+/+ animals (Figure 1C and E). Comparison with PRLR−/− TEBs observed at 8 weeks of age (which displayed normal histology, data not shown) showed that although the typical direct contact between apical and fat cells was maintained, the TEB-like structures were much smaller, with fewer apical cell layers having no distinct cap cell layer (Figure 1F). These histological observations reflect their dormant behavior and indicate that the persistent TEB-like structures were not typical TEBs. Alveolar buds were never seen in PRLR−/− animals. An identical defect was seen in PRL−/− animals (Horseman et al., 1997).

The persistence of TEB-like structures in PRLR−/− mammary glands is intriguing. As the animals age, most of these structures become simple duct ends lacking a distinctive morphology. These aberrant structures probably result from the failure of TEBs to differentiate into alveolar buds and may represent an intermediate structure in which mitogenesis and ductal elongation have been suspended but differentiation to form an alveolar bud has not occurred.

C. OVARIAN PRLRs ARE REQUIRED FOR NORMAL PUBERTAL DEVELOPMENT OF THE MAMMARY GLANDS

Patterning of the mammary epithelium is influenced by its stroma, which is strongly inductive (Sakakura et al., 1976). This suggests that the failure of ductal side branching may be exerted by the mammary stroma. Although initial investigation failed to detect PRLR in the mammary stroma of the mouse or rat (Meister et al., 1992; Ouhtit et al., 1993a,b; Shirota et al., 1995), immunohistochemistry has found low levels in human breast cancer stroma using an antirat PRLR monoclonal (Reynolds et al., 1997) and through in situ hybridization (Mertani et al., 1998). Recently, the stroma of both rat (Camarillo et al., 2001) and mouse (Hovey et al., 2001) mammary gland has been shown to express PRLR. To determine whether PRL acts directly on the mammary epithelial or stromal cells, or indirectly via PRLRs outside the mammary gland, to influence ductal side branching and alveolar bud formation, we transplanted recombined mammary glands formed from epithelium and stroma of both PRLR genotypes, of 129SV background, into RAG1−/− recipients on the C57BL/6 genetic background. Mice homozygous for the inactivated RAG1 allele are immunocompromised and therefore able to accept allografts (Mombaerts et al., 1992). Their mammary glands show typical C57BL/6 morphology of low side branching and little alveolar bud formation. Recombined glands consisted of a fourth mammary
fat pad from a 4-week-old animal (that had been cleared of the undeveloped epithelial rudiment) into which was placed a 1-mm³ portion of mammary gland from a mature animal. The transplant was placed on the muscle wall, under the skin between the fourth and second/third mammary fat pad, via a small, midline incision. Ten weeks after surgery, the transplanted epithelium had filled the fat pad. The transplanted mammary glands, as well as an endogenous gland, were analyzed by whole-mount histology. Whole-mount analysis showed no differences in ductal side branching between any of the various combinations of epithelium and stroma (Figure 1G and H). This indicates that a mammary PRLR, whether in the epithelium or stroma, is not required for normal ductal side branching to occur. Thus, the side-branching phenotype seen in PRLR⁻/⁻ and PRLR⁻/⁻⁻⁻ animals is due to PRL action outside the mammary gland.

All transplants reproduced the highly side-branched ductal pattern of the 129Sv mouse, in contrast to the endogenous glands that showed the typical absence of side branching and alveolar bud formation seen in the C57BL/6 strain. When C57BL/6 stroma is recombined with 129Sv epithelium, the C57BL/6 pattern is reproduced. This indicates that factors within the stroma that are absent from C57BL/6 but present in 129Sv are responsible for the strain-dependent differences in ductal side branching between 129Sv and C57BL/6 (Naylor and Ormandy, 2002). Like the C57BL/6 endogenous gland, none of the transplants produced alveolar buds, suggesting that factors absent from virgin C57BL/6 but present in virgin 129Sv are required for alveolar bud formation. Pregnancy provides this factor in C57BL/6 animals (see below).

D. ROLE OF PROGESTERONE IN SIDE BRANCHING

Transplanted progesterone receptor KO mammary glands show an absence of ductal side branching (Brisken et al., 1998). Progesterone levels are reduced in PRLR⁻/⁻ animals (Clement-Lacroix et al., 1999), suggesting that progesterone is the mediator and that the ovary is the site of PRL action that controls ductal side branching. Treating 6-week-old PRL⁻/⁻ females with progesterone pellets (25 mg/21-day release) for 19 days resulted in the formation of ductal side branches but not alveolar buds (Vomachka et al., 2000). PRL supplementation by pituitary implantation resulted in normal development, as both side branches and alveolar buds formed (Vomachka et al., 2000). In PRLR⁻/⁻ mice, progesterone supplementation also rescued ductal side branching but not alveolar bud formation (Figure 1I). These experiments demonstrate that reduced progesterone levels cause the side-branching deficit in PRL⁻/⁻ and PRLR⁻/⁻ mice but not the failed formation of alveolar buds. Transplanted progesterone receptor KO mice glands do not show extensive side branches during pregnancy (Brisken et al., 1998).
IV. Development During Pregnancy

A. FAILED FIRST LACTATION IN PRLR<sup>+/−</sup> ANIMALS

Heterozygous animals on the 129 SvPas/Ola or 129Sv/C57BL/6 background mated at 6 weeks of age are unable to lactate at their first pregnancy. Aging to 20 weeks prior to mating reduces the severity of the lactational deficit (Ormandy et al., 1997b). The subsequent pregnancy results in lactational capacity sufficient for pup survival and normal growth rate but with a developmental lag of approximately 3 days due to slow onset of growth (Ormandy et al., 1997a). There is some variation in this phenotype, however, as some PRLR<sup>+/−</sup> animals can lactate at the first pregnancy, while others remain incapable of lactation following the second. This heterogeneity was not seen in genetically identical F1 animals. Back cross to the C57BL/6 background greatly increases the severity of the lactational defect, with some PRLR<sup>+/−</sup> animals incapable of lactation, despite multiple pregnancies (Gallego et al., 2001). These observations suggest the presence of a factor enhancing lactation that is absent from the C57BL/6 background.

Analysis of mammary development showed that, prior to day 15 of pregnancy, ductal elongation, branching, and the number of lobules formed were similar between PRLR<sup>+/−</sup> and PRLR<sup>+/+</sup> in response to the hormonal environment of pregnancy (data not shown). From day 15, development of PRLR<sup>+/−</sup> glands stalled and greater development of the lobuloalveoli became increasingly apparent in WT animals. At 1 day postpartum, PRLR<sup>+/−</sup> animals that were incapable of lactation showed lobules mainly at stages 2 and 3, with a few stage 4 lobules at the periphery of the fat pad (Figure 2B). In comparison, PRLR<sup>+/+</sup> animals showed a fat pad densely packed with stage 4 lobules (Figure 2A). Heterozygous animals exhibiting partial lactation showed many more stage 4 lobules than did animals unable to lactate but far fewer than seen in PRLR<sup>+/+</sup> animals (data not shown).

To determine whether the lactogenic defect in PRLR<sup>+/−</sup> mice was epithelial specific, we used epithelial explants from PRLR<sup>+/−</sup> or PRLR<sup>+/+</sup> mice transplanted into the cleared mammary fat pads of Rag1<sup>−/−</sup> recipients. Recombining WT stroma with PRLR<sup>+/−</sup> epithelium failed to rescue lobuloalveolar development during pregnancy, providing direct evidence that the defect lies in the epithelium (Figure 2C and D).

Microdissection of lobules from animals 1 day postpartum demonstrated that the lobules from PRLR<sup>+/−</sup> animals showed the formation of multiple alveoli that had failed to engorge with milk postpartum (Figure 2E and F). H&E-stained serial sections (Figure 2G and H) confirmed that the alveoli were not engorged with milk and showed that although alveoli diameters in PRLR<sup>+/−</sup> animals were smaller, they contained a similar number of epithelial cells as WT animals. Thus,
these PRLR\textsuperscript{+/–} lobules showed similar architecture to WT lobules but were not expanded by milk secretion, suggesting a failure of the final stage of functional differentiation as the cause of failed lactation. Interestingly, the alveoli at the periphery of the fat pad were most developed (Figure 2B). This is a consistent finding in both the fourth and second/third glands, suggesting that initiation of
lactation proceeds in a wave from the periphery towards the nipple in PRLR\(^{+/−}\) animals. No evidence for this was found in WT animals but it may occur too quickly to be detected.

**B. SOCS1 HAPLOINSUFFICIENCY RESCUES THE PRLR\(^{+/−}\) DEFECT**

Although the intracellular signaling pathways activated by PRL are relatively well understood, the mechanisms by which signaling is attenuated are only now being defined. Negative regulation is likely to involve protein tyrosine phosphatases as well as specific inhibitory molecules such as the suppressor of cytokine signaling (SOCS) proteins. The SOCS family of proteins acts in a classical negative-feedback loop to regulate signal transduction by a variety of cytokines (Yoshimura, 1998; Krebs and Hilton, 2000). The eight members (SOCS1–7 and CIS) of this family are characterized structurally by a C-terminal SOCS box, a central src homology 2 (SH2) domain, and an N-terminal region of variable length and limited homology (Hilton et al., 1998). Functionally, SOCS proteins interact with cytokine receptors and/or Jak kinases, thereby inhibiting activation of kinases and signal transducer and activator of transcription (STAT) proteins (Yoshimura, 1998; Krebs and Hilton, 2000).

SOCS1 (also termed JAB or SSI-1) (Endo et al., 1997; Naka et al., 1997; Starr et al., 1997) is induced in response to a broad range of cytokines and interacts with the kinase domain of Jak proteins. SOCS1-deficient mice die from a complex neonatal disease prior to weaning, involving fatty degeneration of the liver, macrophage infiltration of several organs, and multiple hematopoietic defects (Naka et al., 1998; Starr et al., 1998). This multiorgan disease can be prevented by neonatal treatment with neutralizing anti-interferon gamma (IFN) antibodies. It is absent in mice lacking both SOCS1 and IFN genes, indicating that SOCS1 is a key modulator of IFN effects (Alexander et al., 1999; Marine et al., 1999). Thus, additional disruption of the IFN gene allows the effects of SOCS1 gene deficiency to be studied in adult mice.

Since targeted deletion of the IFN gene rescues SOCS1\(^{−/−}\) mice from death at 2 weeks of age (Alexander et al., 1999; Marine et al., 1999), these double KO mice could be used to study the effect of SOCS1 deficiency on mammoipoiesis by comparing them with mice lacking IFN alone. SOCS1\(^{−/−}/\)IFN\(^{−/−}\) mice were crossed to generate females for developmental analysis, while SOCS\(^{+/+}/\)IFN\(^{−/−}\) mice were bred to generate control IFN\(^{−/−}\) females. Loss of IFN had no discernible effect on mammary development, as these mice appeared identical to WT mice at all stages.

SOCS1 deficiency led to increased development of the lobuloalveoli during pregnancy, revealed by whole-mount analysis and histological sectioning. A markedly higher density of lobuloalveoli in mammary glands from SOCS1\(^{−/−}/\)IFN\(^{−/−}\) mice was apparent from day 16 of pregnancy (Figure 2I and J).
of pregnancy, the SOCS1⁻/⁻/IFN⁻/⁻ lobulalveoli displayed dilated lumens, suggesting precocious lactation (Figure 2K and L). Milk protein levels were elevated from day 16 of pregnancy through to day 1 of lactation in SOCS1⁻/⁻/IFN⁻/⁻ mammary glands relative to those from control mice, with the maximal difference occurring at day 18 of pregnancy, confirming precocious lactation (Lindeman et al., 2001).

Stat5 was elevated in SOCS1⁻/⁻/IFN⁻/⁻ mice and higher levels of phosphorylated Stat5 were found in mammary glands at day 1 of lactation relative to controls. However, there was no apparent difference during pregnancy. Furthermore, there was no change in Stat5 DNA-binding activity during pregnancy. Interestingly, substantially less MAP kinase activity (phospho-ERK1 and phospho-ERK2) was found in SOCS1⁻/⁻/IFN⁻/⁻ mammary glands at day 18 of pregnancy and day 1 of lactation, relative to control mammary tissue. The level of total ERK1/2 remained the same, indicating that MAP kinase activity was reduced. It is not known whether SOCS1 directly influences MAP kinase activity but the diminished levels most likely reflect the differentiated state of the epithelium (Lindeman et al., 2001).

To examine whether a reduction in the level of SOCS1 might rescue signal transduction along the PRL pathway, we generated females that were heterozygous for both PRLR and SOCS1 and compared these to either SOCS1⁺/⁻, PRLR⁺/⁻, or WT littermates. We found that six of six double-heterozygous females were capable of lactation after their first pregnancy, whereas four of six PRLR⁺/⁻ females exhibited reduced lactation. Whole-mount and histological analysis of glands from the rescued mice revealed normal morphology of the lobuloalveolar structures in PRLR⁺/⁻/SOCS1⁺/⁻ mice at day 2 postpartum but dramatically reduced development in four PRLR⁺/⁻ females (Figure 2M–P). The rescue of lobuloalveolar development also was achieved in PRLR⁺/⁻/SOCS1⁺/⁻ mice on a different SOCS1 (129Sv) background. Expression of whey acidic protein (WAP) and casein milk protein genes in PRLR⁻/⁻/SOCS1⁺/⁻ mammary glands was restored to the level seen in WT glands, in contrast to the lower levels evident in PRLR⁺/⁻ mice (Lindeman et al., 2001).

C. THE PRLR IS REQUIRED IN THE MAMMARY EPITHELIUM BUT NOT THE MAMMARY STROMA FOR LOBULOALVEOLAR DEVELOPMENT

PRLR⁻/⁻ females are infertile, preventing an analysis of pregnancy on mammary development in these animals. To circumvent this problem, we made recombined mammary glands from epithelial and stromal elements of both PRLR⁻/⁻ and PRLR⁺/⁺ genotypes (from 129Sv background) prior to transfer of the recombined tissue to the abdominal wall of RAG1⁻/⁻ recipients of the C57Bl/6 background. The engrafted animals were mated 8 weeks after surgery
and the transplanted and endogenous glands analyzed 1 day postpartum. Glands formed from WT epithelium and stroma (Figure 3A) showed development identical to endogenous glands. Glands formed from PRLR−/− epithelium and

FIG. 3. PRL acts exclusively via the epithelium to direct lobuloalveolar development during pregnancy. Whole mounts of mammary glands formed by tissue recombination and Tx to Rag1−/− hosts, analyzed 1 day postpartum (A–C). Loss of PRLR in the epithelium results in failed lobuloalveolar development (B) but loss from the stroma is without effect (C), compared to WT recombinations (A). Loss of PRL from the mammary epithelium, but not the endocrine system, was produced by epithelial transplant from the PRL−/− mouse to the Rag1−/− host. This had no effect on ductal or alveolar bud development during puberty (D,E) or on lobuloalveolar development during pregnancy (F,G) but results in reduced cell proliferation 1 day postpartum (H,I). No increase in apoptosis was seen in these glands (not shown), suggesting a role for mammary-produced PRL during the onset of lactation. BrdU, bromodeoxyuridine.
WT stroma (Figure 3B) showed no alveolar development but did form side branches and alveolar buds. Identical results for PRLR−/− epithelium were obtained using the cleared fat pad technique (Brisken et al., 1999), where no beta-casein expression was seen in PRLR−/− glands. Glands formed from PRLR−/− stroma and PRLR+/+ epithelium developed normally (Figure 3C), demonstrating that stromal PRLR is not required for normal development.

These results demonstrate that an epithelial PRLR is not required for alveolar bud formation during pregnancy, in contrast to the virgin state, where PRLR−/− epithelium cannot form alveolar buds. The formation of alveolar buds in PRLR−/− epithelium at pregnancy, or in PRL−/− glands in response to a pituitary transplant, where none form during the virgin estrous cycle, indicates that PRL induces a systemic factor other than progesterone that is permissive for this event. This represents a second indirect effect of PRL on mammary development.

One of the most striking features of these experiments is the close similarity between the mammary glands of PRLRKO and progesterone receptor knockout (PRKO) mice. Both models display failed ductal side branching, persistent TEB-like structures, and alveolar bud dysgenesis in virgin glands. The similarity diverges during pregnancy. Although both show lobuloalveolar development stalled at the alveolar bud stage, PRKO glands do not side branch. Both hormones are essential for development of the alveoli at pregnancy and, clearly, PRL and progesterone cooperate to promote alveolar bud formation from the ductal epithelium postpuberty. Previous investigation has shown that these hormones interact. In mouse mammary cells (Edery et al., 1985) and human breast cancer cells (Ormandy et al., 1997c), PRL and progesterone upregulate each other’s receptors, providing a mechanism for their synergistic interaction during alveolar formation. The nature of the interaction changes in late pregnancy, as progesterone holds PRLR levels in check (Djiane and Durand, 1977), preventing lactation before parturition. This suggests a modulation of the interaction between these hormones with changing mammary epithelial cell phenotype (Vonderhaar, 1987; Vonderhaar and Biswas, 1987). PR and Stat5a also interact, to redirect transcriptional activity (Richer et al., 1998).

Progesterone treatment of PRLR−/− females following mating fully restores the deficits of preimplantation embryo development and implantation but cannot fully sustain fetal growth past midterm. Thus, only 20% of implantations survive long enough to be delivered by Caesarian section. These animals, when successfully fostered, are normal, indicating a maternal or placental defect (Binart et al., 2000). In PRL−/− animals, progesterone supplementation can fully restore fertility. The mammary glands from pregnancies maintained to term by progesterone show normal development in PRL−/− animals (Vomachka et al., 2000) and failed lobuloalveolar development in PRLR−/− animals (Binart et al., 2000). This indicates that the action of placental lactogen (PL), which can act in PRL−/−
but not PRLR−/− animals, is able to fully compensate for PRL, for mammary development and possibly for maintenance of the placenta. PL cannot act in the absence of PRLR. The simplest explanation is that PRLR is the PL receptor or is an essential component of the PL receptor (e.g., as a heterodimer with the growth hormone (GH) receptor) (Herman et al., 2000). If PL acts via the GH receptor homodimer, it must be conditional on PRLR activation.

