Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors

Peter C. Hollenhorst, David A. Jones and Barbara J. Graves*

Department of Oncological Sciences, Huntsman Cancer Institute, 2000 Circle of Hope, University of Utah, Salt Lake City, UT 84112, USA

Received September 16, 2004; Revised and Accepted October 7, 2004

ABSTRACT

Sequence-specific DNA binding proteins that function as transcription factors are frequently encoded by gene families. Such proteins display highly conserved DNA binding properties, yet are expected to retain promoter selectivity. In this report we investigate this problem using the ets gene family, a group of metazoan genes whose members regulate cell growth and differentiation and are mutated in human cancers. We tested whether the level of mRNA can serve as a specificity determinant. The mRNA levels of the 27 paralogous human ets genes were measured in 23 tissues and cell lines. Real-time RT–PCR provided accurate measurement of absolute mRNA levels for each gene down to one copy per cell. Surprisingly, at least 16 paralogs were expressed in each cell sample and over half were expressed ubiquitously. Tissues and complementary cell lines showed similar expression patterns, indicating that tissue complexity was not a limitation. There was no unique, highly expressed gene for each cell type. Instead, one of only eight ets genes showed the highest expression in all samples. DNA binding studies illustrate both overlapping and unique specificities for ubiquitous ETS proteins. These findings establish the parameters of the promoter specificity dilemma within the ets family of transcription factors.

INTRODUCTION

Expansion of the repertoire of functional gene products during evolution has relied upon conservation of protein domains. Consequently, in many eukaryotic genomes, relatively large gene families encode proteins that have highly conserved domains. The functional redundancies of these domains bring into question how individual proteins can participate in biological regulation. The activity of each family member in an individual cell depends on both its molecular properties and relative expression level. Therefore, a catalog of the expression levels of each family member is a necessary backdrop for answering the question of biological specificity.

Specificity is a particularly vexing issue in a gene family that encodes DNA binding transcription factors. The conserved DNA binding domain directs the protein to transcriptional targets, a process that represents the most critical route to specific biological function. The ets gene family illustrates this dilemma. These metazoan genes encode proteins with a well characterized DNA binding domain, termed the ETS domain (1,2). Structural studies illustrate that the mode of DNA binding is strongly conserved among ETS domains and indicate that amino acid sequence differences do not dramatically alter the DNA–protein interface (1). Indeed, site selection experiments with 12 different ETS domains reveal that each prefers a consensus sequence with the same core motif, 5'-GGA(A/T)-3', and additional preferences outside this core often show similarity (2,3). Although preference for sequences flanking this core motif can distinguish some family members in in vitro DNA binding assays, these sequences may not preclude the binding of any ETS protein in vivo. Based on this functional similarity, we propose that multiple ETS proteins could recognize a 5'-GGA(A/T)-3' motif within a particular promoter.

The functional diversity of ETS proteins suggests that target site selection is critical for biological regulation. There are 26 paralogous ets genes in the mouse genome (4), 8 in Drosophila (5) and 10 in Caenorhabditis elegans (6). The human genome has 27 human ets genes, including an apparent ortholog of every mouse ets gene, plus TEL2. The ets genes are subdivided into subgroups by a sequence comparison within the predicted ETS domain (2,7), see also Figure 2. Outside the ETS domain, there is significant sequence divergence, allowing ETS proteins to be either activators or repressors and to respond uniquely to signaling pathways (1,8). The diverse functions of ets family members are also revealed in genetic studies in mouse (1), Drosophila (5) and C.elegans (9), in which mutation of individual ets genes causes distinct phenotypes. In spite of considerable evidence for non-redundant function, biological roles of ets genes are linked to regulation of specific genes only in a few cases.

The targeting of an ETS protein to a specific promoter depends on the active protein concentration and the affinity for the promoter. The apparent affinity is determined by intrinsic affinity for the sequence of the binding site as well as interactions with other proteins, with both processes subject to regulation by post-translational modifications (1,2,10,11). Cooperative DNA binding offers a particularly attractive mechanism to facilitate promoter selectivity. However, the potency of this regulatory strategy can vary. For example, whereas GABPα/GABPβ and PU.1/PIP1 partnerships appear...
specific (12–14), SRF can bind DNA cooperatively with three ETS proteins (ELK1, SAP1, and NET) (1), and PAX5 with five (ETS1, FLI1, GABPα, NET, and ELK1) (15,16). Such promiscuities indicate that deciphering promoter selectivity requires a more global understanding of ets biology, including a comprehensive tally of which ets genes are expressed and which targets are regulated in any cell type.

