

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^{*†}

**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 μ M dexamethasone (dex) in the presence and absence of 1 μ g/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^{-9} . 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 ACA_nnnTGT_nCT, 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 μ M dex in the presence or absence of 1 μ g/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-

matrix 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 (r^2), 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

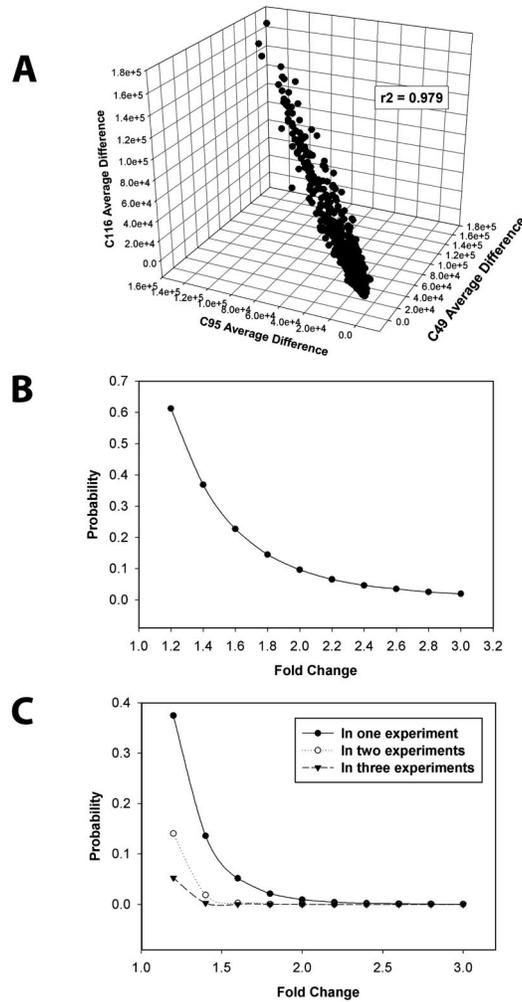


FIG. 1. Statistical analysis of average-difference data from control probesets. As described in the text, 3170 probesets were selected as controls, based on presence in all 12 chips and lack of response to dexamethasone (dex). Panel A contains XYZ linear regression of 3170 probesets from the control 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 $(0.096)^4$ 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 experiment. The probability that a given probeset will behave in this fashion in three independent experiments, as a result of random variation, is $\approx 7 \times 10^{-9}$, 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 \approx 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 ($D = \text{absent}$). 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

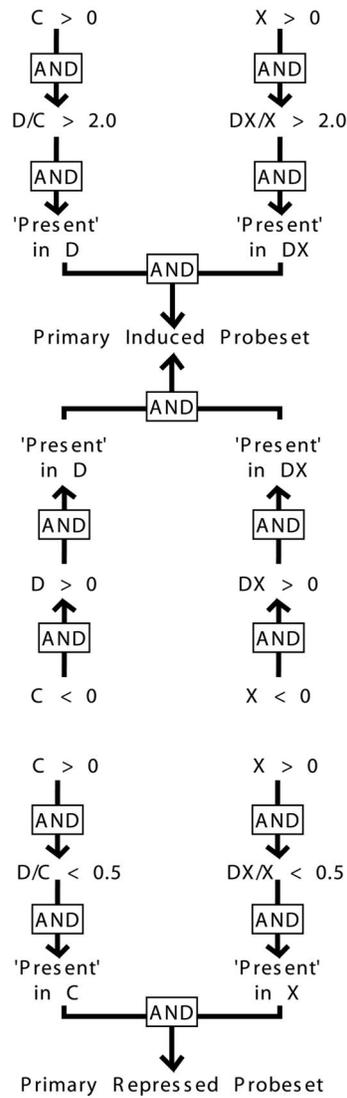


FIG. 2. Graphic representation of the SQL query used to identify induced and repressed probesets. The rationale for this query is discussed in the text and examples of the queries are given in Section IV ("Materials and Methods").