D. MAMMARY PRL PRODUCTION IS NOT REQUIRED FOR NORMAL DEVELOPMENT BUT INFLUENCES PROLIFERATION

PRL is synthesized primarily in the anterior pituitary. Studies utilizing bromocriptine, which inhibits pituitary PRL synthesis, or pituitary isographs, which secrete large amounts of PRL, have established that endocrine PRL is largely responsible for PRL’s reported physiological functions (Freeman et al., 2000). However, PRL is also synthesized in several extrapituitary sites, including mammary epithelial cells (Lkhider et al., 1996; Escalada et al., 1997; Iwasaka et al., 2000), raising the possibility that in addition to PRL’s demonstrated direct and indirect endocrine roles, it may regulate mamnopoisis via an autocrine or paracrine mechanism. We addressed this question by comparing the development of transplanted mammary epithelium with and without a null mutation of the PRL gene, using the endpoints of morphology/histology, cell proliferation, and cell apoptosis.

Deletion of the PRL gene from the epithelium, stroma, or both did not alter ductal side branching (Figure 3D and E) or histology (not shown) in virgin mature animals. The amount of cell proliferation assessed by bromodeoxyuridine (BrdU) staining in these glands did not differ significantly (percentage of epithelial cells positive for BrdU, PRL+/+ epithelium = 1.17 ± 0.09, PRL−/− epithelium = 1.47 ± 0.47; P = 0.56). These data demonstrate that mammary PRL does not regulate mammary gland development in virgin animals and plays no detectable role in epithelial cell proliferation at this stage.

During pregnancy, normal lobuloalveolar development was observed in whole mounts of mammary glands carrying a null mutation of the PRL gene (Figure 3F and G). H&E-stained sections at day 1 postpartum showed the presence of colostrum and oil droplets, indicating normal epithelial secretory function (data not shown). Cell proliferation was assessed by measuring BrdU incorporation in transplanted mammary glands on day 1 postpartum (Figure 3H and I). Both epithelial and stromal cells were scored for BrdU staining; however, the number of proliferating stromal cells was too few to analyze. In mammary glands formed using PRL+/+ epithelium, the percentage of proliferating epithelial cells was 7.94 ± 0.20, compared to 2.82 ± 0.08 in PRL−/− epithelium-derived glands. This represents a 2.8-fold (P < 0.0001) decrease in epithelial cell proliferation in mammary glands unable to produce PRL, suggesting an autocrine
or paracrine mechanism for PRL during mammopoiesis. Apoptosis also was assessed in these transplants using the terminal-deoxy UTP nick end labeling (TUNEL) assay. As expected for this stage of development, rates of apoptosis were very low (i.e., three to five cells per section) in the epithelium of both genotypes. This prevents accurate measurement of the frequency of apoptotic cells but shows that no dramatic increase in apoptosis in PRL−/− glands has occurred.

Perplexingly, no difference in morphology was seen, despite the large difference in proliferation rates that apparently is not to be balanced by increased apoptosis. It is possible that, at this late stage of development, the mammary gland shifts from reliance on endocrine PRL to local PRL influence as part of the changes in regulatory control accompanying the shift from proliferation to milk production. Thus, the effect of an altered proliferation rate may not have had time to exert an effect on morphological endpoints. This is consistent with the role of pituitary PRL during lactation. Although hypophysectomy or treatment with a dopamine agonist will stop lactation, the level of pituitary PRL secretion falls as lactation proceeds, without diminution in milk supply (Tyson et al., 1972). Local PRL may play a role in maintaining lactation during falling pituitary PRL secretion.

Several studies to treat breast cancer using inhibitors of pituitary PRL secretion have either been unsuccessful or have produced inconsistent findings (Vonderhaar, 1999). It has been hypothesized that these studies failed because these compounds do not inhibit extrapituitary synthesis or secretion (Vonderhaar, 1999). Our data provide evidence in support of a role for mammary-produced PRL during late pregnancy but did not detect a difference in virgin glands. The number of proliferating epithelial cells is low outside pregnancy, making any effect of local PRL difficult to detect, so a subtle role for PRL outside pregnancy cannot be excluded. If such a subtle effect occurs, it may be significant across a life span.

V. Model of PRL Action in the Mammary Gland

Figure 4 summarizes PRL’s hormonal actions. PRL acts indirectly to control ductal side branching via an action in the ovary to control progesterone secretion. PRL also acts indirectly via an unknown factor (X) to regulate alveolar bud formation in virgin animals. It acts directly on the mammary epithelium to drive lobuloalveolar development during pregnancy. At this point, alveolar buds form on PRLR−/− ducts. Mammary-produced PRL may influence mammary epithelial proliferation from a stage in late pregnancy. During involution, PRL has a cell-survival action that prevents the second stage of involution. In the transgenic models of mammary cancer examined to date, the absence of PRL or Stat5a reduces the rate of tumor formation.
VI. Mammary Transcriptional Response to PRL

Having defined and described the morphological and functional defects produced in the mammary gland by the loss of PRLR signaling, we turned our attention to understanding the altered transcriptional events that underlie these defects. To discover the genes that PRL regulates during lobuloalveolar development, we utilized high-density oligonucleotide arrays (Affymetrix MGU74A GeneChips) to profile the transcriptional differences between PRLR+/+ and PRLR−−/− epithelial transplants during early pregnancy. Days 2, 4, and 6 of pregnancy were chosen to minimize the effect of the difference in epithelial content between PRLR+/+ glands, which develop normally, and PRLR−−/− glands, in which epithelial development stalls following differentiation to alveolar buds. This approach also allows detection of the early transcriptional response to PRL. We also profiled glands without epithelial transplants, to determine which genes showed an epithelial-specific pattern of expression. Glands from four to six animals were pooled to reduce nonspecific changes in gene expression due to interanimal variation and thus amplify consistent changes in gene expression due to PRLR loss. Data were analyzed using MicroArray Suite 4.0 (MAS 4.0 Affymetrix) and sorted using Excel (Microsoft). Fold
changes calculated by MAS 4.0 for a number of genes were confirmed by quantitative polymerase chain reaction (PCR) using the LightCycler (Roche). Self-organizing maps were constructed using GeneCluster (Whitehead Institute) to investigate genes that have similar patterns of transcript expression across the experiments. Genes were filtered for those that changed the most and were placed into 24 clusters representing similar patterns of change at each of the 3 days (Figure 5A). The most-interesting cluster generated from this analysis was cluster 21, containing 39 transcripts that decrease in the PRLR−/− epithelial transplants, compared to PRLR+/+ transplants at each of the 3 days. Many of these genes are known to be important for mammary gland development or are expressed in the mammary gland during pregnancy, indicating that the self-organizing maps were able to identify a functionally distinct set of genes.

Hierarchical clustering using Cluster and TreeView (Stanford) was employed to investigate the relationship between the types of mammary glands, based on their transcript profiles (Figure 5B). It revealed that PRLR+/+ epithelial transplants are distinct from PRLR−/− epithelial transplants and fat pads cleared of epithelium, the latter two groups being more similar and forming a separate branch. In both the PRLR+/+ and PRLR−/− epithelial transplants, day 6 of pregnancy was more closely related to day 4 than day 2 of pregnancy. This may represent the peak in cell proliferation in the mammary gland during early pregnancy that occurs at day 4 (Traurig, 1967), in response to a rise in progesterone and PRL levels (McCormack and Greenwald, 1974). This indicates that PRLR−/− epithelium does respond, at least in part, to the hormonal changes of early pregnancy, as also seen in the failure of development after formation of alveolar buds in the transplanted PRLR−/− glands, which are not seen in virgin animals.

The genes identified by MAS 4.0 and the GeneCluster program were sorted into groups, depending on their Gene Ontology as annotated in NetAffx (Affymetrix) (Figure 5C). This abbreviated list (to be published in full elsewhere) does not include cDNAs of unknown function or genes associated with expressed sequence tags (ESTs) and, although not discussed here, they are the focus of ongoing investigation. From extensive literature searching, we have found that many of the genes we have identified as decreasing in PRLR−/− transplants by GeneCluster and MAS 4.0 are upregulated during pregnancy, show predominantly epithelial expression, and some have been shown to be important for mammary gland development.

Four milk protein genes (casein alpha, casein beta, casein kappa, and WDNM1) were decreasing in the PRLR−/− epithelium at days 2, 4, and 6 of pregnancy. WDNM1 and β-casein are expressed during early pregnancy and increase during alveolar proliferation (Robinson et al., 1995). The decrease in these markers of epithelial differentiation in the PRLR−/− transcript profiles confirms that our model is able to detect epithelial transcripts important for
FIG. 5. Transcript profiling of PRLR+/+ and PRLR−/− epithelial transplants during early pregnancy. Mammary epithelial transplants from PRLR+/+ and PRLR−/− animals were made to cleared Rag1−/− mammary fat pads and allowed to develop for 6 weeks prior to timed mating and collection of the transplants at days 2, 4, and 6 of pregnancy. Fat pads without epithelial transplants also were collected at these times, to allow epithelial-specific genes to be identified. RNA was extracted, then pooled prior to analyses of gene expression using the Affymetrix mouse U74A chip and MAS 4.0 software. Results are presented as a self-organizing map analysis performed using the Gene Cluster program from the Whitehead Institute and hierarchical clustering using Cluster and Map Viewer from Stanford University. Cluster 21 contained genes decreasing at all time points, some members of which are shown below, including their functional annotation and fold change in expression given by MAS 4.0.
lobuloalveolar development. \(\beta\)-casein is recognized as a classical PRL-regulated gene in both \textit{in vivo} and \textit{in vitro} models. Its appearance in our list of genes confirms that our model can identify PRL-regulated genes involved in mammary epithelial differentiation.

Keratins are traditional markers of the epithelium. However, their regulation is specific to tissue type and state of differentiation of the epithelial cell. Their major role is maintaining the structure of epithelial cells. Keratins have been implicated in cell signaling, stress response, and regulation of other cellular proteins (Coulombe and Omary, 2002). Our study implicates keratins 8, 17, 18, and 19 as PRL-regulated genes required for lobuloalveolar development in the mammary gland. Two pieces of data show that the decrease in keratin expression between PRLR\(^{-/-}\) and PRLR\(^{+/+}\) epithelial transplants is not due to differences in the epithelial content of the glands. First, keratins 5 and 14, markers of myoepithelial cells, did not decrease (Neville and Daniel, 1987). Second, the proportion of expressed genes designated as epithelial specific by the stroma-only profiles remained unchanged between days 2 and 6 of pregnancy (i.e., at \(\approx 15\%\) of all expressed genes at days 2, 4, and 6 in both KO and WT profiles). Interestingly, keratin 18 protein levels are increased during lactation in the human breast, during which keratin 18 and 8 appear as intracellular aggregates rather than as components of the filamentous network seen in the nulliparous state (Michalczyk \textit{et al.}, 2001). While keratins remain useful markers of epithelial cells, their use to correct for variable epithelial content is called into question by this result.

Mammary gland development is influenced not only by systemic hormones such as PRL but also by cell microenvironment. One component of this environment is the extracellular matrix (ECM), which harbors factors that are known to regulate tissue-specific gene expression (Howlett and Bissell, 1993). We have identified a number of ECM components involved in cell adhesion as important for lobuloalveolar development, including two members of the collagen family and laminin.

A number of transcription factors were discovered to be important for PRL-stimulated development of lobuloalveolar cells. These are key molecules in the transcription response to PRL, as they act as turning points in the transcription cascade by activating the transcription of further genes.

GATA binding protein 3 (GATA3) belongs to a family of transcription factors that bind to DNA through a highly conserved zinc finger domain. GATA3 KO mouse embryos show kidney development failure early in embryogenesis. Their embryonic lethality is attributed to noradrenaline deficiency and cardiac failure (Lim \textit{et al.}, 2000). GATA-3, as well as keratin 19 transcript levels, were elevated in ER-positive breast cancer cells lines, when compared to ER-negative breast cancer cell lines. An association was found between ER and GATA-3 expression in hormone-responsive breast cancers. However, estradiol did not
induce GATA-3 expression in MCF-7 cells, suggesting a role for GATA-3 in establishing the hormone-responsive phenotype in breast cancer (Hoch et al., 1999). The level of ER expression is closely correlated with the level of PRLR expression (Ormandy et al., 1997c).

Similarly, the transcription factor activator protein-2 gamma (AP-2γ) was identified as ER factor 1 (ERF-1) in ER-positive breast cancer cell lines (deConinck et al., 1995). Its expression is limited to ER-positive cancer cell lines and is upregulated in breast cancers (Turner et al., 1998). Gel-shift assays suggest that this molecule plays a critical role in regulating ER gene transcription (McPherson et al., 1997). AP-2γ KO mice are embryonic lethal due to failure of trophodermal cell proliferation (Werling and Schorle, 2002).

Claudins are recently discovered integral membrane proteins that are major structural components of tight junction strands. Tight junctions form between epithelial cells to block transport of solutes to neighboring cells and to minimize diffusion of molecules to maintain cellular polarity. Claudin-3 is expressed mainly in the lung and liver, while claudin-7 is primarily found in the lung and kidney (Morita et al., 1999). In order to prevent diffusion of molecules across the mammary epithelium during lactation, tight junction closure is increased. This is mediated by progesterone withdrawal following parturition and requires PRLR activation (Nguyen et al., 2001). Our experiments indicate that PRL not only plays a role in the closure of tight junctions during pregnancy but also may influence the formation of these junctions by regulating transcription of their components.

Connexin-26 is a member of a large family of proteins that form similar junctions between epithelial cells, gap junctions that allow exchange of small ions and metabolites. Connexin-26 mRNA and protein expression are upregulated significantly during pregnancy and remain elevated during lactation (Tu et al., 1998). Furthermore, connexin-26 expression is confined to the alveolar epithelium, specifically localized to where adjacent alveolar cells are in contact (Locke et al., 2000). A functional binding site for the AP-2 transcription factors has been identified in the connexin-26 promoter (Tu et al., 2001), indicating that our transcript-profiling experiment may have found at least one transcription factor cascade following PRL binding to its receptor on mammary epithelial cells.

A number of extracellular ligands also were identified in our screen as transcribed following PRL action on the mammary epithelium during pregnancy. These ligands generally are important during the cell-cell communication necessary for differentiation.

Wnt4 is a member of the Wnt family of secreted glycoproteins implicated in cell-cell signaling. Wnt4 has been shown to act downstream of progesterone to induce ductal side branching during pregnancy (Braken et al., 2000). Overexpression of Wnt4 in the mammary gland by retroviral delivery resulted in
increased ductal side branching and alveolar-like structures in virgin animals, similar to those seen at day 10 of pregnancy in normal animals (Bradbury et al., 1995). This study and our transcript profiles suggest that Wnt4 may play an additional role in the PRL-stimulated development of lobuloalveolar cells.

Amphiregulin is a member of the epidermal growth factor (EGF) family that all bind to the EGF receptor. Amphiregulin can restore ductal proliferation in mammary glands of ovariectomized mice; overexpression of this ligand induces hyperplastic ducts and lobules (Kenney et al., 1996). KO studies have shown that amphiregulin is essential for ductal morphogenesis, suggesting a role in epithelial cell migration or adhesion. During pregnancy, alveoli appear small, dense, and immature in amphiregulin-deficient mammary glands, a phenotype aggravated by the loss of other EGFR ligands, EGF and transforming growth factor alpha (TGFrα) (Luetteke et al., 1999).

Calcitonin, a peptide hormone produced in the thyroid, is known to inhibit osteoclast-mediated bone resorption. Expression of calcitonin mRNA and peptide is induced during mid- to late pregnancy in the rat mammary gland, decreasing at parturition. Calcitonin receptor mRNA is induced during pregnancy, suggesting a paracrine role for this ligand in the mammary gland (Tverberg et al., 2000).

Tumor necrosis factor (TNF) (ligand) superfamily member 11 (Tnfsf11) — also known as receptor activator of nuclear factor-kappa B (NF-κB) ligand (RANKL) and osteoprotegrin ligand (OPGL) — was found to decrease in the PRLR−/− epithelium at all three timepoints. The mammary glands of the RANKL−/− mouse show a phenotype similar to PRLR−/− mammary glands — the mice are unable to lactate because lobuloalveolar cells failed to form during pregnancy. PRL was able to induce RANKL expression in the mammary gland, independent of progesterone and estrogen (Fata et al., 2000). A recent discovery has placed RANKL at the head of a signaling cascade resulting in lobuloalveolar proliferation in the mammary gland. A mouse expressing an inactivated form of the alpha subunit of IκB kinase (IKKα) had a mammary gland defect in lobuloalveolar development (Cao et al., 2001). IκB kinase also is known to activate the transcription factor NF-κB. Another KO mouse with a defect in lobuloalveolar development is that of cyclin D1 (Fantl et al., 1999), a molecule that requires activation of NF-κB for its induction. Overexpression of cyclin D1 in the IKKα-inactivated mouse restored lobuloalveolar development, confirming that cyclin D1 is a molecule acting downstream of IKKα. As RANKL was able to induce NF-κB activation in WT and not IKKα-inactivated mammary epithelial cells, it would seem that RANKL initiates the signaling cascade that results in cyclin D1-induced lobuloalveolar cell development during pregnancy (Cao et al., 2001). Our study has shown that PRL modulates RANKL expression during early pregnancy, suggesting that PRL is the master regulator of the signaling events necessary for lobuloalveolar development.
Thus, PRL acts to induce the transcription of a number of genes in the mammary epithelium that are essential for the complex interactions necessary for lobuloalveolar development and subsequent milk production and secretion. These transcriptional actions are summarized in Figure 6. PRL acts to induce transcription of genes that encode milk proteins at the final stage of differentiation. It also induces transcription of genes important for intracellular structure (keratins), extracellular structure (laminins, collagens), cell permeability (claudins, connexins), cell-cell communication (Rankl, amphiregulin, Wnt4), and the

FIG. 6. The transcriptional response to PRL. PRL acts on the mammary epithelium by binding to its receptor, activating a number of signaling cascades, including the Jak/Stat pathway and the mitogen-activated protein (MAP) kinase pathway. This results in the transcription of genes necessary for epithelial differentiation and formation of lobuloalveolar cells in the mammary gland. Transcript profiles of mammary glands capable of producing lobuloalveolar cells (PRLR^{+/+} epithelial transplants), compared to profiles of mammary glands unable to produce lobuloalveolar cells (PRLR^{-/-} epithelial transplants), identified a number of genes within the mammary epithelium whose function is known. These genes include those important for cell structure (keratins) and components of the extracellular matrix (laminin and collagen) as well as components of junctions necessary for cell permeability (connexin-26, claudin-3 and -7). A number of transcription factors were identified that act to transcribe further genes necessary for differentiation (activator protein (AP)-2 gamma, GATA-3). These genes may include extracellular ligands such as those identified by our screen (Wnt4, amphiregulin, Rankl) that act on neighboring cells to stimulate their differentiation.
continuation of differentiation (transcription factors). Thus, our transcript-profiling experiments have confirmed the morphological phenotype of the PRLR\(^{-/-}\) mouse and have shown that PRL is necessary for lobuloalveolar development in the mammary gland by allowing transcription of genes essential for a number of structures, signals, and transcription factors necessary for cell differentiation.