To catalog the expression pattern of each ets gene, a quantitative RT–PCR strategy was developed that measured absolute levels of ets mRNA in 23 human tissues and cell lines. Each cell sample contained mRNAs for approximately two-thirds of the 27 human ets paralogs. About half of the ets genes were expressed in all cell types examined and, thus, were classified as ubiquitous. Our results placed previously reported tissue specificities into a broader context and uncovered additional cell type specific expression. These findings highlight the severity of the promoter selectivity problem in the ets gene family and provide direction for its investigation.

MATERIALS AND METHODS
PCR primers and templates
Gene specific oligonucleotides used for RT–PCR (Table 1) were designed by the Primer III software at http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi (17). Default parameters were used except a GC clamp of 2 bp. Primers were designed to flank an intron to detect any genomic DNA contamination. In genes that are alternatively spliced, have alternate transcription start sites, or display alternative polyadenylation sites, primers were designed to amplify a common element in all alternative products. Blast searches were used to ensure that primers were specific for each individual ets gene. Primer sets amplified a single product of the correct size from a complex cDNA pool [total human RNA reverse transcribed with a mixture of random hexamers and oligo-d(T)] as measured by melting curve analysis on the light cycler system (Roche) and agarose gel electrophoresis.

Only one primer set, FEV, displayed multiple products. Sequencing revealed templating from FEV cDNA and ERG cDNA. We were unable to identify a better primer set possibly due to high GC content around the only two introns in FEV. Only the FEV product was observed in prostate and small intestine. The ERG product was observed in endothelial cells.

Gene specific PCR products from RT–PCR were cloned into the Smal site of puc19. Purified plasmid DNA was linearized, gel purified, then quantified by UV spectrometry to generate reagents for standard curves. One plasmid concentration from each standard curve was compared by real-time PCR with primers specific to the puc19 backbone. Consistent UV

Table 1. Primers used for real-time PCR

<table>
<thead>
<tr>
<th>Locus</th>
<th>Other names</th>
<th>5′ Primer</th>
<th>3′ Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S rRNA</td>
<td></td>
<td>GGTTGAATTTTGGGACCCGGGC</td>
<td>GACCTTTGTTTCCCCGAGAAGGC</td>
</tr>
<tr>
<td>EL1A</td>
<td>ETV4, PEA3</td>
<td>CCCTCTAACACCCAGCTTGCG</td>
<td>GAGAGCCCTCCTGTTTGAGA</td>
</tr>
<tr>
<td>EL1F</td>
<td></td>
<td>TCAAGGAAAAAGGGGAGCTGCG</td>
<td>CACCCAGAAATGCGCACCC</td>
</tr>
<tr>
<td>ELK1</td>
<td></td>
<td>ATTCACCTTCCGACACCGGG</td>
<td>TTACCTCGTAGGTAGCTGCG</td>
</tr>
<tr>
<td>ELF4</td>
<td></td>
<td>CCGTGGGACCGACGCGGC</td>
<td>GCTTTTCCTCGGCGGCGC</td>
</tr>
<tr>
<td>ELF2</td>
<td></td>
<td>CCAACGACCAAACAGAAGGG</td>
<td>CGGAGAAGGGGTCTTGAGC</td>
</tr>
<tr>
<td>ESE2</td>
<td></td>
<td>GCACTAAGATGTTTGAGAAG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>ESE1</td>
<td>ERF</td>
<td>GGCATCTTCCTGCTTGACG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>ESE2-b</td>
<td></td>
<td>GCACTACCTCATCGAAATGGG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>ESE3</td>
<td></td>
<td>ACCACGACCAAACAGAAGGG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>ETS1</td>
<td></td>
<td>CAAGGTGTTGATGAGAAGGG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>ETS2-a</td>
<td></td>
<td>CACCGTGATGAGAAGGGG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>ETS2-b</td>
<td></td>
<td>CAGGTGTATGAGAAGGGG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>ELF2</td>
<td></td>
<td>AAAGGGGCGGACGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>GABPA, E4F, ELG</td>
<td></td>
<td>CAGCTGCTCTGGACGCTGG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>MEF</td>
<td>ETV4, ELF4</td>
<td>GCACAGCCATCTACTCTGGG</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>SAP2, ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>SAP2, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
<tr>
<td>NET</td>
<td>ELK3, ERP</td>
<td>TCCGCTCTGCTCTGACGAC</td>
<td>GACCTGCTCTGGACGCTGG</td>
</tr>
</tbody>
</table>