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 $\Delta\Delta C_t$ 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 $ACA_{nnn}TGTnCT$ 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

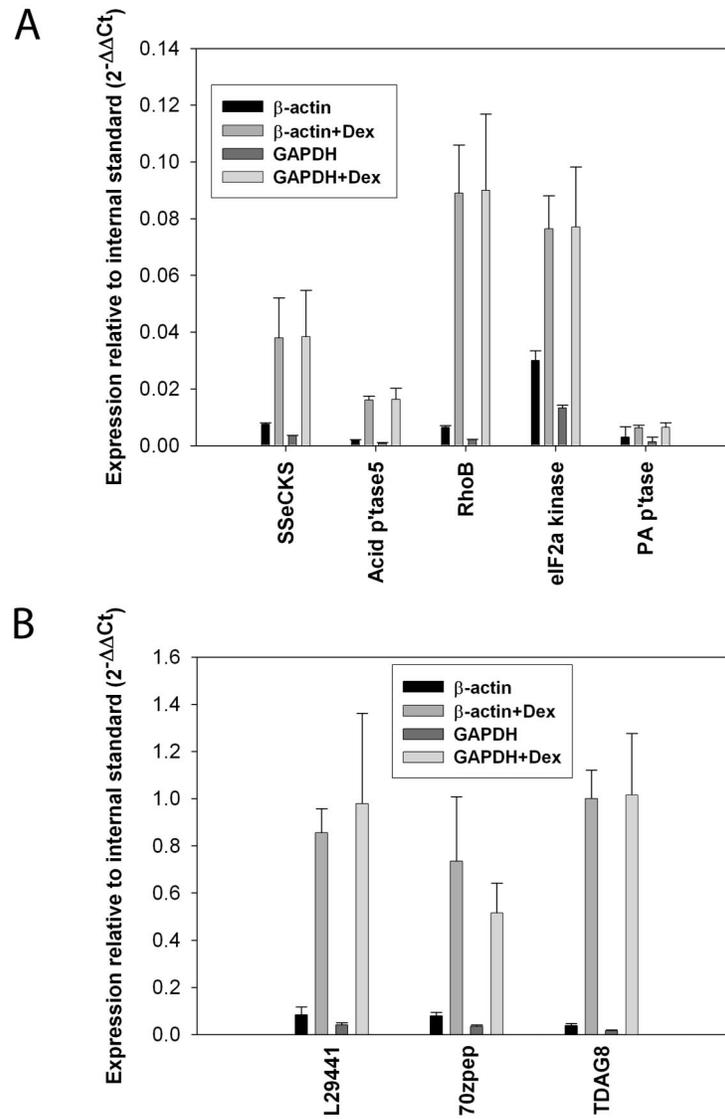


FIG. 3. Real-time polymerase chain reaction (PCR) analysis of glucocorticoid-induced genes. TaqMan[®] analysis was used to measure mRNAs corresponding to the eight glucocorticoid-induced genes shown in Table I. Experiments were performed in triplicate and $\Delta\Delta\text{Ct}$ values calculated relative 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 I
Analysis of Potential Glucocorticoid Receptor Binding Sites Within 4 kb of the Transcription Start Site of Glucocorticoid Target and Control Genes

	Description	Sequence	Location	Strand
Induced probeset				
95022_at	AB020886: SSeCKS	ACAGCTTGTGCT	-2286	+
100151_at	L29441: Overexpressed in testicular tumor		No	
92356_at	M90388: Protein tyrosine phosphatase (70zpep)	ACACCTTGTCT	+1349	-
98859_at	M99054: Acid phosphatase 5, tartrate resistant	ACAGCTTGTCT	-1097	+
96553_at	U39827:G protein-coupled receptor TDAG8		No	
101030_at	X99963: rhoB	ACAATATGTAC	-3721	-
94941_at	AJ243533: GCN2 eIF2alpha kinase		No	
98508_s_at	D84376: Phosphatidic acid phosphatase	ACAAAATGTACT	+3575	+
Control probeset				
100144_at	X07699: Nucleolin	ACAGTCTGTGCT	+1661	+
		ACACCTTGTACT	+1333	-
100156_at	D26090: CDC46		No	
100131_at	X15830: Secretory granule neuroendocrine protein 1		No	
100600_at	M58661: CD24a		No	
104606_at	M55561: CD80	ACAAGCTGTCT	-648	+
94837_at	U67328: NIPI-like protein (NIPIL (A3))		No	
94892_r_at	M27938: Male-enhanced antigen 1		No	
94896_at	D90151: CArG-binding factor-A		No	

[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.]