VII. Conclusion

The combination of epithelial/stromal recombination with transcript profiling has provided an opportunity to uncover the transcriptional program underlying the formation of lobuloalveoli in the mammary gland in response to pregnancy. Some of these genes (Figure 6; to be published elsewhere in detail) have well-established roles in mammary development, demonstrating the success of this approach and the likely importance of the novel genes that we currently are analyzing. Which of these effects are direct and which are mediated via the modulation of transcription factor activity remain to be elucidated, as does the exact temporal sequence of events. The growing understanding of the development of various cell lineages within the mammary gland will be central to fully understanding the global changes in gene expression that we now can observe. We look forward to future advances that will allow the separation of these cell types.

REFERENCES


Fantl V, Edwards PA, Steel JH, Vonderhaar BK, Dickson C 1999 Impaired mammary gland development in cycl-1(-/-) mice during pregnancy and lactation is epithelial cell autonomous. Dev Biol 212:1–11


Howlett AR, Bissell MJ 1993 The influence of tissue microenvironment (stroma and extracellular matrix) on the development and function of mammary epithelium. Epithelial Cell Biol 2:79–89


McPherson LA, Baichwal VR, Weigel RJ 1997 Identification of ERF-1 as a member of the AP2 transcription factor family. Proc Natl Acad Sci USA 94:4342–4347
Nandi S 1958 Endocrine control of mammary gland development and function in the c3h/he crg1 mouse. J Natl Cancer Inst 21:1039–1063
Naylor M, Ormandy C 2002 Mouse strain specific patterns of ductal side-branching are elicited by stromal factors. Dev Dynam 225:101–105

Robinson GW, McKnight RA, Smith GH, Hennighausen L 1995 Mammary epithelial cells undergo secretory differentiation in cycling virgins but require pregnancy for the establishment of terminal differentiation. Development 121:2079–2090


Microarray Analysis and Identification of Novel Molecules Involved in Insulin-like Growth Factor-1 Receptor Signaling and Gene Expression

JOELLE DUPONT, SANDRA E. DUNN, J. CARL BARRETT, AND DEREK LEROITH

*Physiologie de la Reproduction et des Comportements, Unité Mixte de Recherches 6073, Institut National de la Recherche Agronomique-Centre National de la Recherche Scientifique, Université F. Rabelais, Tours 37380 Nouzilly, France; †British Columbia Research Institute for Children’s and Women’s Health, Vancouver, British Columbia, Canada; ‡Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892-2440; ¶Clinical Endocrinology Branch, National Institutes of Health, Bethesda, Maryland 20892-1758

ABSTRACT

The insulin receptor (IR) and the insulin-like growth factor-1 receptor (IGF-1R) are members of the same subfamily of receptor tyrosine kinases. The two receptors phosphorylate many of the same substrates and activate the same signaling modules, including the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3’ kinase (PI3K) signaling pathways. Although the IR and IGF-1R share some redundant functions in metabolism, cell growth, differentiation, and apoptosis, they also exhibit distinct physiological roles. Some of these may be due to differences in tissue distribution, receptor structure, formation of hybrid receptors, or mechanisms of ligand binding. However, the divergent effects of insulin and IGF-1 also may be explained by specificity in the intracellular signals generated by insulin and IGF-1. In particular, the IR and IGF-1R are capable of triggering their own biological responses by using specific or preferential substrates, molecular adapters, or signaling pathways. In a recent study, we used cDNA microarray analysis to identify genes differentially regulated by insulin and IGF-1. Mouse NIH-3T3 fibroblasts expressing either the wild-type human IGF-1R or IR were stimulated with either IGF-1 or insulin, respectively. We identified 39 genes differentially regulated by insulin and IGF-1. Most of these genes had not been reported previously to be responsive to insulin or IGF-1. The genes induced by IGF-1 generally were involved in mitogenesis or differentiation, while the genes found to be induced by insulin did not conform to any particular category. In a separate study, immortalized breast epithelial cells were stimulated with IGF-1 and a cDNA microarray analysis was used to generate a profile of IGF-1-regulated genes. A number of genes known to be involved in angiogenesis were found to be regulated by IGF-1. These results strongly suggest that this technology may be extremely useful in identifying groups of genes that are specifically regulated by different ligands and their activated receptors.

I. Introduction

Insulin and insulin-like growth factor-1 (IGF-1) are peptide hormones that are homologous in primary structure but differ in their physiological effects.
Insulin, produced by the beta (β) cells of the pancreas, stimulates the uptake of glucose and amino acids, inhibits gluconeogenesis, and promotes lipogenesis. IGF-1 is involved primarily in cell growth, survival, apoptosis, and differentiation. Insulin and IGF-1 mediate their biological effects by binding to their respective receptors, the insulin receptor (IR) and the IGF-1 receptor (IGF-1R). Although IR and IGF-1R are structurally and functionally similar, targeted gene knockouts in mice showed that they have both overlapping and distinct physiological roles (Nakae et al., 2001). One of the major areas of interest in this field is to understand how the specificity of IR and IGF-1R signaling is defined. In this review, we briefly compare and contrast the IR and IGF-1R signaling pathways and discuss various mechanisms that could explain the divergent physiological functions mediated by the two receptors. We then describe recent experiments using cDNA microarray analysis that have identified specific differences at the level of gene expression.

II. Structure of the Insulin and IGF-1 Receptors

The IR and IGF-1R are both comprised of two extracellular alpha (α) subunits containing ligand-binding sites and two transmembrane β subunits transmitting the ligand-induced signal (Yarden and Ullrich, 1988). More specifically, IGF-1R and IR αβ subunits consist of three domains: 1) a juxtamembrane domain, with motifs required for recruiting the major signaling adapter proteins; 2) a tyrosine kinase domain, essential for catalytic activity of the receptor; and 3) the carboxyl-terminal domain, which has several important residues for IGF-1R and IR signaling (Figure 1). As a consequence of this high level of homology, hybrid receptors, comprised of an insulin αβ-hemireceptor and an IGF-1 αβ-hemireceptor, can form in tissues and cultured cells expressing both the IR and the IGF-1R (Federici et al., 1997). Such hybrid receptors may play a role in the divergent actions of insulin and IGF-1.

A. EXTRACELLULAR (LIGAND-BINDING) DOMAIN

Despite the structural similarities between IGF-1 and insulin, the IR and IGF-1R have 100- to 1000-fold higher binding affinity for their cognate ligands. The α subunits have been shown to confer ligand-binding specificity (Schumacher et al., 1991). Some studies using chimeric receptors have shown that the high affinity of the IR for insulin is determined by regions adjacent to the cysteine-rich domain (Gustafson and Rutter, 1990; Schumacher et al., 1991). On the other hand, the high-affinity IGF-1 binding by the IGF-1R is determined by its cysteine-rich domain within the α subunit.
B. CYTOPLASMIC DOMAIN

Structural differences in the cytoplasmic domain of the β subunits of the IR and IGF-1R may contribute to the divergence of these two signaling pathways. The highest degree of homology between the two receptors is found within the tyrosine kinase domain (about 84%), whereas the region of greatest divergence between the IR and IGF-1R is found within the juxtamembrane domain (about 61%) and the carboxyl-terminal domain (about 56%) (Ullrich et al., 1986; Ullrich and Schlessinger, 1990). Chimeric receptors consisting of the ligand-binding domain of IR and the cytoplasmic domain of IGF-1R functioned more like the IGF-1R than the IR (Lammers et al., 1989). Similarly, chimeric IGF-1R containing the carboxyl-terminal β subunit domain of the IR more closely resembled the IR than the IGF-1R (Tartare et al., 1994). To eliminate interactions of ligands with endogenous receptors, other chimeras were generated in which...
the extracellular portion of the neurotrophin receptor was fused to the intracellular portions of IR or IGF-1R (Siddle et al., 2001). These chimeric molecules were stably expressed in 3T3-L1 fibroblasts (Kalloo-Hosein et al., 1997) or 3T3-L1 adipocytes (Urso et al., 1999, 2001) at levels comparable to those of endogenous IR or IGF-1R and activated by nerve growth factor (NGF). The TrkC-IR chimeric receptor was more effective in stimulating physiologically relevant metabolic responses, whereas the TrkC-IGF-1R was more effective in promoting mitogenesis (Urso et al., 1999). Thus, the intracellular domains of the IR and IGF-1R are likely to mediate at least part of the observed receptor specificity.

III. Signal Transduction via IR and IGF-1R

A. COMMON SIGNALING PATHWAYS

Many of the intracellular signaling events mediated by activation of the IR and IGF-1R are remarkably similar (White, 1994; Cheatham and Kahn, 1995; LeRoith et al., 1995) (Figure 2). Some of the shared substrates that become phosphorylated by the IGF-1R and IR include members of the insulin receptor substrate (IRS) family of proteins (IRS-1, -2, -3, and -4) (Sun et al., 1991; Lavan and Lienhard, 1993; Patti et al., 1995; Fantin et al., 1998), Gab-1 (Winnay et al., 2000), and Shc (Pellicci et al., 1992). Upon stimulation by insulin or IGF-1, tyrosine-phosphorylated IRS and Shc proteins form signaling complexes between phosphotyrosine-containing binding motifs (YXXM) and Src homology 2 (SH2) domains found in molecules such as growth factor receptor binding-2 protein (Grb2) (Lowenstein et al., 1992; Skolnik et al., 1993) and the p85 regulatory subunit of the phosphatidylinositol 3′ kinase (PI3K) (Backer et al., 1992). The phosphotyrosine residues on IRS-1 also form docking sites for other signaling molecules, including Syp (SHPTP2) (Xiao et al., 1994), Fyn (Sun et al., 1996), Nck (Lee et al., 1993), and Crk (Beitner-Johnson et al., 1996).

By binding to Grb2, IRS proteins couple the IR and IGF-1R to the Ras/mitogen-activated protein kinase (MAPK) pathway. This pathway regulates cell growth, differentiation, and proliferation in response to insulin and IGF-1 (Blenis, 1993; Crews and Erikson, 1993). Various protein tyrosine phosphatases can regulate the activities of the IR and IGF-1R signaling systems.

B. SPECIFICITY

1. Proximal Substrates

To understand the mechanisms involved in the distinct physiological functions of insulin and IGF-1, some investigators searched for specific substrates for
IR or IGF-1R. Najjar and coworkers identified pp120, a plasma membrane glycoprotein, which is a substrate for the IR but not for the IGF-1R (Najjar et al., 1997; Soni et al., 2000). Phosphorylation of pp120 is required for its function in insulin endocytosis (Formisano et al., 1995), bile acid transport (Sippel et al., 1994), tumor suppression (Kleinerman et al., 1995), and its inhibitory effect on the mitogenic actions of insulin (Soni et al., 2000). Interestingly, when the carboxyl-terminus of the IGF-1R is replaced by an equivalent region of the IR, the chimeric IGF-1R then can bind to and phosphorylate pp120, decreasing its effect on cell growth (Soni et al., 2000). Mutation of the tyr1316 in the IR, which is not conserved in the IGF-1R, abrogates the insulin-induced tyrosine phosphorylation of pp120 and its ability to suppress the mitogenic action of insulin (Soni et al., 2000).

Some of the other substrates of the IR and IGF-1R are differentially phosphorylated in response to IGF-1 or insulin, indicating that they may mediate specific effects for both ligands. The molecular adapter Grb14 binds specifically to the regulatory kinase loop of the IR and inhibits catalytic activity (Kasus-Jacobi et al., 1998). It recently was shown that Grb14 is three to 10 times less effective at inhibiting the catalytic activity of the IGF-1R than the IR (Bereziat et al., 1998).
et al., 2002). Rother and coworkers showed that the specificity of signaling may be explained by the preferential use of different substrates by the IR and IGF-1R (Rother et al., 1998). In particular, the IR was coupled preferentially to IRS-2, whereas the IGF-1R was coupled preferentially to IRS-1. This conclusion was confirmed by ablation of the IRS-1 and IRS-2 genes in mice (Araki et al., 1994; Tamemoto et al., 1994; Withers et al., 1998).

In Chinese hamster ovary (CHO) cells stably expressing either the human IR or IGF-1R, it was shown that there are differences in the complement of SH2-containing proteins recruited to IRS-1 by the two receptors (Amoui et al., 2001). In particular, the IGF-1R appears to couple IRS-1 preferentially to Grb2, whereas the IR appears to couple IRS-1 preferentially to the p85 subunit of PI3K (Amoui et al., 2001). In other recent studies, Olefsky and coworkers showed that both the IGF-1R and IR can function as G protein-coupled receptors and engage different G-protein partners. The IGF-1R utilizes Gαi, whereas the IR does not (Dalle et al., 2001). In contrast, the IR signals through Goq/11, whereas the IGF-1R does not (Imamura et al., 1999; Dalle et al., 2001). Using the two-hybrid system, Grb10 was found to associate preferentially with the IR in mouse fibroblasts expressing either the IR or IGF-1R (Laviola et al., 1997). Using the same technique, the protein 14-3-3β was found to bind to the IGF-1R but not to the IR (Furlanetto et al., 1997).

Recently, Ligensa and coworkers identified a new PDZ (postsynaptic density protein-95, disc large, zonula occlusions-1) domain-containing protein (IGF-1 receptor interacting protein-1, IIP-1) that interacts with the C-terminal tail of the IGF-1R but not the IR (Ligensa et al., 2001). Furthermore, the most distal three amino acids in the C-terminal tail of the IGF-1R appear to be crucial for the interaction of IIP-1 with IGF-1R. Indeed, a mutated IR tail carrying the terminal three amino acids of the IGF-1R is able to bind to IIP-1, whereas mutating any of the terminal three amino acids in the IGF-1R tail to the corresponding three amino acids in the IR abolishes the interaction with IIP-1 (Ligensa et al., 2001). Overexpression of IIP-1 in MCF-7 cells does not affect either IGF-1-dependent proliferation or IGF-1-mediated protection from apoptosis but significantly reduces cell motility (Ligensa et al., 2001). Hermanto and coworkers also identified a novel IGF-1R-interacting molecule called RACK1 (Hermanto et al., 2002). RACK1 associates specifically with the IGF-1R but not with the IR, both in yeast and in HEK293T and NIH-3T3 cells overexpressing either the IGF-1R or the IR (Hermanto et al., 2002). RACK-1 is involved in IGF-1R-mediated regulation of cell growth and transformation (Hermanto et al., 2002). However, the interaction of RACK1 with the IGF-1R but not the IR seems to be specific to certain cell types. Indeed, Kiely and coworkers showed that endogenous RACK-1 could interact with both the endogenous IGF-1R and IR in Chinese ovary siemens (COS) cells (Kiely et al., 2002). Nevertheless, these different
receptor-specific adaptor proteins, particularly IIP-1 and RACK1, might contribute to the biological specificity of the two hormones.

2. Signaling Pathways

Some evidence suggests that the IR and IGF-1R may phosphorylate the same substrates but use different signaling pathways to mediate the same or different biological effects. For example, insulin induces the expression of vascular endothelial growth factor (VEGF) via the PI3K/Akt pathways in NIH-3T3 cells overexpressing the human IR, whereas IGF-1 induces VEGF expression via the mitogen extracellular kinase (MEK)/MAPK pathway in NIH-3T3 cells overexpressing human IGF-1R (Miele et al., 2000). In rat hepatic stellate cells, insulin and IGF-1 both stimulate cellular proliferation. However, both PI3K and extracellular signal-regulated kinase (ERK) are involved in IGF-1-induced mitogenesis, whereas insulin stimulated mitogenesis through a PI3K-dependent and ERK-independent pathway (Svegliati-Baroni et al., 1999). Interestingly, glycogen synthesis was more effectively stimulated by the IR than by the IGF-1R, although both receptors mediated similar activation of the Akt/protein kinase B (PKB) protein kinase in hepatocytes and in 3T3-L1 fibroblasts (Park et al., 1999). The insulin-specific stimulation of glycogen synthesis appears to involve a rapamycin-sensitive pathway in hepatocytes (Park et al., 1999).