**Table 1.** Primers used for real-time PCR

*a* Gene loci marked with a and b indicate two alternate primer sets used to measure mRNA levels of the same gene. Underlined names indicate HUGO ID (62).

*b* Alternative names for human gene or mouse ortholog. Gene names used in this report were selected based on common use or to differentiate subgroups.

*c* 5′ and 3′ indicate gross position relative to direction of transcription. All sequences are given in 5′ to 3′ orientation.

*d* Primers were used for reverse-transcription except where denoted with an asterisk.
spectrometry and DNA dilution was judged by <10% variation between a sample point from each standard curve.

cDNA preparation

RNA was prepared from cell lines and primary umbilical vein endothelial cells by Trizol extraction according to instructions (Invitrogen). BD Biosciences Clontech provided whole tissue total RNA. These tissue RNAs were pooled from between 3 and 45 individuals with the exception of heart, brain and stomach, which were individual samples. Reverse transcription reactions were performed at 55°C with Superscript III reverse transcriptase (Invitrogen) according to the instructions except that no dithiothreitol was used. Each reaction included sequence specific primers for &lt;14 ets genes plus 18S rRNA to create a cDNA pool. An aliquot of 1 pmol of the most 3' primer for each gene was included in the reaction. In controls using a subset of ets mRNAs, reverse transcription with alternative 3' primers revealed minor efficiency differences (standard errors &lt;5% of the mean). Reverse transcriptase processivity was observed over a distance of up to one kilo base, the longest distance from the reverse primer that was used for PCR. cDNA products were further processed by digesting RNA for 20 min with 4 U of RNase H (Fermentas), and then purified using a PCR cleanup kit (Qiagen). For the purposes of quantification, we assumed that RNA preparation and reverse transcription was 100% efficient.

Real-time PCR

Real-time PCR was performed with the LightCycler FastStart DNA Master SYBR Green I system (Roche). PCR was performed according to instructions with 3 mM MgCl2, an annealing time of 5 s and an extension time of 12 s. Annealing temperature was 63°C, except for assays with primer sets ETS2-a, ER81-a, SAP1 and ESE3-a that were performed at 57°C. Each cDNA assay included a primer set and the cDNA template, derived from 30 ng of total RNA. Each experiment included a minus-template control and five assays to create a standard curve that contained 10^{-3}, 10^{-2}, 10^{-1}, 10^0 and 10^1 copies of the gene specific linear plasmid as template. All primers generated standard curves with excellent linear fits (R values = 1.00) and showed a single sharp melting peak. cDNA levels in each sample were measured using the Fit Points Method of the LightCycler Software. A noise band was set in the log-linear phase of each sample curve. The software plotted the cycle number of the crossing point of each standard versus the copy number present in the standard. The copy number of each cDNA sample was extrapolated from this standard curve. Simple repetitions of a subset of measurements revealed excellent reproducibility with standard errors that averaged &lt;5% of the mean and never exceeded 15%. In light of the ~2-fold error inherent in this assay (Figure 1), we judged this experimental repetition with its minimal error to be unnecessary. Three criteria required for the accuracy shown in Figure 1 included the use of gene specific oligonucleotides rather than random hexamers or oligo d(T) for reverse transcription, the use of plasmids rather than PCR products for standard curves, and the use of primer sets that gave a product with a single sharp peak in the melting curve.