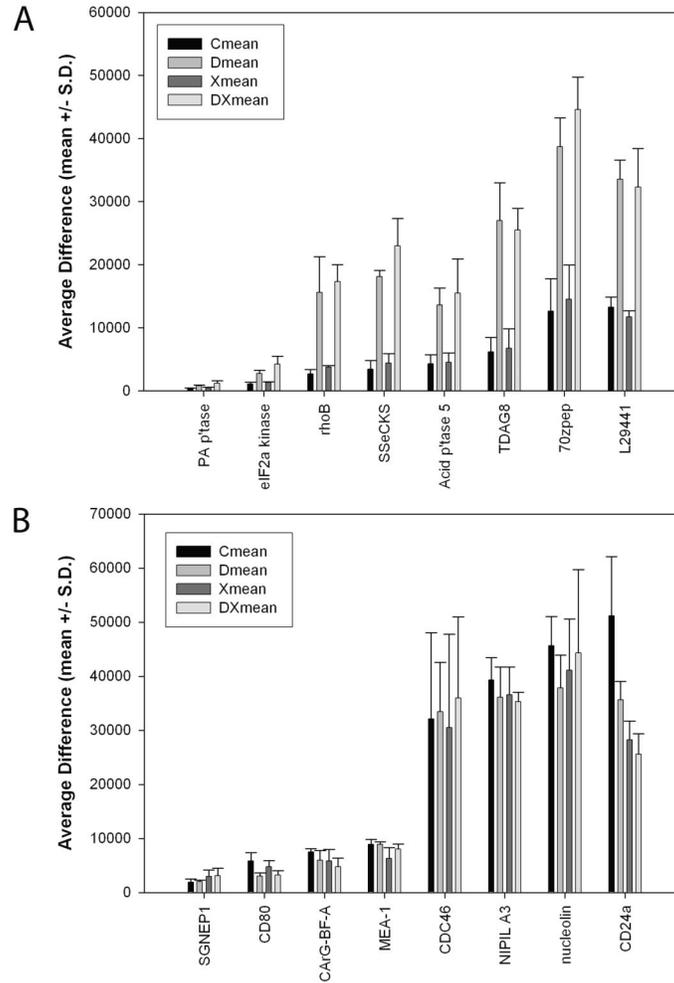


FIG. 4. Expression data for glucocorticoid-induced and control promoters. Using the Celera database, we were able to recover detailed and complete promoter sequence for eight glucocorticoid-induced genes and eight control genes, which were selected more or less at random. The mean and standard deviation (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; CArG-BF-A, CArG-binding factor-A; MEA-1, male enhanced antigen 1; CDC46; NIPIL A3, NIPI-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 probe will respond in a predictable fashion. In our experiments, the probability that any pair-wise comparison of a single probe will yield two numbers that vary by > 2 -fold is almost 10%. Thus, for a given probe, the probability that $D/C > 2.0$ or < 0.5 is ≈ 0.1 . Likewise, the probability that $DX/X > 2.0$ or < 0.5 is ≈ 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 1×10^{-4} . For three independent experiments, the probability would be $1 \times 10^{-4} \times 0.1 \times 0.1$ or 1×10^{-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 glucocorticoid-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 (PKC α), which was induced by glucocorticoids (data not shown). PKC α 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) PKC α , which is stimulated by diacylglycerol; and 3) SSeCKS, which, when phosphorylated by PKC α , 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 $\mu\text{g/ml}$ 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' GGCCAGTGAATTGTAATACGACTCACTATA-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 morpholinoethane 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):

```
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 $d49/c49 \geq 2$ and $d49ac = 2$ and $dx49ac = 2$); The following query was used to combine induced probesets in all three experiments:

```
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:
```

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

D. REAL-TIME PCR

Applied Biosystems assays-by-design 20 \times 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 ΔC_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 $^{\circ}$ C for 30 minutes, AmpliTaq activation 95 $^{\circ}$ C for 10 minutes, denaturation 95 $^{\circ}$ C for 15 seconds, and annealing/extension 60 $^{\circ}$ C for 1 minute (repeat 40 times) on ABI7700. Triplicate C_T values were analyzed using the comparative $C_T(\Delta\Delta C_T)$ 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

- Beato M** 1987 Induction of transcription by steroid hormones. *Biochim Biophys Acta* 910:95–102
- Choi JW, Lee SY, Choi Y** 1996 Identification of a putative G protein-coupled receptor induced during activation-induced apoptosis of T cells. *Cell Immunol* 168:78–84
- Diamond MI, Miner JN, Yoshinaga SK, Yamamoto KR** 1990 Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. *Science* 249:1266–1277
- Gronemeyer H** 1992 Control of transcription activation by steroid hormone receptors. *FASEB J* 8:2524–2529
- Groner B, Ponta H, Beato M, Hynes NE** 1983 The proviral DNA of mouse mammary tumor virus: its use in the study of the molecular details of steroid hormone action. *Mol Cell Endocrinol* 32:101–116
- 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
- Lin X, Tombler E, Nelson PJ, Ross M, Gelman IH** 1996 A novel src- and ras-suppressed protein kinase C substrate associated with cytoskeletal architecture. *J Biol Chem* 271:28430–28438
- Lin X, Nelson PJ, Gelman IH** 2000 SSeCKS, a major protein kinase C substrate with tumor suppressor activity, regulates G(1)-S progression by controlling the expression and cellular compartmentalization of cyclin D. *Mol Cell Biol* 20:7259–7272
- Liu A, Cerniglia GJ, Bernhard EJ, Prendergast GC** 2001 RhoB is required to mediate apoptosis in neoplastically transformed cells after DNA damage. *Proc Natl Acad Sci USA* 98:6192–6197
- Miner JN, Yamamoto KR** 1991 Regulatory crosstalk at composite response elements. *Trends Biochem Sci* 16:423–426
- Miner JN, Yamamoto KR** 1992 The basic region of AP-1 specifies glucocorticoid receptor activity at a composite response element. *Genes Dev* 6:2491–2501
- Pittner RA, Fears R, Brindley DN** 1985 Interactions of insulin, glucagons and dexamethasone in controlling the activity of glycerol phosphate acyltransferase and the activity and subcellular distribution of phosphatidate phosphohydrolase in cultured rat hepatocytes. *Biochem J* 230:525–534
- Rhee K, Reisman D, Bresnahan W, Thompson EA** 1995 Glucocorticoid regulation of G1 cyclin-dependent kinase genes in lymphoid cells. *Cell Growth Diff* 6:691–698
- Ringold GM, Dobson DE, Grove JR, Hall CV, Lee F, Vannice JL** 1983 Glucocorticoid regulation of gene expression: mouse mammary tumor virus as a model system. *Recent Prog Horm Res* 39:387–424
- Snyder JM, Mendelson CR, Johnston JM** 1981 The effect of cortisol on rabbit fetal lung maturation *in vitro*. *Dev Biol* 85:129–140
- Suzuki N, Suzuki T, Uchida A, Thompson EA, Hosoya T** 1992 Effect of dexamethasone on nucleolar casein kinase II activity and phosphorylation of nucleolin in lymphosarcoma P1798 cells. *J Steroid Biochem Mol Biol* 42:305–312
- 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
- Tonko M, Ausserlechner MJ, Bernhard D, Helmberg A, Kofler R** 2001 Gene expression profiles of proliferating vs G1/G0 arrested human leukemia cells suggest a mechanism for glucocorticoid-induced apoptosis. *FASEB J* 15:639–699
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW** 1995 Serial analysis of gene expression. *Science* 270:368–369

- Vieira PL, Kalinski P, Wierenga EA, Kapsenberg ML, de Jong EC** 1998 Glucocorticoids inhibit bioactive IL-12p70 production by *in vitro*-generated human dendritic cells without affecting their T cell stimulatory potential. *J Immunol* 161:5245–5251
- Wassler MJ, Foote CI, Gelman IH, Shur BD** 2001 Functional interaction between the SSeCKS scaffolding protein and the cytoplasmic domain of β 1,4-galactosyltransferase. *J Cell Sci* 114:2291–2300