To explain the specificities in the function of the IR and IGF-1R, some investigators suggested that downstream kinases may mediate the specific effects of the IR vs. the IGF-1R. For example, Nakae et al. reported that the transcription factor forkhead homologue to rhabdomyosarcoma (FKHR) is differentially regulated by insulin and IGF-1 in hepatocytes. The phosphorylation of one threonine residue in particular (Thr-24) appears to be induced by insulin but not by IGF-1. As this residue can be phosphorylated by PKB \textit{in vitro} and PKB is also activated by IGF-1 in these cells, the authors proposed that a PKB-like kinase specifically activated by insulin may mediate this effect (Nakae et al., 2000). More recently, it has been shown that although both insulin and IGF-1 induce proliferation of murine skin keratinocytes, the action of insulin — but not IGF-1 — is mediated specifically via a protein kinase C delta (PKC\(\delta\)) and involves activation of the sodium/potassium (\(Na^+/K^+\)) pump (Shen et al., 2001). Thus, PKC\(\delta\) is a multifunctional serine kinase that represents a divergence point in IR and IGF-1R signaling. In this same cell type, insulin and IGF-1 stimulate the translocation of different glucose transporters, although they both increase glucose uptake (Shen et al., 2001). Thus, insulin and IGF-1 can mediate the same or different biological responses by utilizing different signaling pathways or different intracellular mediators.
IV. Induction of Specific Genes by the IR and IGF-1R

Some reports have shown that insulin and IGF-1 can act on the same genes but with different outcomes. For example, in murine skin keratinocytes, insulin stimulates the expression of differentiation markers, whereas IGF-1 inhibits them (Wertheimer et al., 2000). Also, in the developing eye lens of the chicken, the level of delta-crystallin induced by IGF-1 is greater and occurs more quickly than that induced by insulin (Alemany et al., 1989). It has been shown that low concentrations of IGF-1 (10 nM) increase the expression of uncoupling protein 3 (UCP-3) by 2-fold, whereas much higher concentrations of insulin (860 nM) are necessary to obtain the same effect in human neuroblastoma SH-SYSY cells (Gustafsson et al., 2001).

cDNA microarray analysis recently has been established as a powerful tool to study the effects of hormones on cellular metabolism and gene regulation on a genomic scale. Until now, this technology was used to define the effects of IGF-1 on gene expression in different cell lines (Liu et al., 2001; Oh et al., 2002) but not to compare the different gene-expression profiles induced by insulin and IGF-1. We used cDNA microarray expression profiling to identify genes that are regulated differently by IGF-1 and insulin in mouse fibroblast NIH-3T3 cells (Dupont et al., 2001b) as a first step towards understanding the molecular basis for the different functions of the IGF-1R and the IR.

A. DIFFERENTIAL REGULATION OF GENE-EXPRESSION PATTERNS BY INSULIN AND IGF-1 IN NIH-3T3 FIBROBLASTS

The biological and physiological comparison of the IR and IGF-1R is complicated by the fact that each ligand can cross-react with the other receptor and hybrid receptors can form when both receptors are expressed in the same cells. To circumvent these problems, we have compared the effect of insulin and IGF-1 in NIH-3T3 fibroblasts overexpressing either human IR (IR cells) (Levy-Toledano et al., 1993) or human IGF-1R (NWTb3 cells) (Blakesley et al., 1995, 1996). NWTb3 and IR cells were incubated in the presence or absence of IGF-1 (50 nM) or insulin (50 nM) for 90 minutes, respectively. Of the 2221 genes on the mouse cDNA microarrays, we found that the expression levels of 30 were significantly induced by IGF-1 but not by insulin. In contrast, only nine genes and one expressed sequence tag (EST) were upregulated specifically by insulin but not by IGF-1 (Tables I and II). We confirmed the IGF-1- and insulin-induced regulation for 10 of these genes by Northern analysis (Figure 3). The genes that were identified as regulated by IGF-1 and insulin are involved in various cellular functions, including proliferation, differentiation, apoptosis, cellular processes, and metabolism (Tables I and II). Interestingly, most of these genes were not known previously to be regulated by either IGF-1 or insulin. Indeed, only three genes — the Jun oncogene (Chiou and Chang, 1992; Monnier et al., 1994), α5
integrin (Palmade et al., 1994), and the early growth response-1 transcription factor (EGR-1) (Jhun et al., 1995) — had been reported to be induced by IGF-1. Furthermore, more than half of the genes upregulated by IGF-1 are associated with mitogenesis and differentiation, whereas none of the genes specifically upregulated by insulin are associated with these processes. IGF-1, but not insulin, induced the expression of two cytokine receptors (interleukin (IL) receptors 3 and 4) that have been reported to be involved in the regulation of cell growth (Keegan et al., 1994). IGF-1 also induced the expression of glial cell line-derived neurotrophic factor (GNDF), which is known to be crucial for the development and the maintenance of various neurons (Airaksinen and Saarma, 2002). IGF-1 increased the expression of the Wee-1-like kinase, which is involved in cell-cycle progression (Helmbrecht et al., 2000), and the EGR-1 transcription factor, which is known to enhance cell proliferation. These results suggest that IGF-1-induced cellular proliferation is a tightly regulated process.

Our study also suggested that insulin and IGF-1 are involved in the apoptosis process. IGF-1 treatment increased expression of the T-cell death-associated gene (TDAG)-51 and Daxx (Fas-binding) genes, whereas insulin increased expression of apoptotic protease-activating factor-1 (APAF-1) and seven in absentia homologue-1B (SIAH-1B) (Tables I and II). Importantly, IGF-1 is capable of increasing the expression of antiapoptotic genes such as Twist (Maestro et al., 1999). Thus, the induction of IGF-1- or insulin-specific genes could explain the specificity of the biological effects of these two hormones.

B. TWIST EXPRESSION IS SPECIFICALLY INDUCED BY IGF-1

In a separate study, we studied Twist, one of the genes that was specifically induced by the IGF-1-responsive gene (Dupont et al., 2001a). Twist belongs to the basic helix-loop-helix family of transcription factors, which play a central role in cell-type determination and differentiation in both vertebrates and invertebrates (Olson and Klein, 1994). IGF-1 treatment increased the abundance of Twist mRNA in NWTb3 cells, whereas insulin failed to increase Twist mRNA in IR cells. The IGF-1-induced increase in Twist expression requires activation of IGF-1R, since Twist mRNA expression was not induced in response to IGF-1 in parental NIH-3T3 cells, which express few IGF-1Rs, nor in the NKR (NIH-3T3 cells expressing an IGF-1 receptor with lysine-to-arginine substitution) cell line, which overexpresses the dominant-negative human IGF-1R (Kato et al., 1993). We also showed that injection of IGF-1 via the inferior vena cava increased Twist mRNA expression in muscle. We used various pharmacological inhibitors and a MEK-1 dominant-negative construct to investigate which IGF-1R signaling pathway was involved in the induction of Twist gene expression. These experiments demonstrated that the MEK/MAPK pathway plays a critical role in IGF-1-induced Twist expression. Using an antisense strategy, we
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Clone number</th>
<th>IGF-1</th>
<th>Insulin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mitogenesis and differentiation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interleukin 3 receptor, α-chain</td>
<td>IL-3Ra</td>
<td>445664</td>
<td>5.32</td>
</tr>
<tr>
<td>Colony stimulating factor, macrophage</td>
<td>mCSF</td>
<td>634838</td>
<td>4.12</td>
</tr>
<tr>
<td>Glial cell line-derived neurotrophic factor</td>
<td>GNDF</td>
<td>425671</td>
<td>3.96</td>
</tr>
<tr>
<td>Integrin α-5 (fibronectin receptor)</td>
<td>Iα5</td>
<td>476908</td>
<td>3.55</td>
</tr>
<tr>
<td>Early growth response-1</td>
<td>EGR-1</td>
<td>608153</td>
<td>3.65</td>
</tr>
<tr>
<td>Jun oncogene</td>
<td>JUN</td>
<td>949554</td>
<td>3.01</td>
</tr>
<tr>
<td>Twist gene homolog</td>
<td>TWIST</td>
<td>479367</td>
<td>2.95</td>
</tr>
<tr>
<td>Forkhead homolog 14</td>
<td>FKH-14</td>
<td>541099</td>
<td>2.91</td>
</tr>
<tr>
<td>Wee 1-like protein kinase</td>
<td>Wee-1</td>
<td>539548</td>
<td>2.75</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 10</td>
<td>IGF-BP10</td>
<td>557055</td>
<td>2.41</td>
</tr>
<tr>
<td>Sex-determining region Y (SRY)-box containing gene 2</td>
<td>SRY-2</td>
<td>351033</td>
<td>2.39</td>
</tr>
<tr>
<td>Interleukin 4 receptor alpha</td>
<td>IL-4Ra</td>
<td>721594</td>
<td>2.30</td>
</tr>
<tr>
<td>Mouse mRNA for dbpa murine homolog</td>
<td>DBPA</td>
<td>602275</td>
<td>2.29</td>
</tr>
<tr>
<td>Expressed sequence tags, moderately similar to MAK16 (S. cerevisiae)</td>
<td>MAK16</td>
<td>537328</td>
<td>2.27</td>
</tr>
<tr>
<td>Ngfi-A binding protein 2</td>
<td>NGFI-A BP-2</td>
<td>476298</td>
<td>2.31</td>
</tr>
<tr>
<td>MAD (mothers against decapentaplegic) homolog 5 (Drosophila)</td>
<td>MAD5</td>
<td>551401</td>
<td>2.24</td>
</tr>
<tr>
<td>Early development regulator</td>
<td>EDR</td>
<td>616348</td>
<td>2.22</td>
</tr>
<tr>
<td>Ets variant gene 6 (TEL oncogene)</td>
<td>TEL</td>
<td>402134</td>
<td>2.21</td>
</tr>
<tr>
<td><strong>Apoptosis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mas musculus TDAG51 (T-cell death-associated gene)</td>
<td>TDAG51</td>
<td>694076</td>
<td>9.00</td>
</tr>
<tr>
<td>Mas musculus Fas-binding protein (Daxx)</td>
<td>Daxx</td>
<td>736796</td>
<td>5.99</td>
</tr>
<tr>
<td><strong>Cellular processes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murine mRNA for replacement variant histone H3.3</td>
<td>vH3.3</td>
<td>618380</td>
<td>3.30</td>
</tr>
<tr>
<td>Kinesin heavy chain member 1A</td>
<td>Kin1A</td>
<td>492514</td>
<td>2.83</td>
</tr>
<tr>
<td>Mouse chromatin nonhistone high-mobility group protein (HMG-I(Y))</td>
<td>HMG-1(Y)</td>
<td>616054</td>
<td>2.64</td>
</tr>
</tbody>
</table>
also showed that Twist is positively involved in the antiapoptotic effects of the IGF-1R. These studies show that a gene that is regulated by IGF-1 receptor activation may, in turn, regulate the function of the IGF-1 receptor function.

C. IGF-1-INDUCED GENES AND CANCER PROGRESSION

IGF-1-regulated genes were studied in a preneoplastic, immortalized breast cell line, 184htert. The advantage of utilizing this cell line is that genes may be identified that are affected by IGF-1 and the IGF-1 receptor signaling pathways at an early stage in the progression of cancer. These cells were created by using retroviral technology to introduce the human telomerase reverse transcriptase gene into normal breast epithelial cells. Of the ~2000 known genes on the microarray chip, 156 (8%) were regulated by IGF-1. These genes exhibited various patterns of regulation; whereas some were either up- or downregulated at early time points, others were regulated in a biphasic manner. The IGF-1-responsive genes could be subdivided into various categories (e.g., transcription factors, cell cycle-related genes, genes involved in cancer progression, signaling-related genes, extracellular matrix genes, genes related to metabolism). Interestingly, IGF-1 regulated a large number of genes involved in angiogenesis. Many genes known to stimulate angiogenesis were upregulated.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Clone number</th>
<th>IGF-1</th>
<th>Insulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mas musculus mRNA for eRF1</td>
<td>eRF-1</td>
<td>572924</td>
<td>2.34</td>
</tr>
<tr>
<td>DEAD (aspartate-glutamate-alanine-aspartate) box polypeptide 5</td>
<td>DEAD5</td>
<td>537478</td>
<td>2.22</td>
</tr>
<tr>
<td>Splicing factor, arginine/serine 3 (SRp20)</td>
<td>SRp20</td>
<td>595904</td>
<td>2.41</td>
</tr>
<tr>
<td>Metabolism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murine Glvr-1 mRNA</td>
<td>GLVR-1</td>
<td>335579</td>
<td>4.88</td>
</tr>
<tr>
<td>Glycerol phosphate dehydrogenase 1, mitochondrial</td>
<td>GPDH</td>
<td>351221</td>
<td>2.74</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear factor erythroid-derived 2, like 2</td>
<td>NF-E2</td>
<td>635541</td>
<td>2.90</td>
</tr>
<tr>
<td>Immediate early protein Gly96</td>
<td>Gly96</td>
<td>579574</td>
<td>2.46</td>
</tr>
</tbody>
</table>

[Adapted from Dupont J, Khan J, Qu BH, Metzler P, Helman L, LeRoith D 2001 Insulin and IGF-1 induce different patterns of gene expression in mouse fibroblast NIH-3T3 cells: identification by cDNA microarray analysis. Endocrinology 142:4969–4975.]
by IGF-1, whereas inhibitors of angiogenesis such as plasminogen activator inhibitor-1 (PAI-1) and metalloproteases were inhibited by IGF-1 (Table III). The effects of IGF-1 on many of these genes — including c-fos, VEGF, Fas ligand, cyp1A1, cyp1B1, interleukin-1β, and uPA — were validated by other techniques. Many of the genes that are regulated by IGF-1 are also responsive to the hypoxia-inducible factor-1alpha (HIF-1α) and cAMP response binding protein (CREB) transcription factors. Indeed, IGF-1 induced nuclear translocation of HIF-1α and the phosphorylated form of CREB, thereby inducing gene expression.

Thus, this study demonstrated that IGF-1 regulates the expression of many genes involved in cancer progression. This new information may be helpful when considering gene targeting for therapeutic uses in the treatment of cancer.

<table>
<thead>
<tr>
<th>Morphogenesis and development</th>
<th>Symbol</th>
<th>Clone ID</th>
<th>IGF-1</th>
<th>Insulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse alpha-B crystallin mRNA</td>
<td>CRYαB</td>
<td>605970</td>
<td>1.56</td>
<td>2.28</td>
</tr>
<tr>
<td>Calponin H1, smooth muscle</td>
<td>CNNh1</td>
<td>557012</td>
<td>1.27</td>
<td>2.10</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>APAF-1</td>
<td>657503</td>
<td>1.33</td>
<td>2.20</td>
</tr>
<tr>
<td>Seven in absentia 1B</td>
<td>SIAH-1B</td>
<td>618379</td>
<td>1.30</td>
<td>2.04</td>
</tr>
<tr>
<td>Cellular processes</td>
<td>TAU</td>
<td>552102</td>
<td>1.36</td>
<td>2.23</td>
</tr>
<tr>
<td>Integrin alpha 6</td>
<td>α6</td>
<td>584662</td>
<td>1.63</td>
<td>2.05</td>
</tr>
<tr>
<td>Cytochrome P450 2d10</td>
<td></td>
<td></td>
<td>1.46</td>
<td>2.34</td>
</tr>
<tr>
<td>Others</td>
<td>PRL-R</td>
<td>520835</td>
<td>0.72</td>
<td>3.74</td>
</tr>
<tr>
<td>Delta-aminolevulinate dehydratase</td>
<td>DAH</td>
<td>518879</td>
<td>1.53</td>
<td>2.13</td>
</tr>
<tr>
<td>Expressed sequence tags, highly similar to</td>
<td></td>
<td>539102</td>
<td>1.79</td>
<td>2.13</td>
</tr>
<tr>
<td>envelope (ENV) polyprotein precursor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Adapted from Dupont J, Khan J, Qu BH, Metzler P, Helman L, LeRoith D 2001 Insulin and IGF-1 induce different patterns of gene expression in mouse fibroblast NIH-3T3 cells: identification by cDNA microarray analysis. Endocrinology 142:4969–4975.]
IV. Conclusion and Future Directions

The purpose of this review is to bring the possible applications of this exciting new technology to the attention of researchers. While cDNA microarray analysis is

<table>
<thead>
<tr>
<th>Gene</th>
<th>B3</th>
<th>C43</th>
<th>IR</th>
<th>IGF-1</th>
<th>INSULIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDNF</td>
<td></td>
<td></td>
<td></td>
<td>4.2 ± 0.2</td>
<td>1.0</td>
</tr>
<tr>
<td>EGR-1</td>
<td></td>
<td></td>
<td></td>
<td>8.4 ± 0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>mCSF</td>
<td></td>
<td></td>
<td></td>
<td>6.2 ± 0.2</td>
<td>2.1</td>
</tr>
<tr>
<td>TDAG51</td>
<td></td>
<td></td>
<td></td>
<td>4.3 ± 0.3</td>
<td>1.7</td>
</tr>
<tr>
<td>GLVR-1</td>
<td></td>
<td></td>
<td></td>
<td>3.8 ± 0.1</td>
<td>1.2</td>
</tr>
<tr>
<td>TWIST</td>
<td></td>
<td></td>
<td></td>
<td>2.7 ± 0.3</td>
<td>1.2</td>
</tr>
<tr>
<td>eRF1</td>
<td></td>
<td></td>
<td></td>
<td>2.5 ± 0.4</td>
<td>1.3</td>
</tr>
<tr>
<td>Wee-1</td>
<td></td>
<td></td>
<td></td>
<td>3.5 ± 0.6</td>
<td>1.6</td>
</tr>
<tr>
<td>PRLR</td>
<td></td>
<td></td>
<td></td>
<td>1.0 ± 0.2</td>
<td>2.7</td>
</tr>
<tr>
<td>ICα6</td>
<td></td>
<td></td>
<td></td>
<td>1.4 ± 0.3</td>
<td>2.7</td>
</tr>
</tbody>
</table>

FIG. 3. Confirmation of the specific gene expression by IGF-1 or insulin using Northern blot analysis. Northern blot analysis was performed using RNA from cells expressing IGF-1 receptors (B3 and C43) or insulin receptors (IR), following stimulation. The specificity of stimulated gene expression correlated with the microarray results seen in Tables I and II. Abbreviations: CSF, colony-stimulating factor; EGR, early growth response; GDNF, glial cell line-derived neurotrophic factor; GLVR, gibbon ape leukemia virus receptor; PRLR, prolactin receptor; TDAG, T-cell death-associated gene. [Reprinted with permission from Dupont J, Khan J, Qu BH, Metzler P, Helman L, LeRoith D 2001 Insulin and IGF-1 induce different patterns of gene expression in mouse fibroblast NIH-3T3 cells: identification by cDNA microarray analysis. Endocrinology 142:4969–4975. Copyright The Endocrine Society.]
associated with a number of technical hurdles and experimental flaws, its utility has enormous potential, if used carefully, with multiple controls and constant validation. This approach may prove to be useful as a rapid screening test to identify the many genes that are differentially regulated in different tissues and systems. In particular, it may lead to the identification of genes not previously known to be affected by a particular process. Furthermore, in studying the progression of disease states and variations in gene expression, in various models, it may prove to be the more rapid and economical method. Numerous other applications undoubtedly will emerge as more investigators utilize this technology.