Figure 1. Absolute values for mRNA levels are accurate to one copy per cell. The expression of ETS2, ER81, E1AF, ESE2 and SPIB in 23 cell samples was measured by real-time RT–PCR with gene specific primers (Table 1). Values were converted to mRNA copies per cell by the use of a standard curve and 18S rRNA as an internal control. Each cDNA measurement was repeated with an independent primer set. The fold difference (value 1/value 2) between the two measurements was 2-fold or less in the gray area. Of 61 measurements equal to or greater than one copy per cell, 59 (97%) showed &lt;2-fold error. This error was not significantly different for each individual gene and can therefore be considered gene independent.

Use of 18S rRNA as a standard

The mean of two measurements of the reverse transcribed product of 18S rRNA was used to standardize cDNA copy number for each cell sample. The standard error between these measurements was &lt;1% of the mean. The 18S rRNA copy number per cell was estimated only for cell lines. Cell number was counted using plates prepared in parallel to those used for RNA harvest. The number of cell equivalents present in each real-time PCR reaction was used to calculate 18S rRNA copy number per cell. The 18S rRNA copy number per cell ranged from $4 \times 10^2$ to $4 \times 10^3$ with a mean of $2 \times 10^2$. This estimate was similar to a previous estimate of $3 \times 10^2$ ribosomes in a HeLa cell (18).

Chromatin immunoprecipitation

Crosslinking of $1 \times 10^7$ cells was performed as described previously (19) for 15 min at room temperature. Nuclei were prepared from fixed cells as described previously (20). Nuclei were resuspended in 0.5 ml of sonication buffer (19) and sonicated four (HCT116 cells) or six (Jurkat cells) times for 30 s resulting in chromatin averaging 1000 bp. Chromatin was diluted with 5 ml dilution buffer, 20 mM Tris (pH 7.9), 2 mM EDTA, 150 mM NaCl, 1% Triton X-100 and mammalian protease inhibitors (Sigma). Chromatin was precleared with 300 µl of a 50% slurry of a 1:1 mixture of pre-blocked Protein A and G agarose beads (Upstate Biotechnology) for 1 h at 4°C. One milliliter of precleared chromatin was rotated overnight at 4°C with 10 µl of the rabbit polyclonal antibodies ETS1 (21), ETS2 (22), ETS1/2 (sc-351 Santa Cruz), ELK1 (sc-355 Santa Cruz) or rabbit IgG (Santa Cruz). Chromatin/antibody was rotated with 60 µl of the Protein A/G mixture for 6 h. Agarose beads were washed six times with IP wash buffer, 10 mM Tris (pH 7.9), 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl, 0.05% SDS, 0.25% NP-40. Immunoprecipitated DNA
was prepared and PCR was performed as described previously (23) with the following primers; albumin 5′ ggtacctgctgccagaatc; 3′ gattcctgccacatcagaaag; 5′ EGR1 ctatttgagccacactttgg; 3′ gattcctgccacatcagaaag; CDC2L2 primers as described previously (24). Real-time PCR was used to analyze enrichment of albumin and CDC2L2 DNA using the same primer sets except that 3′ albumin was ctccttgctcctgtccatgct.

RESULTS

Measurement of ets mRNA levels by real-time RT–PCR

To facilitate comparison of mRNA levels across multiple cell samples for multiple genes we measured both absolute and relative levels of mRNA by real-time RT–PCR. Data were tabulated in two dimensions for comparative purposes. The first, termed the ets family profile, compared the relative expression of each ets gene in a single cell type and required the measurement of absolute levels of different cDNAs from the same cell sample. The second dimension, the ets gene profile compared the relative expression of a single ets gene across multiple cell types. To compare different samples, ets mRNA levels were normalized to the levels of 18S rRNA. The reported units, mRNA copy number per cell, assumed a hypothetical cell that contains $2 \times 10^6$ 18S rRNA molecules (see Materials and Methods).

Control experiments established the sensitivity and accuracy of our approach. Real-time PCR is often used to monitor relative changes in the expression level of a single gene (25). In contrast, we needed to measure absolute levels of different genes. To test the accuracy of real-time PCR for absolute measurements, two distinct primer sets for five different ets gene cDNAs interrogated cDNAs from 23 different cell samples. At the level of sensitivity of one copy per cell, the error between primer sets was <2-fold (Figure 1). Lower levels of ets gene expression could be biologically relevant, however, values below this level displayed a dramatic decrease in accuracy and were not reported. Similar controls in yeast studies have revealed the same level of error (2-fold or less) (26). We found several specific experimental criteria that were necessary to reduce error to this level (see Materials and Methods). We are not aware of any previous work using mammalian genes in which these controls were performed, therefore, this report can provide guidelines for the analysis of additional metazoan gene families.

ets family profiles in human cell samples

By measuring the mRNA levels for the 27 human ets genes, ets family profiles were created for 15 tissues and eight cell lines (Figure 2, columns). In the tissues, an average of 22 genes ($\bar{x} = 22 \pm 3$) was expressed (Figure 2, left columns). The cellular complexity of a tissue could cause an overestimation of the number of ets family members expressed. To evaluate the severity of this problem, eight established human cell lines were analyzed (Figure 2, right columns). A similar abundance of ets gene expression was observed with 19 genes detected per cell line ($\bar{x} = 19 \pm 2$). Furthermore, cell lines with a matching tissue sample displayed similar expression patterns (Figure 3). The most highly expressed genes illustrated this trend: ETS1 in thymus and Jurkat cells, SPIB in spleen and Raji cells, ESE1 in colon and HCT116 cells, and PDEF in prostate and PC3 cells. Consistent with the co-expression of many ets genes in one cell type, HUVECs, which are primary endothelial cells lacking tissue-based complexity, expressed 19 ets genes, a total similar to the cell lines. Nevertheless, some cell type diversity is detectable in the tissue samples. For example, endothelial and hematopoietic cells likely reside in all tissues, consistent with the detection of PU.1, ERG, and FLI1 mRNA in all tissues. (These genes were abundantly expressed in these cell types as discussed below). In another example, the mRNA copy numbers for particular genes were usually slightly higher in cell lines than in matching tissues (Figure 3). This increased expression could be either a hallmark of transformed cells or simply due to cell type complexity in tissues diluting apparent mRNA levels. In spite of minor concerns for tissue complexity, the overall similarity between the tissue and cell line data suggested that tissues provided a valid profile of ets gene expression in a particular cell type.

Taken together, the analyses of tissue and cell lines revealed abundant expression of the ets gene family with, on average, 21 ets genes expressed in each cell type ($\bar{x} = 21 \pm 3$). On average, 11 genes ($\bar{x} = 11 \pm 4$) were expressed at levels >10 copies per cell and 1 ($\bar{x} = 1 \pm 1$) expressed at a level >100 copies per cell. Most individual mRNAs in mammalian cells are estimated to be present at levels <10 copies per cell with an upper limit of 500 copies per cell, except for rare cases (27,28). Thus, ets human paralogs are expressed at levels similar to estimates for most cellular mRNAs. Contrary to the simple expectation, there was no unique, predominant ets gene expressed in each cell type. Instead, only eight genes were scored as the highest expressing gene in a cell sample: ETS1, ETS2, ESE1, ESE3, PU.1, E1AF, GABPα and ERG (Figure 2, bold values). Furthermore, ETS1, ETS2 and ESE1 dominated this category. In summary, the co-expression of numerous ets genes in every cell sample indicates that many ETS proteins must be considered potential ETS-binding site regulators in any human system.

ets gene profiles in human cell samples

The distribution of expression of a single gene across diverse cell types (ets gene profile) provides clues to function. With our comprehensive approach, all human family members were analyzed in the same set of cell samples (Figure 2, rows). The ets family profiles predicted that a high number of ets genes would be expressed in all cells. Indeed, 14 of the 27 ets genes were expressed in at least 22 of the 23 cell samples and classified as ubiquitous (Figure 2, summarized in Table 2). The ubiquitously expressed ets genes tended to be in the ETS, ELF, TCF and ERF subgroups. The ubiquitous ets genes varied in expression levels, with some at high levels, such as ETS2 and GABPα (mean expression of 61 and 47, respectively), and some at low levels, such as ELF1 and ERF (mean expression of 12 and 6, respectively) (Figure 2).