REFERENCES


| TABLE III |

Genes Specifically Affected by IGF-1 in Breast Epithelial Cells

<table>
<thead>
<tr>
<th>Genes Specifi cally Affected by IGF-1 in Breast Epithelial Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-cell lymphoma-2 interacting killer</td>
</tr>
<tr>
<td>c-fos</td>
</tr>
<tr>
<td>Cytochrome P450 1A1 and 1B1</td>
</tr>
<tr>
<td>Ferrodoxine reductase</td>
</tr>
<tr>
<td>GADD45 (growth arrest and DNA damage)</td>
</tr>
<tr>
<td>Interleukin 1 beta</td>
</tr>
<tr>
<td>Jun B</td>
</tr>
<tr>
<td>Low-density lipoprotein-related protein</td>
</tr>
<tr>
<td>Plasminogen activator inhibitor-1</td>
</tr>
<tr>
<td>Transferrin</td>
</tr>
<tr>
<td>Vascular endothelial growth factor</td>
</tr>
</tbody>
</table>


Blenis J 1993 Signal transduction via the MAP kinases: proceed at your own RSK. Proc Natl Acad Sci USA 90:5889–5892


Dupont J, Khan J, Qu BH, Metzler P, Helman L, LeRoith D 2001b Insulin and IGF-1 induce different patterns of gene expression in mouse fibroblast NIH-3T3 cells: identification by cDNA microarray analysis. Endocrinology 142:4969–4975


Furlanetto RW, Dey BR, Lopaczynski W, Nisley SP 1997 14-3-3 proteins interact with the insulin-like growth factor receptor but not the insulin receptor. Biochem J 327:765–771


Kiely PA, Sant A, O’Connor R 2002 RACK1 is an insulin-like growth factor I- (IGF-I) receptor-interacting protein that can regulate IGF-1-mediated Akt activation and protection from cell death. J Biol Chem 277:22581–22589


LeRoith D, Werner H, Beitner-Johnson D, Roberts CT Jr 1995 Molecular and cellular aspects of the insulin-like growth factor I receptor. Endocr Rev 16:143–163


Olson EN, Klein WH 1994 bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. Genes Dev 8:1–8


Soni P, Lakkis M, Poy MN, Fernstrom MA, Najjar SM 2000 The differential effects of pp120 (Ceacam 1) on the mitogenic action of insulin and insulin-like growth factor 1 are regulated by the nonconserved tyrosine 1316 in the insulin receptor. Mol Cell Biol 20:3896–3905


Withers DJ, Gutierrez JS, Towery H, Burks DJ, Ren JM, Previs S, Zhang Y, Bernal D, Pons S, Shulman GI, et al. 1998 Disruption of IRS-2 causes type 2 diabetes in mice. Nature 391:900–904


Consequences of Elevated Luteinizing Hormone on Diverse Physiological Systems: Use of the LHβCTP Transgenic Mouse as a Model of Ovarian Hyperstimulation-induced Pathophysiology

RACHEL J. MANN, RUTH A. KERI, AND JOHN H. NELSON

Department of Pharmacology, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

ABSTRACT

Chronically elevated luteinizing hormone (LH) induces significant pathology in the LHβCTP transgenic mouse model, which uses the bovine gonadotropin alpha (α)-subunit promoter to direct transgene expression specifically to gonadotropes in the anterior pituitary. Previously, it was shown that female LHβCTP mice are infertile due to anovulation, develop granulosa cell tumors, and undergo precocious puberty from elevated LH and steroid hormones that fail to completely repress the α-subunit promoter. This chapter will discuss recent studies that further elucidate the impact of chronically elevated LH on diverse physiological systems. Granulosa cell tumors induced by elevated LH are strain dependent and prevented when transgenics are treated with human chorionic gonadotropin (hCG) surges. A granulosa cell tumor-associated transcriptome is generated, revealing several possible gene candidates for ovarian granulosa cell tumorigenesis. Primordial follicles in LHβCTP transgenics become depleted and oocytes exhibit increased rates of meiotic segregation defects, although meiotic competency is acquired normally. Anovulation can be rescued in transgenics by superovulation, though pregnancy fails at midgestation due to maternal factors. Uterine receptivity defects prevent implantation of normal embryos following induction of pseudopregnancy. Transgenics develop Cushing-like adrenocortical hyperfunction with increased corticosterone production following induction of adrenal LH receptor expression. Elevated LH acts as a tumor promoter in the gonads and the adrenal gland, when expressed in conjunction with the inhibin-α SV40 transgene. Finally, chronic elevated LH promotes mammary tumorigenesis. The understanding of multiple clinical pathologies — including ovarian cancer, perimenopausal reproductive aging, premature ovarian failure, polycystic ovarian syndrome, Cushing’s syndrome, and breast cancer — may be enhanced through further study of this useful transgenic mouse model.

I. Introduction

A. PHYSIOLOGICAL ROLE OF LUTEINIZING HORMONE AND THE HYPOTHALAMIC-PITUITARY-GONADAL AXIS

Luteinizing hormone (LH) is a member of the glycoprotein hormone family, which includes follicle-stimulating hormone (FSH), thyroid-stimulating hormone
(TSH), and chorionic gonadotropin (CG). These hormones are heterodimers composed of a common \( \alpha \) subunit and a unique \( \beta \) subunit, which confers hormone specificity (Bousfield *et al*., 1994). LH is involved primarily in gametogenesis and steroid hormone production in both males and females.

To fully understand the function of LH physiologically, one must understand the context in which it exists. This context includes where it comes from, what drives its production, and what regulates its repression. The major system that controls reproductive physiology in mammals is the hypothalamic-pituitary-gonadal (HPG) axis. Gonadotropin-releasing hormone (GnRH), a decapeptide, is secreted in a pulsatile fashion from neurons located in the hypothalamus into the hypothalamic-hypophyseal portal system (Page, 1994). It then exerts its effect through GnRH receptors located on the surface of gonadotrope cells in the anterior pituitary. The seven transmembrane GnRH receptor is a G protein-coupled receptor that stimulates multiple signaling pathways, including modulation of intracellular calcium concentrations and activation of protein kinase C and other protein kinases (e.g., mitogen-activating protein kinase, MAPK) (Kraus *et al*., 2001). To what extent these paths cross-talk, or differentially regulate LH\(\beta\) subunit versus the \(\alpha\) gonadotropin subunit, is still under investigation (Gajewska and Kochman, 2001). GnRH signaling stimulates both the synthesis and secretion of the gonadotropins LH and FSH from the anterior pituitary.

LH exerts its effect by binding the G protein-coupled LH receptor, initiating signaling primarily through the second messenger, cyclic AMP (cAMP) (Ascoli *et al*., 2002). In females, LH stimulates ovarian thecal cell production of androstenedione, which diffuses to granulosa cells and undergoes conversion to estrogen by aromatase. FSH controls aromatase activity and also induces the appearance of LH receptors in the granulosa cells of preovulatory follicles, further augmenting estrogen production (Mathews *et al*., 1987). During the preovulatory surge, LH binds to LH receptors on granulosa cells, stimulating meiotic maturation of the oocyte and release of the oocyte from the follicle. Thus, granulosa cells with both FSH and LH receptors are capable of responding to LH directly, causing luteinization and release of progesterone (Gore-Langton and Armstrong, 1994). These steroids exert regulatory feedback onto the HPG axis by acting on the hypothalamus and gonadotrope cells, primarily in a negative (repressive) manner (Haisenleder *et al*., 1994). In males, the role of LH in spermatogenesis via stimulation of androgen production has been well established. Activation of LH receptors on testicular Leydig cells stimulates cAMP release, inducing synthesis of testosterone from cholesterol. In adults, LH-stimulated testosterone is the primary determinant of spermatogenesis (Sharpe, 1994). Due to the LH\(\beta\)CTP sexually dimorphic phenotype, the remainder of this review will focus primarily on the impact of LH hypersecretion in females.
A highly regulated system has evolved to ensure that LH is produced at appropriate concentrations, which, in females, change with time in a cyclic nature. Abnormalities leading to over- or underactivation of the HPG axis have extreme consequences on reproductive health (de Roux and Milgrom, 2001). Too little LH results in underdevelopment of the gonads and subsequent decline (or cessation) of gametogenesis. Too much LH causes precocious puberty and excessive steroid hormone production, which can lead to cell immortalization and tumorigenesis. In humans, inappropriate LH regulation can result in inability to produce a genetically related family, a hardship that is not only physiologically difficult to overcome but also potentially psychologically debilitating. Thus, it is important that we understand the physiologic role of LH and its ability to induce reproductive pathology upon dysregulation.

B. GENERATING LH HYPERSECRETION IN LHβCTP TRANSGENIC MICE

Many approaches can be taken to generate mice with elevated serum levels of LH. However, most of these methods rely on pharmacological means to elevate LH to nonphysiological levels. In addition, it is difficult to devise protocols for chronic administration of exogenous LH that mimic endogenous LH pulse patterns. By using a transgenic approach, our laboratory not only achieved a physiological means of elevating LH but also limited its synthesis and secretion strictly to gonadotrope cells in the anterior pituitary. This approach leaves intact the HPG axis and permits studies that depend on the responsiveness of this axis. Studying the impact of chronically elevated LH in the whole animal permits a physiological method of disease modeling in multiple, diverse systems, as described in this review.

The LHβCTP transgenic mouse was generated by designing a transgene (depicted in Figure 1) that incorporates the C-terminal peptide (CTP) of the hCGβ subunit fused to the C-terminus of the bovine LHβ cDNA and its first intron. The 87-bp CTP fragment extends the half-life of the resulting LHβCTP fusion product by 2- to 3-fold (Risma et al., 1995). An SV40 polyadenylation signal was added at the C-terminus of the transgene. Transgene expression is driven by the 360-bp (−315 to +45) bovine gonadotropin α-subunit promoter, which has been previously characterized and targets expression specifically to gonadotrope cells in the anterior pituitary (Kendall et al., 1991; Hamernik et al., 1992). This promoter retains responsiveness to GnRH, estrogens, and androgens in transgenic mice (Keri et al., 1991; Hamernik et al., 1992; Clay et al., 1993).

C. PHENOTYPE OVERVIEW

Some aspects of the impact of chronic elevated LH on female LHβCTP transgenic mice have been reviewed previously in greater detail (Nilson et al.,
2000). Originally described in Risma et al. (1995), serum LH is elevated in female transgenics at 39.7 ng/mL vs. 2.7 ng/mL in controls. Expression of the transgene led to elevated serum LH levels by at least 2 weeks of age, though serum levels also changed from 2–6 weeks, presumably in response to puberty and the “reawakening” of the HPG axis (Risma et al., 1997). Testosterone in transgenic females is elevated 5-fold over control females by 2 weeks of age; estradiol also is elevated by 3 weeks of age by 2- to 5-fold (Risma et al., 1995, 1997).

Despite pathologically elevated levels of steroid hormones, the bovine gonadotropinα-subunit promoter driving expression of the LHβCTP transgene does not become repressed. In contrast, the endogenous mouse LHβ-subunit gene becomes completely silent during the transition from the neonatal period to adulthood (Abbud et al., 1999). Interestingly, the endogenous mouse gonadotropinα-subunit gene remains active in addition to the transgene. Abbud et al. showed that α-subunit promoter activity in the presence of elevated steroid hormones is due to independence from GnRH regulation and, as a result, a lack of responsiveness to estradiol-negative feedback. Since this bovine α-subunit promoter previously was shown to be responsive to GnRH and estradiol in transgenic studies using a chloramphenicol acetyl transferase (CAT) reporter gene (Keri et al., 1991; Hamernik et al., 1992), loss of α-subunit promoter responsiveness to GnRH and estradiol must be mediated by the pathological effects of exposure to high levels of LH early in life, possibly through the reprogramming of key signal transduction cascades.

Female LHβCTP transgenic mice are infertile due to chronic anovulation, while males appear to have normal fertility. However, male founder transgenics appeared to be slightly subfertile due to breeding delays and had significantly
smaller testes than controls (Risma et al., 1995). Transgenic females enter puberty precociously at 21 days of age vs. 30 days in controls (Risma et al., 1997). By 3 weeks of age, transgenic females have accelerated folliculogenesis and multifollicular ovaries. By 5–6 weeks of age, transgenic ovaries become significantly enlarged and contain multiple fluid- or blood-filled cysts (Risma et al., 1997). Prolonged luteal life span with elevated serum progesterone is induced in transgenics following hemiovariectomy and mating with vasectomized males. Morphological analyses of ovaries from older animals reveal ovarian granulosa cell tumors or theca-interstitial cell tumors by 4 months of age (Risma et al., 1995).

Transgenic females develop renal abnormalities — including enlarged bladders, dilated ureters, and hydronephrosis — sometimes associated with acute pyelonephritis that can lead to death (Risma et al., 1995). This renal pathology probably is due to exposure to chronically elevated steroids, since hydronephrosis has been observed in rats chronically administered estradiol (Corriere and Murphy, 1968) and is well documented during pregnancy in women (Andriole, 1975). Interestingly, this renal pathology appears to occur most frequently in CF-1 female transgenics but only rarely in [C57BL/6xCF-1] F1 hybrids (Nilson et al., 2000). Thus, female transgenics on C57BL/6 background live longer and go on to develop pituitary hyperplasia by 4 months of age, with pituitary adenoma formation by 10–12 months of age (Nilson et al., 2000).

II. Pathophysiology Induced by Elevated LH

A. OVARY PATHOLOGY: MODELING OVARIAN TUMORIGENESIS

1. Granulosa Cell Tumors Are Strain Dependent

Initial studies with LHβCTP transgenics were performed on mice with a combined genetic background of CF-1, C57BL/6, and SJL. These mice developed granulosa cell tumors on an occasional basis. However, when the transgene was bred over multiple generations into a pure CF-1 background, the tumor phenotype became 100% penetrant by 5 months of age. The dependency of tumor formation on genetic background suggested that additional modifiers existed that could act in conjunction with elevated LH to cause or prevent the formation of granulosa cell tumors. In Keri et al., this dependency was investigated. Figure 2 shows the granulosa cell tumor that develops when the transgene is on an undiluted CF-1 background. Transgenic F1 hybrid strains — generated by breeding the CF-1 transgenic mice to CD-1, C57BL/6, or SJL — failed to form granulosa cell tumors. Instead, they uniformly developed cystic ovaries with an extensively luteinized phenotype, reminiscent of a luteoma of pregnancy observed in women (Piana et al., 1999). Therefore, apparent dependence of
granulosa cell tumor formation on a specific genetic profile may explain the rarity of these tumors in women (Wynder et al., 1969).

At 5 months of age, all F1 hybrid transgenics had elevated LH, even relative to transgenics on the CF-1 background. An altered hormonal profile, ovarian cysts, and precocious puberty are phenotypes induced by the transgene that are independent of strain background. The fact that the granulosa cell tumor phenotype was lost after only one generation of breeding suggests that at least one gene with recessive effects must be involved in the differential ovarian response to elevated LH (Keri et al., 2000).

To estimate the number of recessive genes involved in granulosa cell tumor strain dependency, backcrosses between the C57BL/6xF1 transgenic hybrids (♂) and CF-1 nontransgenic (♀) were performed. Of the 138 transgenic females generated from this study, 19 harbored at least one granulosa cell tumor. The fit between the predicted and observed (13.8%) frequency of tumor appearance supports a three-gene model (Keri et al., 2000). The expected (predicted) frequency of tumor formation for two-, three-, and four-gene models are, respectively, 25%, 12.5%, and 6.25% (predicting 34.5, 17.3, and 8.6 tumors out of 138 evaluated). Subsequent to this publication, more transgenic females have been evaluated, altering the final observed frequency to 30/295 (10.2%), further
supporting that three unlinked, recessive genes may discriminate tumor susceptibility in response to elevated LH (R.A. Keri, personal communication).

Ongoing studies to determine the identity of the genes required for granulosa cell tumors are being performed using microsatellite markers and bulk segregant analysis to identify contributing loci. Recent preliminary data indicate phenotype linkage to mouse chromosomes 10 and 11 (R.A. Keri, personal communication).

2. LH Surges Prevent Granulosa Cell Tumor Formation

While granulosa cell tumors comprise only 10% of ovarian tumors, they exhibit the potential for malignancy and recurrence, making them clinically significant (Wynder et al., 1969, Fontanelli et al., 1998; Lee et al., 1999). Efforts to identify the molecular mechanisms leading to development of these tumors are important, as they may reveal possible new therapeutic targets.

Clinical studies have implicated elevated gonadotropins levels in tumorigenesis. The strongest correlative evidence comes from postmenopausal women, who represent the largest cohort of patients with granulosa cell tumors (Amsterdam and Selvaraj, 1997).

Studies using other types of transgenic mice support this view as well. Mice deficient in inhibin α develop granulosa cell tumors at an early age (Matzuk et al., 1992). The lack of inhibin α leads to an increase in LH, FSH, and estrogens (Kumar et al., 1999). When these animals also are made genetically deficient for GnRH (hpg/hpg), tumors fail to form, further underscoring the importance of LH and FSH (Kumar et al., 1996). In contrast, mice deficient in FSH and inhibin continue to develop tumors, although with increased latency and decreased penetrance (Kumar et al., 1999). These data suggest that elevated LH, independent of FSH, can induce granulosa cell tumor formation when mice also are deficient for inhibin α, while FSH may accelerate this process (Owens et al., 2002).