Cell type specificity was observed to some degree for 16 of the 27 human ets genes (Figure 2, summarized in Table 2). Some genes, such as SPIB and SPIC, were detected only in a few cell samples. Other genes, such as NET and ER71, were expressed in all cell samples, but were present at higher levels in certain cell types (endothelial cells and testis, respectively).
Some cell type specific genes (ESE1, ESE3, PU.1, ET51 and ERG) were the most highly expressed ets genes in a cell sample. Others, such as ER71 or SPIB, never exceeded the levels of other ets genes in the same sample. These differences may reflect diverse strategies for target promoter recognition or differences in the quantity of in vivo binding sites.

The cell type specificities of many genes were consistent with previous in vivo studies (for literature review, see Table 2). Examples of concordance include the ESE subgroup in certain epithelia containing tissues (7), PDEF in prostate (29), PU.1 in myeloid cells (30), SPIB in B cells (31), SPIC in spleen (32), ER71 in testis (33) and FEV in small intestine and prostate (34). In addition, our findings also correlated with functional tests. A conditional lymphoid deletion of ET51 has a decreased number of B and T cells (35, 36), and ET51 was highly expressed in B and T cells. The NET deletion mouse shows increased target gene expression in the vasculature (37), and this putative repressor was expressed at the highest levels in endothelial cells. Members of the PEA3 subgroup, ER81 and ERM, were expressed at higher levels in testis, lung and brain, consistent with roles in male fertility (38) and in branching morphogenesis in lung and brain (39, 40). One interesting observation was the higher expression of ESE1, ESE3 and PDEF in tissues than in cell lines (Figure 3). This difference may be related to a role for these genes in terminal differentiation rather than cell growth. Indeed, ESE1 is proposed to be critical for intestinal epithelia differentiation based on genetic disruption in the mouse (41). This concordance with previous findings, including functional data, supports the validity of our methods and conclusions.

Our survey of ets family expression also revealed new specificities. For example, members of the ERG subgroup have been reported to be present in a number of cell types, including endothelial cells (42–46). However, ERG was expressed at an

---

**Table 2.** Expression profiles for ets family demonstrate extensive co-expression. The expression of 27 human ets genes was measured by real-time RT–PCR with gene specific primers (Table 1). Horizontal lines separate ets genes into subgroups that are defined by similarity in the ETS domain. The mRNA copy number per cell was estimated as mRNA molecules per 2 × 106 18S rRNA molecules in the same sample. Values <1, indicated with an asterisk, could not be measured accurately (Figure 1). Each column represents values from a single RNA sample. Values for ESE2, ET52, SPIB, E1AF and ER81 are the mean of mRNA level of the same cDNA sample with two independent primer sets. Since simple repetition gave much lower error (see Materials and Methods) than that inherent in the assay (Figure 1) such measurements were not deemed valuable and the error for all values should be assumed to be ~2-fold. Values in bold indicate the most highly expressed ets gene in a cell sample.
extremely high level only in endothelial cells, whereas FLI1 was expressed at high levels in both endothelial and hematopoietic cells. This endothelial specific expression correlated with genetic studies implicating ERG and FLI1 in endothelial cell differentiation (47,48). In a second example, the PEA3 subgroup showed dramatically higher expression in cell lines than in tissues. For example, E1AF showed minimal expression in tissues, but was among the most highly expressed ets genes in certain cell lines. This expression pattern could be explained by the reported expression of the PEA3 subgroup in multiple tumor types (49–51) in conjunction with the fact that the majority of the tested cell lines are derived from tumors.