In Owens et al. (2002), subtractive gene expression profiling was employed, comparing normal ovaries, LHβCTP (CF-1) granulosa cell tumors, and LHβCTP (F1 hybrids) luteomas. This subtractive method (diagrammed in Figure 3) was employed to evaluate the differential gene expression profile that distinguishes a granulosa cell tumor from a normal or luteoma-bearing ovary. Equal amounts of total RNA were pooled from at least four mice for each experimental group, to minimize changes due to interindividual variation. Global changes in gene expression profiles were assessed with Affymetrix Mu11K oligodeoxynucleotide microarrays containing approximately 11,000 genes and expressed sequence tags (ESTs). Samples were normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and analyzed using self-organizing maps (SOMs) to reveal informative patterns of gene expression (Tamayo et al., 1999). The data were filtered using Affymetrix parameters to exhibit changes that were at least 3-fold, since
FIG. 3. Subtractive gene-expression profiling compares normal ovaries, LHβCTP (CF-1) granulosa cell tumors, and LHβCTP (F1 hybrids) luteomas. (A) Eight experimental groups are represented by the circles; B6/C = (C57BL/6, Wt ♀ × CF-1, Tg ♂). (B) The expression-profiling technique is diagrammed with four of the eight experimental groups for simplicity. The small percentage of genes from each comparison group is overlapped, identifying genes associated only with the formation of a granulosa cell tumor. [Reprinted with permission from Owens GE, Keri RA, Nilson JH 2002 Ovulatory surges of hCG prevent hormone-induced granulosa cell tumor formation leading to the identification of tumor-associated changes in the transcriptome. Mol Endocrinol 16:1230–1242. Copyright The Endocrine Society.]
these are highly reproducible in replicate samples (Cho et al., 1998). Clusters were generated of 47 genes whose expression dramatically decreased and 75 whose expression increased significantly.

To identify consistent patterns in the tumor-associated transcriptome, analysis of RNA from tumors and luteomas was repeated on a new sample set containing three ovaries in each group. This further limited the cluster of genes with dramatic expression changes to 26 reduced and 46 increased. These genes, listed in Tables I and II, may represent common expression alterations that contribute to granulosa cell tumor formation in LHβCTP mice (Owens et al., 2002).

Many of the genes identified have been associated previously with ovarian tumors (e.g., c-fos, N-myc, platelet-derived growth factor (PDGF)-A, inhibin βA and βB), further implicating them as contributors to the tumor phenotype (Versnel et al., 1994; Gercel-Taylor and Taylor, 1996; Luthra and Chapekar, 1998; deKretser et al., 2000). Further verification of a subset of the identified cluster genes was performed by semiquantitative reverse transcription-polymerase chain reaction (RT-PCR) on granulosa cell tumors and luteomas obtained from individual animals (as opposed to the RNA pooling done previously). Expression differences in LH-R, steroidogenic acute regulator protein (StAR), inhibin βB, and Wnt-4 observed between luteomas and granulosa cell tumors through microarray analyses were recapitulated.

Owens et al. further evaluated the role of elevated LH in granulosa cell tumor development by determining whether the lack of ovulatory surges of LH in genetically predisposed LHβCTP (CF-1) mice might contribute to the granulosa cell tumor phenotype. LHβCTP CF-1 mice were treated with ovulatory doses of hCG every fourth day for 5 months, beginning at 2 weeks of age. Transgenic female littermates in control groups received saline injections. All transgenic animals (n = 4) receiving hCG injections failed to develop granulosa cell tumors while, as expected, all littermate controls (n = 3) receiving saline injections developed tumors by 5 months of age (Owens et al., 2002). Ovaries from all four hCG-treated animals developed luteomas, a histological phenotype indistinguishable from those seen in F1 hybrid transgenics (C57BL/6xC57BL/6xCF-1). This result indicates that long-term restoration of ovulatory surges of hCG is capable of antagonizing the genetic predisposition for granulosa cell tumor formation in CF-1 transgenic mice. Interestingly, the ovarian transcriptome of CF-1 transgenic mice treated with hCG closely resembled that of the F1 hybrid transgenics (C57BL/6xC57BL/6xCF-1). These results are shown in the right column of Tables I and II.

Figure 4 depicts the finding that chronically elevated LH initiates a molecular pathway that, in conjunction with a genetic predisposition, leads to the development of granulosa cell tumors. However, in the absence of this genetic predisposition, or in the presence of ovulatory-like surges of hCG, chronically
TABLE I
Gene Expression Decreases Associated with Granulosa Cell Tumors

<table>
<thead>
<tr>
<th>Accession number</th>
<th>Gene name</th>
<th>Average-fold decrease (tumor vs. luteoma)</th>
<th>Reproduced in hCG-induced luteoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA087277</td>
<td>Secreted frizzled related protein 4</td>
<td>400</td>
<td>Yes</td>
</tr>
<tr>
<td>Z27088</td>
<td>Relaxin precursor</td>
<td>300</td>
<td>Yes</td>
</tr>
<tr>
<td>AA122502</td>
<td>NG-CAM related adhesion molecule precursor</td>
<td>150</td>
<td>Yes</td>
</tr>
<tr>
<td>M14757</td>
<td>Multidrug resistant protein</td>
<td>100</td>
<td>No</td>
</tr>
<tr>
<td>M89797</td>
<td>Wnt-4</td>
<td>100</td>
<td>Yes</td>
</tr>
<tr>
<td>W15994</td>
<td>Laminin B receptor</td>
<td>49.0</td>
<td></td>
</tr>
<tr>
<td>D17433</td>
<td>Prostaglandin F receptor</td>
<td>43.0</td>
<td></td>
</tr>
<tr>
<td>L36062</td>
<td>StAR</td>
<td>33.0</td>
<td></td>
</tr>
<tr>
<td>D14636</td>
<td>PEBP2a1 protein</td>
<td>32.0</td>
<td></td>
</tr>
<tr>
<td>D32137</td>
<td>MOPG</td>
<td>31.0</td>
<td></td>
</tr>
<tr>
<td>W41963</td>
<td>Acetyl-CoA synthetase</td>
<td>16.0</td>
<td>No</td>
</tr>
<tr>
<td>M62766</td>
<td>HMG-CoA reductase</td>
<td>14.0</td>
<td>Yes</td>
</tr>
<tr>
<td>Z22532</td>
<td>Syndecan-1</td>
<td>13.0</td>
<td>No</td>
</tr>
<tr>
<td>I13593</td>
<td>Prolactin receptor</td>
<td>11.5</td>
<td>Yes</td>
</tr>
<tr>
<td>D42048</td>
<td>Squalene epoxidase</td>
<td>9.5</td>
<td>No</td>
</tr>
<tr>
<td>X61940</td>
<td>Growth factor-inducible immediate early gene</td>
<td>9.0</td>
<td>Yes</td>
</tr>
<tr>
<td>AA066425</td>
<td>Tumor-associated antigen</td>
<td>8.5</td>
<td>No</td>
</tr>
<tr>
<td>M81310</td>
<td>Luteinizing hormone receptor</td>
<td>8.0</td>
<td>Yes</td>
</tr>
<tr>
<td>X56304</td>
<td>Tenascin</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>U49507</td>
<td>Lisch 7</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>M28698</td>
<td>Cytokeratin 19</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>AA016727</td>
<td>Farnesyl pyrophosphate synthetase</td>
<td>7.0</td>
<td>No</td>
</tr>
<tr>
<td>I05781</td>
<td>Cytosolic epoxide hydrolase</td>
<td>6.0</td>
<td>Yes</td>
</tr>
<tr>
<td>M37761</td>
<td>Calycyclin</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>M31419</td>
<td>204 interferon activatable protein</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>AA000961</td>
<td>Hemoglobinase precursor</td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

[Reprinted with permission from Owens GE, Keri RA, Nilson JH 2002 Ovulatory surges hCG prevent hormone-induced granulosa cell tumor formation leading to the identification of tumor-associated changes in the transcriptome. Mol Endocrinol 16:1230–1242. Copyright The Endocrine Society.]

Elevated levels of LH result in the formation of a luteoma rather than a granulosa cell tumor. Further molecular clarification of the signaling pathways involved in these two different outcomes may help refine our understanding of human granulosa cell tumors as well as the molecular pathways regulated by LH.
B. FOLLICLE AND OOCYTE DEVELOPMENT: MODELING PERIMENOPAUSAL REPRODUCTIVE AGING AND PREMATURE OVARIAN FAILURE

1. Primordial Follicles Are Depleted in LHβCTP Mice

In human females, the nonrenewable follicular reserve of dormant primordial follicles becomes depleted with age and usually is exhausted between ages 45 and 55 (Richardson and Nelson, 1990). While depletion occurs gradually throughout the reproductive life span, as primordial follicles are recruited into the growing follicular pool, during the period of perimenopause, the loss of primordial follicles accelerates at a rate twice that seen previously (Richardson et al., 1987). If not for this acceleration, women would remain fertile well into their eighties (Faddy et al., 1992). During perimenopause, women experience transient elevations in the levels of both LH and FSH (Sherman et al., 1976; Marcus et al., 1993). It is thought that this altered endocrine milieu exerts a toxic effect on the primordial follicle pool, accelerating its depletion (Richardson et al., 1987).

Accelerated primordial follicle depletion also is seen in the disorder known as premature ovarian failure (POF). Although of unknown origin, POF clearly is associated with early and accelerated follicle atresia (Rebar, 1982; Cohen and Speroff, 1991). POF is characterized by primary or secondary amenorrhea, with elevated levels of serum gonadotropins and early menopause (Bione et al., 1998). Women with POF are infertile due to anovulation, experience depletion of their follicular reserve, and show hormonal profiles similar to older women who are postmenopausal (Russell, 1997).

In Flaws et al. (1997), ovaries from LHβCTP mice were evaluated for follicular depletion. By scoring serial sections of ovaries obtained from transgenics and control littermates, a 45% decline in primordial follicles was observed in transgenics by 5 weeks of age. By 3 months, ovaries from transgenic mice contained 68% fewer primordial follicles, when compared to controls (Figure 5). As the primordial follicle pool becomes depleted, primary follicle numbers begin to decline as well. By 3 months, primary follicles are 45% depleted in transgenics vs. controls (Flaws et al., 1997). These data suggest that elevated LH leads to depletion of the primordial and primary follicle pool and thus alters the timing of reproductive senescence. Further evaluation of the molecular mechanism of this depletion may increase our understanding of primordial pool depletion observed during perimenopause and POF.

2. Oocyte Meiotic Competency Is Not Altered

The altered hormonal milieu and associated ovarian pathology in transgenic females occur around the time of precocious puberty (21 days), which is also the
<table>
<thead>
<tr>
<th>Accession number</th>
<th>Gene name</th>
<th>Average-fold decrease (tumor vs. luteoma)</th>
<th>Reproduced in hCG-induced luteoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>X13586</td>
<td>2,3-bisphosphoglycerate mutase</td>
<td>130</td>
<td>Yes</td>
</tr>
<tr>
<td>U81603</td>
<td>Eya2 homolog</td>
<td>40.0</td>
<td></td>
</tr>
<tr>
<td>W35058</td>
<td>Frizzled 1</td>
<td>39.0</td>
<td></td>
</tr>
<tr>
<td>U85610</td>
<td>Indian hedgehog protein</td>
<td>27.0</td>
<td></td>
</tr>
<tr>
<td>AB004048</td>
<td>Neuronatin</td>
<td>26.0</td>
<td></td>
</tr>
<tr>
<td>J04946</td>
<td>Angiotensin converting enzyme</td>
<td>22.0</td>
<td></td>
</tr>
<tr>
<td>AA111277</td>
<td>Visinin-like protein</td>
<td>22.0</td>
<td></td>
</tr>
<tr>
<td>U37459</td>
<td>Glia-derived neurotrophic growth factor (GDNF)</td>
<td>21.0</td>
<td></td>
</tr>
<tr>
<td>M69069</td>
<td>MHC class I mRNA</td>
<td>21.0</td>
<td>No</td>
</tr>
<tr>
<td>X68837</td>
<td>Secretogranin II</td>
<td>17.0</td>
<td>Yes</td>
</tr>
<tr>
<td>M29464</td>
<td>Platelet-derived growth factor A-chain</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>X69620</td>
<td>Inhibin β-B subunit</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>AA105452</td>
<td>Glia-derived nexin precursor</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>M15525</td>
<td>Laminin B1</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>X03919</td>
<td>N-myc</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td>X69619</td>
<td>Inhibin β-A subunit</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>J056439</td>
<td>IGFBP-2</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>M18194</td>
<td>Fibronectin</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>AA097626</td>
<td>Pol polyprotein</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>X04322</td>
<td>Melanoma-inhibitory-activity protein</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>V00727</td>
<td>c-fos oncogene</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>X13945</td>
<td>L-myc</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>M70642</td>
<td>FiSP-12</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>U17961</td>
<td>p62 mRNA</td>
<td>7.0</td>
<td>No</td>
</tr>
<tr>
<td>AA035915</td>
<td>Ras-like protein TC21</td>
<td>6.5</td>
<td>Yes</td>
</tr>
<tr>
<td>U79766</td>
<td>Ajuba</td>
<td>6.0</td>
<td>No</td>
</tr>
<tr>
<td>AF004326</td>
<td>Angiopoietin-2</td>
<td>6.0</td>
<td>Yes</td>
</tr>
<tr>
<td>W48402</td>
<td>SIR2</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>X89627</td>
<td>17-beta-hydroxysteroid dehydrogenase</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>AA064226</td>
<td>RAB-11B</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>M31131</td>
<td>N-cadherin</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>X70853</td>
<td>Fibulin C</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>W82053</td>
<td>EF-2</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>W41733</td>
<td>Protein kinase C substrate</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>X75285</td>
<td>Fibulin 2</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>X70854</td>
<td>Fibulin D</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>Z28532</td>
<td>Follistatin</td>
<td>4.0</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE II**

Gene Expression Increases Associated with Granulosa Cell Tumors
time during which oocyte meiotic competency is acquired. Along with LH, testosterone becomes elevated in transgenics by 2 weeks of age (Risma et al., 1997). Abnormally elevated levels of androgens are thought to induce follicular atresia and oocyte degeneration (Hsueh et al., 1994). Other studies have suggested that elevated androgens can cause premature activation of meiotic prophase I-arrested oocytes (Wasserman et al., 1994). In addition to androgen-induced follicular demise, elevated levels of LH may have a direct detrimental impact on oocyte development (Jacobs and Homburg, 1990). For example, increased LH concentrations during the follicular phase may result in the inappropriate activation of meiotic prophase I-arrested oocytes (Wasserman et al., 1994).

Meiotic competency, or the ability to resume meiosis after release from the ovarian follicle, is indicated by germinal vesicle breakdown (GVBD), followed by polar body (PB) extrusion after overnight culture. No differences could be observed between transgenics or control littermates: at 15 days of age, oocytes arrest at GVBD, while at 21 days, meiotic competency has been acquired and oocytes extrude a polar body, presumably arrested at metaphase of meiosis I (Mann et al., 1999). Thus, despite development within the context of elevated LH, estrogen, and testosterone, and the depletion of primordial follicles, acquisition of meiotic competency in oocytes from transgenics overtly appears to remain unaltered.

---

**TABLE II**

(continued)

<table>
<thead>
<tr>
<th>Accession number</th>
<th>Gene name</th>
<th>Average-fold decrease (tumor vs. luteoma)</th>
<th>Reproduced in hCG-induced luteoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA063914</td>
<td>Tubulin alpha chain</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>L04538</td>
<td>Amyloid precursor-like protein</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>AF119416</td>
<td>GM3 synthase</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>U79748</td>
<td>DPC4</td>
<td>3.5</td>
<td>No</td>
</tr>
<tr>
<td>X70853</td>
<td>Fibulin C</td>
<td>3.0</td>
<td>Yes</td>
</tr>
<tr>
<td>W13162</td>
<td>CDK4</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>W43968</td>
<td>Myosin heavy chain 1B</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>W81863</td>
<td>Extensin precursor</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Z22784</td>
<td>Troponin 1</td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

[Reprinted with permission from Owens GE, Keri RA, Nilson JH 2002 Ovulatory surges hCG prevent hormone-induced granulosa cell tumor formation leading to the identification of tumor-associated changes in the transcriptome. Mol Endocrinol 16:1230–1242. Copyright The Endocrine Society.]
3. Elevated LH Increases the Rate of Oocyte Meiotic Segregation Defects

While acquisition of meiotic competency occurs normally in transgenics, elevated LH does significantly impact the organization of chromosomes on the meiotic I spindle during oocyte maturation (Hodges et al., 2002). This study suggests that the altered endocrine milieu of LHβCTP transgenics results in an increase in meiotic nondisjunction, similar to that observed during human reproductive aging.

Errors in chromosome segregation during human meiotic divisions result in the loss of a significant number of conceptions. The high incidence of meiotic errors in humans has a strong correlation to maternal age (Hassold and Chiu, 1985), though the mechanisms underlying these errors remain unclear. Data show that the age-related increase in oocyte meiotic errors is characteristic of all racial groups (Hassold and Chiu, 1985) and that most meiotic errors have their genesis at meiosis I (Hassold and Hunt, 2001). Studies on human trisomies have further elucidated the impact of age on meiotic nondisjunction but have not revealed the molecular mechanisms involved (Hodges et al., 2002). Recently, immunofluorescent studies have suggested that early events in meiosis I may contribute to subsequent nondisjunction (Battaglia et al., 1996; Volarcik et al., 1998).

Congression failure, or the inability of chromosomes to move to the equator of the meiosis I (MI) spindle due to gross aberrations in spindle morphology
and/or chromosome alignment, is another age-related phenomenon (Volarcik et al., 1998). Since female meiosis (unlike mitosis or male meiosis) is not subject to chromosomal monitoring checkpoint mechanisms of the metaphase/anaphase transition, this alignment failure may be linked to eventual nondisjunction (LeMaire-Adkins et al., 1997; Woods et al., 1999; Burke, 2000; Shah and Cleveland, 2000).

Because the altered endocrine environment is known to disrupt folliculogenesis in the LHβCTP female, Hodges et al. evaluated oocytes from LHβCTP transgensics (and other mouse models) for evidence of congression failure and subsequent indications of nondisjunction. On both a heterogeneous background and the C57BL/6 inbred strain, oocytes from the first follicular wave in LHβCTP were analyzed for chromosome alignment at metaphase I (6 hours in culture) and metaphase II (16–18 hours in culture). Table III shows the results of these studies (Hodges et al., 2002). A high level of congression failure was observed at both MI (18.2% vs. 0–1% in controls) and meiosis II (MII) metaphase (38.6% vs. 0% in controls) in LHβCTP mice. In addition, meiotically immature oocytes from

FIG. 5. Chronically elevated LH depletes primordial follicle pools. Ovaries from control (nontransgenic) and transgenic (LHβCTP) mice were collected and sectioned at the ages indicated. Every fifth section was sampled to estimate the number of primordial follicles per ovary. *Significantly different from control. [Reprinted with permission from Flaws JA, Abbud R, Mann RJ, Nilson JH, Hirshfield AN 1997 Chronically elevated luteinizing hormone depletes primordial follicles in the mouse ovary. Biol Reprod 57:1233–1237.]
nontransgenics (partially competent 18- to 20-day-old mice) did not exhibit congression failure. These studies show that congression failure is not a symptom of meiotic immaturity but, rather, a reflection of oocyte growth in an altered environment. It also was observed that the degree of meiotic disturbance appeared to correlate to the extent of ovarian pathology, another indicator of the abnormal endocrine environment.

Both microinjection of kinetochore antibodies into mouse oocytes (Simerly et al., 1990) and immunodepletion of kinetochore-associated motor protein, CENP-E, in Xenopus egg extracts (Wood et al., 1997) produce a congression failure phenotype. Thus, Hodges et al. postulated that the congression failure observed in LHβCTP oocytes (and reproductively aged human oocytes) may be due to endocrine affects on production of one or more microtubule motor proteins.

Hodges et al. also asked if congression failure causes an increase in oocyte aneuploidy through nondisjunction and premature sister chromatid segregation (PSCS). Table IV shows that LHβCTP-derived oocytes do exhibit a significant increase in total hyperploidy and/or PSCS (18% vs. 5% controls). These data, along with data from another congression failure mutant (not shown), support the conclusion that increased frequency of congression failure at MI results in increased impairment of segregation of homologous chromosomes at anaphase I.

### Table III

<table>
<thead>
<tr>
<th>Hours in culture</th>
<th>Meiotic stage</th>
<th>Total cells scored</th>
<th>Normal MI alignment (%)</th>
<th>Congression failure (%)</th>
<th>Congression with outliers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (26–28 day)</td>
<td>8 MI</td>
<td>147</td>
<td>137 (93)</td>
<td>0</td>
<td>10 (7)</td>
</tr>
<tr>
<td></td>
<td>10 MI</td>
<td>227</td>
<td>210 (93)</td>
<td>3 (1)</td>
<td>14 (6)</td>
</tr>
<tr>
<td></td>
<td>16–18 MI</td>
<td>59</td>
<td>58 (98)</td>
<td>0</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Partially competent (18–20 day)</td>
<td>8 MI</td>
<td>43</td>
<td>35 (81)</td>
<td>0</td>
<td>8 (19)</td>
</tr>
<tr>
<td></td>
<td>16–18 MI</td>
<td>42</td>
<td>45 (98)</td>
<td>0</td>
<td>1 (2)</td>
</tr>
<tr>
<td>LHβCTP (26–28 day)</td>
<td>6 MI</td>
<td>88</td>
<td>67 (76)</td>
<td>16 (18)</td>
<td>5 (6)</td>
</tr>
<tr>
<td></td>
<td>16–18 MI</td>
<td>57</td>
<td>28 (49)</td>
<td>22 (39)</td>
<td>7 (12)</td>
</tr>
</tbody>
</table>

These results suggest that changes in the endocrine milieu that occur during human reproductive aging and the perimenopausal period may contribute significantly to the age-related increase in oocyte meiotic nondisjunction, possibly due to errors in chromosomal alignment at MI, such as those seen during congression failure.

C. INFERTILITY: MODELING POLYCYSTIC OVARIAN SYNDROME

Clinical observations have shown that hypersecretion of LH impairs fertility in women by disrupting ovarian function and maintenance of pregnancy (Regan et al., 1990; Balen et al., 1993; Shoham et al., 1993). Polycystic ovarian syndrome (PCOS) affects up to 10% of women of reproductive age and is a leading cause of anovulatory infertility (Adams et al., 1986). Women with PCOS typically present with elevated LH and androgens, chronic anovulation, and polycystic ovaries (Franks, 1995). However, PCOS is an etiologically complex disorder that often is associated with hirsutism, obesity, insulin resistance, and pregnancy complications (Legro et al., 1998). Elucidating the mechanism by which elevated LH induces female reproductive disorders such as PCOS has been difficult due to the complexity of these syndromes. Generating a mouse model that recapitulates many of the features of PCOS would be useful to allow for greater understanding of its treatment and prevention.

Women with elevated LH who are diagnosed with PCOS have difficulty conceiving, frequently requiring the use of fertility medications to induce ovulation. Unfortunately, those who do conceive experience an elevated miscarriage rate of 30–64% (compared to a rate of 12% in women with normal LH) (Hamilton-Fairley and Franks, 1990; Regan et al., 1990). Because this increased miscarriage rate in women with PCOS often is attributed to poor oocyte quality (Brzyski et al., 1995), we evaluated oocyte, embryo, and maternal reproductive health in LHβCTP transgenics (Mann et al., 1999).

### TABLE IV

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Normal chromosomes (%)</th>
<th>PSCS (%)</th>
<th>Hyperploid (%)</th>
<th>PSCS &amp; hyperploid (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>44</td>
<td>42 (95)</td>
<td>2 (5)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LHβCTP</td>
<td>44</td>
<td>36 (82)</td>
<td>2 (5)</td>
<td>3 (7)</td>
<td>3 (7)</td>
</tr>
</tbody>
</table>

As discussed previously, oocytes from LHβCTP transgenic mice acquire meiotic competency normally, despite early exposure to elevated LH and androgens. Subsequent studies revealed abnormalities in LHβCTP meiotic chromosomal segregation (Hodges et al., 2002). It is possible that a pool of these oocytes would develop abnormally and result in pregnancy failure if evaluated. However, when superovulation was used to recruit a large group of oocytes, which then were fertilized and transferred to pseudopregnant, nontransgenic recipients, they developed into normal, live-born pups.

LHβCTP females are infertile primarily because of anovulation, without ovulating spontaneously, though they will mate, as exhibited by the presence of repetitive vaginal plugs. However, both a single dose of hCG and a superovulatory regimen of pregnant mare serum gonadotropin (PMSG), followed by hCG 48 hours later, induces LHβCTP ovulation. This suggests that chronically elevated LH causes a loss of normal LH ovulatory surges. Following superovulation and mating, LHβCTP transgenics will become pregnant but they experience pregnancy failure at midgestation. By 12 days post coitus (dpc), transgenic pregnancy resorption approaches 100%, whereas nontransgenic resorption reaches only 21% (elevated slightly from normal, presumably due to large litters from superovulation). The rate of pregnancy loss is plotted in Figure 6.

As discussed, embryos from transgenic females that were transferred to nontransgenic recipients developed normally to term. This result indicates that pregnancy failure in LHβCTP transgenics is maternal, not embryonic, in origin. However, when embryos from nontransgenics were transferred into pseudopregnant LHβCTP transgenics, they did not implant. This implantation failure is due to a lack of appropriate uterine receptivity, as demonstrated by the failure of transgenic uteri to undergo decidualization when stimulated with corn oil. Monitoring serum levels of progesterone throughout pseudopregnancy revealed some alterations, suggesting that the signaling pathway initiated by the mating stimulus and resulting in uterine receptivity may be abnormal in LHβCTP females. Superovulation appears to overcome this receptivity defect and reveals a subsequent midgestation disorder.

Levels of LH, estradiol, testosterone, and progesterone were measured throughout superovulation-induced pregnancy and compared to pregnant nontransgenics. Progesterone levels were not significantly different from nontransgenics following superovulation, suggesting that corpus luteum failure was not responsible for the midgestation pregnancy resorption phenotype. Testosterone does not become elevated until 14 dpc, after the onset of pregnancy failure. Interestingly, estradiol becomes elevated 4-fold over nontransgenics from 8–14 days gestation, during the critical window of pregnancy failure. Elevated estradiol during midgestation has been shown to be toxic to embryos (Huggett and Pritchard, 1945). Elevated estradiol causes midgestation pregnancy failure in mice lacking the 5α-reductase type I gene (Mahendroo et al., 1997). Thus, it is
possible that excess estradiol is the cause of LH/CTP midgestation pregnancy failure. Interestingly, estrogen toxicity has been used to explain some human miscarriages (Trout and Seifer, 2000). LH/CTP pregnancy defects could be rescued, with live pups born when animals were ovariectomized following embryo transfer and treated with estrogen and progesterone during pregnancy. While inefficient, the success of this regimen demonstrated that the reproductive abnormalities experienced by LH/CTP females are reversible (Mann et al., 1999).

This study suggests that pregnancy failure in women with elevated LH, such as those diagnosed with PCOS, may be attributed to a hostile maternal environment contributing to lack of uterine receptivity and/or pregnancy loss. Future studies to determine the molecular mechanisms involved in LH/CTP pregnancy failure may help us to further understand the complexities behind increased miscarriage rates in women with PCOS.
D. ADRENAL DISTURBANCES: MODELING CUSHING’S SYNDROME

Although most of the androgens in PCOS are secreted from the ovaries, in approximately 50% of cases, excessive production of adrenal androgens also occurs (Azziz, 1996). The underlying cause of this adrenocortical disturbance remains unclear.

The primary regulator of glucocorticoid production in the adrenal cortex is adrenocorticotropin (ACTH) (Orth and Kovacs, 1998). In humans, ACTH drives the adrenal gland to produce the glucocorticoids cortisol and corticosterone. However, corticosterone is the only glucocorticoid produced in the mouse (Spackman and Riley, 1978).

In addition to ACTH, several other hormones (e.g., LH, prolactin, insulin-like growth factor-1 (IGF-1)) have been implicated in the regulation of adrenal androgen production (Orth and Kovacs, 1998). There is some evidence that LH acts directly at the adrenal gland, although this is controversial (Parker and Odell, 1980). A case of postmenopausal Cushing’s syndrome has been described in which the adrenocortical hyperfunction was found to be LH dependent and responsive to treatment with GnRH agonist (Lacroix et al., 1999). In addition, a significant proportion of women with chronic anovulation have elevated serum levels of LH and adrenal-derived androgen precursors such as dehydroepiandrosterone-sulfate (DHEA-S) but normal ACTH levels (Hoffman et al., 1984).

In Kero et al. (2000), adrenal gland function and LH receptor expression were evaluated in LH/H252CTP transgenics. They presented evidence for a novel mechanism that might explain the adrenal-derived androgens in PCOS and some LH-dependent Cushing’s syndrome cases.

Adrenal glands from 5-month-old transgenic females were evaluated and found to be morphologically altered, compared to nontransgenic littermates. Adrenal glands from transgenics weighed significantly more than controls (6.0 ± 0.86 mg vs. 3.3 ± 0.3 mg, respectively). Histological signs of cortical stimulation included increased cortical width and centripetal extension of lipid-depleted cells, widening the zona reticularis. There were also foci of acute and chronic inflammatory cells (Kero et al., 2000). Notably, lymphocyte infiltration of the adrenal gland has been observed in some patients with Cushing’s syndrome (Willenberg et al., 1998).

While at 1 month of age, both transgenics and controls had a typical area of the adrenal X-zone visible, by 5 months of age, the X-zone had disappeared in transgenics. A poorly characterized area, the X-zone is thought to be involved in steroid hormone production (Hu et al., 1999). Interestingly, in normal male mice, the X-zone disappears after puberty, possibly due to testosterone (Asari et al., 1979). However, in normal females, the X-zone remains until after the first pregnancy (Holmes and Dickson, 1971).
To determine whether the altered adrenal morphology reflected a change in adrenal steroidogenesis, serum corticosterone levels were measured and determined to be elevated 14-fold in female transgenics, when compared to nontransgenic controls. This elevation decreased to nearly undetectable levels following adrenalectomy, indicating that the primary source of the corticosterone was the adrenal gland. In addition, transgenic ovariectomy resulted in decreased corticosterone levels to that of nontransgenic controls. Interestingly, gonadectomy, which induces elevated LH in both males and females, did not increase corticosterone levels in nontransgenics. This suggests that the polycystic ovaries found in transgenics play a significant role in adrenal hyperfunction and that chronically elevated LH alone cannot account for the increased adrenal steroidogenesis (Kero et al., 2000).

The steroidogenic capacity of adrenal glands from 3-month-old transgenics was assessed further through culturing dispersed adrenal cells. Unstimulated cells from transgenic females produced elevated levels of cAMP, progesterone, and corticosterone, relative to nontransgenics. In addition, transgenic-derived adrenal cells responded to both hCG (in a dose-dependent manner) and ACTH treatment. However, cells from nontransgenics responded only to ACTH. These data indicate that, in addition to ACTH, an LH analog stimulates adrenal corticosterone production in transgenics. This increased steroidogenic capacity is associated with increased LH receptor expression in the transgenic adrenal gland. Significant and specific LH receptor binding in adrenal homogenates was observed in transgenic females. RT-PCR was used to demonstrate that long-term elevation of serum LH concentrations is associated with the presence of LH receptor mRNA in adrenal glands. LH receptor expression was observed in female transgenics by 3 months of age and in gonadectomized nontransgenics after 5 months (although at lower levels) (Kero et al., 2000). In situ hybridization localized LH receptor expression in transgenic adrenal glands across the whole cortex, including the zona glomerulosa, zona fasciculata, and zona reticularis. Thus, the presence of the cystic ovary, in conjunction with elevated LH, seems to be a requirement for achieving significant adrenal LH receptor expression along with functional responsiveness of adrenal cells to hCG.

Besides serving as a model for PCOS, these findings of increased corticosterone production in LHβCTP transgenics suggest that these mice may provide a useful model for Cushing-like adrenocortical hyperfunction (Kero et al., 2000). Other manifestations of Cushing’s syndrome (e.g., obesity) have been observed in these mice. LHβCTP mice become obese, weighing 30% more than their nontransgenic littermates (R.A. Keri, J. Kero, I.T. Huhtaniemi, J.H. Nilson, unpublished data). Interestingly, in rare cases of ACTH-independent Cushing’s syndrome, the disease becomes more prominent during pregnancy and may improve or remit spontaneously after delivery (Da Motta et al., 1991; Buescher et al., 1992; Close et al., 1993).
The mechanism involved in the development of this Cushing-like adrenocortical hyperfunction is unclear. It is known that LHβCTP mice have 2- to 3-fold elevated circulating estradiol (Risma et al., 1995). Estradiol is known to elevate prolactin, which is upregulated in transgenics by 2.5-fold (Kero et al., 2000). Prolactin is an important regulator of LH receptor expression at both the level of transcription and translation (Huhtaniemi and Catt, 1981; Gafvels et al., 1992; Pakarinen et al., 1994). This may explain the ovarian dependence for induction of adrenal LH receptor expression and subsequent stimulation of corticosterone production by elevated LH seen in transgenics. This represents a novel mechanism behind the altered adrenocortical function observed in transgenics, which results in elevated glucocorticoid production. As these mice appear to model some aspects of both PCOS and Cushing’s syndrome, further studies on adrenocortical function are warranted in humans with chronically elevated gonadotropin levels (Kero et al., 2000).

E. LH ACTS AS A TUMOR PROMOTER IN LHβCTP/INHαTAG DOUBLE TRANSGENICS

Previous studies on mice transgenic for the inhibin-α promoter directing expression of the SV40 T-antigen showed that 100% of these mice developed granulosa cell or Leydig cell tumors by 5–8 months of age (Kananen et al., 1995, 1996a). When gonadectomized, these mice developed adrenal gland tumors with 100% penetrance by 6–8 months (Kananen et al., 1996b). While the adrenal tumor dependency on gonadectomy suggested a role for increased levels of gonadotropins — and, indeed, was shown to be gonadotropin dependent (Kananen et al., 1997) — these studies did not differentiate between the actions of FSH and LH. Thus, a new line of double transgenics was created to introduce the LHβCTP transgene onto the inhαTag line (Kero et al., 2002).

In the double-transgenic mice, gonadal tumorigenesis starts earlier and progresses faster, compared to inhαTag transgenics. At 3 months of age, LHβCTP female ovaries are enlarged and multicystic, while ovaries from double transgenics are even larger and present with highly proliferating tumors of granulosa cell origin. Ovaries from nontransgenics and inhαTag transgenics are normal at this time point. The presence of elevated LH in female double transgenics enhanced the aggressiveness of their ovarian tumors. While gonadal tumors in inhαTag transgenics never metastasized to other tissues, double transgenics were observed to generate metastases to the lungs and liver.

Interestingly, these studies revealed some previously unseen phenotypes in male LHβCTP transgenics. While serum LH levels originally were not observed to be elevated in males (Risma et al., 1995), this study found slight, but significant, elevation of LH at 9–12 weeks of age (5.18 ± 0.45 μg/mL vs. 1.35 ± 0.34 μg/mL). Additionally, some Leydig cell hyperplasia was observed. As
previously identified, testicles from LHβCTP males were slightly, but significantly, reduced in size (Risma et al., 1995). This can be explained by the known effects of elevated LH with concomitantly low FSH before puberty, leading to early differentiation of testicular cells and reduced testicular size (Gaytan et al., 1994). Three-month-old male double transgenics had increased testicular weight and showed massive hemorrhagic and invasive tumors originating from Leydig cells, while inhαTag transgenic males showed only microscopic tumor foci at this age.

Histological signs of adrenal gland tumorigenesis in all double-transgenic females were observed in the presence of the gonad. These tumors clearly originated from the adrenal cortex. In contrast, adrenal tumors were observed in inhαTag transgenics only after gonadectomy. Not surprisingly, no evidence of adrenal tumorigenesis was seen in male double transgenics, presumably due to their only slightly elevated LH levels. These findings suggest that chronically elevated LH, in the presence of normal to decreased FSH, can act as a tumor promotor on the inhαTag transgenic background.

The progression of ovarian tumorigenesis in double transgenics is associated with elevated LH receptor and inhibin-α gene expression. Interestingly, high serum levels of LH fail to induce downregulation of LH receptor in both LHβCTP and double transgenics. Indeed, LH receptor is upregulated, indicating that the normal desensitization of LH receptor by LH is altered in these animals (Dufau, 1998). Chronic activation of the LH receptor function could act as a protooncogene, as shown with some other G protein-coupled receptors (Allen et al., 1991; Parma et al., 1993). It is possible that the enhanced tumorigenesis seen in double transgenics could be explained by LH receptor stimulation of increased cAMP production and subsequent activation of the cAMP response element (CRE) contained within the inhibin-α transgene promoter (Pei et al., 1991).

F. MAMMARY TUMORIGENESIS: MODELING BREAST CANCER

Mammary gland development is a hormonally regulated process. Progression of the rudimentary mammary ductal system present at birth to an extensive network following puberty and finally into a differentiated, milk-producing organ during pregnancy and lactation is mediated by estrogen, progesterone, and prolactin (Hennighausen and Robinson, 1998). These hormones also regulate the occurrence and timing of reproductive events that influence a woman’s risk of developing breast cancer. While many mouse models of mammary cancer exist, few address the role that hormones play in spontaneous tumor formation.

Known risk factors for development of breast cancer include early age at menarche, late age of menopause, late age at first full-term pregnancy, or nulliparity (Kvale, 1992; Stoll et al., 1994; Armstrong et al., 2000). Interestingly, parity has a dual effect on breast cancer. While full-term pregnancy actually
reduces long-term tumor incidence in humans and rodents (MacMahon et al., 1982; Grubbs et al., 1986), a short-term increased risk in tumor formation has been observed immediately following pregnancy (Hsieh et al., 1994; Lambe et al., 1994).

In Milliken et al. (2002), the impact of LH-mediated alterations in the hormonal milieu of LHβCTP transgenics on the development of the mouse mammary gland was investigated. Initial studies revealed that mammary gland development is accelerated in LHβCTP transgenics. Precocious puberty occurs in transgenics around postnatal days 21–22, at least 5 days earlier than nontransgenic littermates (Risma et al., 1997; Keri et al., 2000). To assess the impact of early ovarian activity on mammary gland morphogenesis, Milliken et al. examined glands at 3 and 5 weeks of age. By 3 weeks of age, accelerated development of the mammary gland was observed in transgenic mice (Figures 7A and B). This accelerated growth of the transgenic mammary gland was more apparent by 5 weeks of age (Figures 7C and D). While the ductal network in nontransgenics progressed through only half of the mammary gland fat pad by 5 weeks of age,
the entire fat pad of the transgenic mouse was filled with ducts that displayed abundant alveoli and a loss of terminal end buds (Milliken et al., 2002). This pattern of development closely resembles that seen at mid- to late pregnancy in normal mice (Hennighausen and Robinson, 1998).

Further histological examination of adult transgenic mammary glands revealed extensive epithelial hyperplasia with considerable alveolar development and accumulation of lipid droplets. This hyperplasia was determined to be due to a 12-fold increase in proliferation, as measured by incorporation of bromodeoxyuridine (BrdU). However, a corresponding change in the rate of apoptosis in these glands was not observed. The histological evidence of pregnancy-like morphology was supported by the expression pattern of molecular markers of pregnancy-induced differentiation within mammary glands of transgenics. Milk protein genes, whey acidic protein (WAP), β-casein, and Westmead DMBA-8 nonmetastatic cDNA 1 (WDNM1) expression levels in transgenics corresponded to those observed in nontransgenics at day 14 of pregnancy. These findings suggest that mammary glands from virgin transgenic animals exposed to excessive LH develop a phenotype that simulates midpregnancy (Milliken et al., 2002).

Ovariectomy was performed to determine whether the mammary hyperplasia observed in transgenics was due to direct action of LH or whether it required LH-induced ovarian hyperstimulation. Mammary glands collected after 21 days postovariectomy showed complete regression of the gland, compared to sham surgery controls. These results indicate that hyperplasia is reversible and requires ovary-derived factors.

While transgenics develop spontaneous mammary tumors (mostly mammary intraepithelial neoplasias (MINs)) with 50% penetrance by 41 weeks (Figure 8), treatment with the mammary carcinogen, 7,12-dimethylbenz(a)anthracene (DMBA) accelerates tumor formation. Transgenics treated with DMBA develop invasive mammary carcinomas with squamous metaplasia beginning at 13.5 weeks posttreatment (with 100% penetrance by 20 weeks), while only 20% of nontransgenic controls treated with DMBA developed tumors by 56 weeks postexposure. These data demonstrate that LH/H9252 transgenics are predisposed to mammary carcinogenesis, when compared to their nontransgenic littermates.

Levels of estrogen, progesterone, and prolactin in transgenics were measured at several different ages (starting at 5 weeks), to identify the hormonal profile that supports the development of mammary hyperplasia and subsequent tumor development. Significant increases in estrogen and progesterone were detected in transgenics at 5 weeks of age, while prolactin became significantly elevated later in life, perhaps due to chronic elevations in estrogen, which increases lactotrope secretion of prolactin (Ascoli and Segaloff, 1996). Tumor-bearing mice have significantly higher levels of both estrogen and prolactin, while progesterone remains comparable to levels seen at nontumor-bearing transgenics at 20 weeks.
It is known that growth of mammary tumors can be classified as either hormone dependent or independent. Analysis of human breast tumors has revealed a significant correlation between expression of estrogen and progesterone receptors and hormone growth dependency (DePlacido et al., 1990; Ormandy et al., 1997). Immunohistochemical analysis of the spontaneous mammary tumors formed in transgenics for progesterone receptor (PR) indicated that while normal tissue adjacent to tumor cells did express nuclear PR, tumor cells did not. Unfortunately, detection of estrogen receptor (ER) expression was not possible at that time.

These studies have demonstrated that chronic exposure to elevated LH leads to precocious mammary gland development and ovary-dependent mammary hyperplasia. Hyperplasia is due to an increase in epithelial cell proliferation. It
PHYSIOLOGICAL CONSEQUENCES OF ELEVATED LUTEINIZING HORMONE  369

will be important to determine whether precocious puberty in LHβCTP transgenics significantly contributes to their development of mammary tumors, considering that early puberty is a risk factor for breast cancer in humans (Apter et al., 1989).

Despite histological and molecular markers suggesting the mammary glands are in a midpregnancy-like state, LHβCTP transgenics have accelerated tumor formation. This finding contrasts with studies indicating a protective effect of pregnancy against mammary tumor formation. The apparent discrepancy may suggest that the protection achieved through pregnancy is regulated by the extent and timing of hormone elevation.

Further studies are being performed to dissect the molecular mechanisms that govern the hormone-induced mammary tumorigenesis seen in this model. Gene-expression profiling has revealed many genes that behave similarly in transgenic and pregnant mammary glands, as compared to virgin nontransgenics. However, there are also many genes that behave differentially between the transgenic and pregnant glands. Understanding these similarities and differences may reveal novel mechanisms by which to develop therapies to treat and prevent human breast cancer.

III. Conclusions

LH hypersecretion in LHβCTP transgenic mice leads to the development of a diverse spectrum of physiological pathologies. Modeling ovarian tumorigenesis, granulosa cell tumors, which are strain dependent, develop in transgenics by 5 months of age and are preventable when animals are exposed to regular hCG surges. Modeling perimenopausal reproductive aging and POF, primordial follicles are depleted by 45% in LHβCTP transgenics at 5 weeks of age. While oocyte meiotic competency is acquired normally, elevated LH increases the rate of meiotic segregation defects, which contributes to nondisjunction-related aneuploidy in females. Modeling infertility and PCOS, transgenics are infertile due to anovulation and undergo pregnancy failure at midgestation following superovulation. Transgenics exhibit a lack of uterine receptivity, despite the induction of pseudopregnancy through mating. Modeling Cushing’s syndrome, adrenal glands from transgenics exhibit adrenocortical hyperfunction, with increased production of corticosterone following induction of LH receptor in response to elevated circulating LH. Modeling breast cancer, chronic exposure to elevated LH leads to mammary gland tumorigenesis.

The diversity of physiological systems ultimately impacted by LH hypersecretion is revealed through the LHβCTP transgenic mouse model. Future studies investigating the downstream signaling pathways and molecular mechanisms involved in the generation of these LH-induced pathologies should support the
notion that the LHβCTP transgenic mouse model may be a useful resource for both transcriptome and proteome profiling.

ACKNOWLEDGMENTS

We would like to thank Dr. Christine Quirk and Tehnaz Parakh for their advice and help. We also acknowledge support from the National Institutes of Health, grant CA 086387 (J.H.N.), and the Department of Defense, grant DAMD 17-01-1-0195 (R.A.K.).

REFERENCES


Apter D, Reinila M, Vihko R 1989 Some endocrine characteristics of early menarche, a risk factor for breast cancer, are preserved into adulthood. Int J Cancer 44:783–787


Dufau ML 1998 The luteinizing hormone receptor. Annu Rev Physiol 60:461–496


Hassold T, Chiu D 1985 Maternal age specific rates of numerical chromosome abnormalities with special reference to trisomy. Hum Genet 70:11–17


Kananen K, Riliaanawati K, Paukku T, Markkula M, Rainio EM, Huhtanen I 1997 Suppression of gonadotropins inhibits gonadal tumorigenesis in mice transgenic for the mouse inhibin


Kero J, Mikola M, Nilson JH, Keri RA, Poutanen M, Huhtaniemi I 2002 High levels of luteinizing hormone stimulate gonadal and adrenal tumorigenesis in mice transgenic for the mouse inhibin/α-subunit promoter/Simian virus 40 T-antigen fusion gene. Oncogene, in press


Mahendroo MS, Cala KM, Landrum CP, Russell DW 1997 Fetal death in mice lacking 5α-reductase type I caused by estrogen excess. Mol Endocrinol 11:917–927

Mann RJ, Keri RA, Nilson JH 1999 Transgenic mice with chronically elevated luteinizing hormone are infertile due to anovulation, defects in uterine receptivity, and midgestation pregnancy failure. Endocrinology 140:2592–2601

Nilson JH, Abbud RA, Keri RA, Quirk CC 2000 Chronic hypersecretion of luteinizing hormone in transgenic mice disrupts both ovarian and pituitary function, with some effects modified by the genetic background. Rec Progr Horm Res 55:69–91
Owens GE, Keri RA, Nilson JH 2002 Ovulatory surges of hCG prevent hormone-induced granulosa cell tumor formation leading to the identification of tumor-associated changes in the transcriptome. Mol Endocrinol 16:1230–1242
Regan L, Owen EJ, Jacobs HS 1990 Hypersecretion of luteinising hormone, infertility and miscarriage. Lancet 336:1141–1144
Risma KA, Hirshfield AN, Nilson JH 1997 Elevated luteinizing hormone in prepubertal transgenic mice causes hyperandrogenemia, precocious puberty, and substantial ovarian pathology. Endocrinology 138:3540–3547
Shoham Z, Jacobs HS, Insler V 1993 Luteinizing hormone: its role, mechanism of action, and detrimental effects when hypersecreted during the follicular phase. Fertil Steril 59:1153–1161
Spackman DH, Riley V 1978 Corticosterone concentrations in the mouse. Science 200:87
Trout SW, Seifer DB 2000 Do women with unexplained recurrent pregnancy loss have higher day 5 serum FSH and estradiol values? Fertil Steril 74:335–337
AUTHOR INDEX

A
Aronow, Bruce, 263

B
Barrett, J. Carl, 325
Brasier, Allan R., 95
Bush, Craig R., 155

C
Chen, Lu, 155
Chi, Ping, 155
Copland, John A., 25

D
Dasen, Jeremy S., 249
Dasu, M.R.K., 75
Davies, Peter J., 25
Dunn, Sandra E., 325
Dupont, Joelle, 325

F
Faustman, Denise L., 131
Fields, Alan P., 155
Finnerty, Celeste, 155

G
Graves, Paul R., 1
Guo, Huiping, 155
Gustafson, William C., 155

H
Handwerger, Stuart, 263
Harris, Jessica, 297
Hayashi, Takuma, 131
Haystead, Timothy A.J., 1
Herndon, D., 75
Horsemann, Nelson D., 297
Huang, Erich, 55

J
Johnson, Betty H., 175
Ju, Bong Gun, 249

K
Kelly, Paul A., 297
Keri, Ruth A., 343

L
LeRoith, Derek, 325
Lindeman, Geoffrey J., 297
Luxon, Bruce A., 25, 155

377
AUTHOR INDEX

M
Mann, Rachel J., 343

N
Naylor, Matthew, 297
Nesic, O., 75
Nevins, Joseph R., 55
Nilson, John H., 343
Nordeen, Steven K., 199

O
Olson, Lorin E., 249
Ormandy, Christopher J., 297

P
Perez-Polo, J.R., 75

R
Reya, Tannishtha, 283
Robertson, Fiona, 297
Rosenfeld, Michael G., 249

S
Shipley, Gregory L., 25
Svrakic, N.M., 75

T
Thompson, E. Aubrey, 155
Thompson, E. Brad, 175
Tian, Bing, 95
Tollkuhn, Jessica, 249

U
Urban, Randall J., 25

V
Visvader, Jane, 297

W
Wan, Yihong, 199
Wechsler-Reya, Robert J., 227
West, Mike, 55
Wood, Christopher G., 25
SUBJECT INDEX

A
Adrenal disturbances, 343
Affymetrix
gene chip technology, 155
oligonucleotide array analysis, 199
Aging, 343
Alveolar bud formation, 297
Apoptosis
defective, 131
glucocorticoid-evoked, 175
Autoimmune diseases, 131
Autoregulation, nuclear factor kappa B, 95

B
Basal gene expression, CEM-C1 subclones, 175
Biochips, 75
Biomedical applications, clustering, 75
Breast cancer
estrogen receptor, 55
metagene analysis, 55
modeling of, 343
progestins and, 199

C
Cancer progression, insulin-like growth factor-1-regulated genes and, 325
Categorical reprogramming, in trophoblast differentiation, 263
Cellular phenotypes, 55
CEM cell lines, 175
Cerebellar granule cells, 227
Cerebellum, gene expression in, 227
Chemokine expression, neutrophilic, 95
Clone, glucocorticoid-resistant, 175
CreloxP, retroviral promoter-trapping system, 199
Cushing’s syndrome, 343

D
Database searching, protein identification, 1
Diabetes, type 1, 131
DNA microarray, 25
chip analysis, 75
villous cytotrophoblast differentiation, 263

E
Electrophoresis, gel, 1
Eosinophilic chemokine expression, 95
Estrogen receptor, breast cancer, 55

F
Follicle development, 343

G
Gel electrophoresis, 1
Gene-array analysis, glucocorticoid effects on leukemic cells, 175
Gene array technology, 25, 199
Gene Chips®, 75, 155
Gene-clustering algorithms, 75
Gene expression, 25, 75
acute lymphoblastic leukemic cell line, 175
elicited by glucocorticoids and progestins, 199
during human trophoblast differentiation, 263
insulin-like growth factor-1-receptor signaling and, 325
medulloblastoma, 227
normal and malignant cerebellum, 227
profilin, 55, 227
Gene induction, insulin receptor and insulin-like growth factor-1 receptor, 325
Gene regulation, glucocorticoids and progestins, 199
Genetic networks, 75, 95
Genome-scale gene expression, 55
Genomic analysis, glucocorticoid-regulated promoters, 155
sequences, target genes, 155
Glucocorticoid apoptosis evoked by, 175
gene expression elicited by, 199
-regulated probesets, 155
-regulated promoters, T-lymphoma cells, 155
-resistant clone, 175
Granule cells cerebellar, 227
precursor proliferation, 227
Granulocyte-macrophage colony, impaired formation, 131
Groucho-related corepressor, 249

Hematopoietic stem cells, self-renewal, 283
Hesx1, 249
Hormone-inducible promoters, 199
Hormone regulation, 199
Hypothalamic-pituitary-gonadal axis, luteinizing hormone and, 343

Infertility, 343
Insulin, structure of, 325
Insulin-like growth factor-1 receptor, 325

Leukemia, 175
Lobuloalveolar development, 297
Luteinizing hormones, diverse physiological systems and, 343
Lymph node metastasis, breast cancer, 55
Lymphoid cells, CEM, 175
Lymphoid leukemia, glucocorticoid-resistant, 175

Major histocompatibility complex region, 131
Mammary ducts, 297
glands, prolactin receptor knockout mice, 297
tumorigenesis, 343
Mass spectrometry, 1
Medulloblastoma, 227
Metagene analysis, breast cancer, 55
Microarray technology, 25, 75

Neutrophilic chemokine expression, 95
Nuclear factor kappa B, 95

Oocyte development, 343
Organogenesis, pituitary, 249
Ovarian failure, premature, 343
tumorigenesis, modeling of, 343
Perimenopausal reproductive aging, 343
Phenotypes, cellular, 55
Phosphoproteome analysis, 1
Pit-1, 249
Pituitary
development, 249
multiple patterning and growth defects, 249
organogenesis, 249
Placental differentiation, 263
Polycystic ovarian syndrome, 343
Precursor proliferation, granule cell, 227
Pregnancy, mammary gland defects, prolactin receptor knockout mice, 297
Principal component analysis, 75
Progenitor cells, Wnt signaling in development, 283
Progesterone, 297
Progestins, gene expression elicited by, 199
Prolactin receptor, 297
release, 297
Promoter analysis, nuclear factor kappa B-inducible genes, 95
glucocorticoid-regulated, 155
hormone-inducible, 199
Prop-1, 249
Proteasome, 131
Protein identification, 1
phosphorylation sites, 1
radiolabeling, 1
Proteomics, 1

Real-time quantitative reverse-transcription polymerase chain reaction, 25
Retroviral promoter-trapping system, Cre/loxP, 199
Rpx-Hesx1, in pituitary development, 249

Self-organizing maps, 75
Signal transduction, 1
via insulin receptor and insulin-like growth factor-1 receptor, 325
Singular value decomposition, 75
Smooth muscle cell physiology, 1
Stem cells hematopoietic, 283
Wnt signaling in development, 283
Subproteomes, 1

Time-course experiments, clustering of, 75
Tissue samples, microarray analysis of, 25
T-lymphoma cells, 155
Transcriptional regulator, 95
response, mammary, 297
Trophoblast differentiation, 263
Tumor necrosis factor alpha, apoptosis induced by, 131

Villous cytotrophoblast differentiation, 263
Viral infection, nuclear factor kappa B-dependent gene network in, 95

Wnt signaling, 283

Yeast genetics, 1