In vivo ETS DNA binding specificity

The discovery that over one-half of ets genes are ubiquitous brings into question whether each ETS protein has unique targets and whether these targets would change in different cells. Using chromatin immunoprecipitation, we tested the in vivo promoter occupancy of three ubiquitous ETS proteins ETS1, ETS2 and ELK1 on the CDC2L2 and EGR1 promoters with reported specificity for ETS1 (24) or ELK1 (52), respectively. The antibody for ELK1, but not for ETS1 or ETS2, detected the EGR1 promoter in both Jurkat and HCT116 cells (Figure 4). Antibodies for ETS1 and ETS2, but not ELK1, detected the CDC2L2 promoter in both cell lines (Figure 4). These reciprocal findings indicated that specificity is not necessarily reflective of mRNA levels as all three genes are expressed in both cell lines (Figure 2). The presence of both ETS1 and ETS2 at the CDC2L2 promoter had not been previously reported. Interestingly, this dual occupancy does not correlate with the relative mRNA levels for these ets genes in the two cell lines. ETS1 expression is higher than that of ETS2 in Jurkat cells, whereas ETS2 expression is higher than ETS1 in HCT116 (Figure 2). These binding data demonstrate that there is promoter selectivity in spite of extensive co-expression of many ets genes in each cell type. Furthermore, an understanding of ETS protein association with any particular promoter requires consideration of all ETS proteins present in that cell type.
Table 2. Classification of ubiquitous and cell type specific ets genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>U*</th>
<th>Tissue or cell line specificitiesb</th>
<th>Previously reported expressionc</th>
<th>Mouse gene deletion phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELF1</td>
<td>+</td>
<td>None</td>
<td>Many tissues (57,63)</td>
<td></td>
</tr>
<tr>
<td>MEF</td>
<td>+</td>
<td>None</td>
<td>Many tissues (64)</td>
<td></td>
</tr>
<tr>
<td>NERF</td>
<td>+</td>
<td>None</td>
<td>Many tissues (63)</td>
<td></td>
</tr>
<tr>
<td>ELK1</td>
<td>+</td>
<td>None</td>
<td>Many tissues (65,66)</td>
<td>Reduced c-fos expression in brain (67)</td>
</tr>
<tr>
<td>NET</td>
<td>+</td>
<td>HUVEC</td>
<td>Many tissues (65,66), endothelia (68)</td>
<td>Vascular defects, upregulate egr-1 (37)</td>
</tr>
<tr>
<td>SAP1</td>
<td>+</td>
<td>None</td>
<td>Many tissues (65,66)</td>
<td></td>
</tr>
<tr>
<td>ER71</td>
<td>+</td>
<td>Testis</td>
<td>Testis (33)</td>
<td></td>
</tr>
<tr>
<td>ERF</td>
<td>+</td>
<td>None</td>
<td>Many tissues (69)</td>
<td></td>
</tr>
<tr>
<td>PE1</td>
<td>+</td>
<td>None</td>
<td>Many tissues (70)</td>
<td></td>
</tr>
<tr>
<td>ERO</td>
<td>U</td>
<td>HUVEC</td>
<td>Thymus (45), HUVEC (46)</td>
<td>Hemorrhage, Decreased Lymphocytes (48)</td>
</tr>
<tr>
<td>FEV</td>
<td>U</td>
<td>Prostate</td>
<td>Prostate, small intestine (34)</td>
<td></td>
</tr>
<tr>
<td>FLI1</td>
<td>U</td>
<td>HUVEC, HL-60, Jurkat, Raji, lung</td>
<td>Thymus, ovary, bone marrow, spleen, heart (43,44,71)</td>
<td>Intestinal differentiation defects (41)</td>
</tr>
<tr>
<td>ESE1</td>
<td>U</td>
<td>Colon, lung</td>
<td>GI tract, prostate, kidney, ovary, lung, pancreas, liver (72)</td>
<td></td>
</tr>
<tr>
<td>ESE2</td>
<td>U</td>
<td>Kidney</td>
<td>Salivary, mammary, kidney, prostate, lung (73)</td>
<td></td>
</tr>
<tr>
<td>ESE3</td>
<td>U</td>
<td>GI Tract, lung, prostate, kidney, PC3, HCT116</td>
<td>Salivary, prostate, colon, mammary, lung, kidney, pancreas, trachea (7)</td>
<td></td>
</tr>
<tr>
<td>ETS1</td>
<td>U</td>
<td>Lung, Jurkat</td>
<td>Thymus, lung, heart, gut, spleen (2)</td>
<td>Reduced B, T, and NK cells (35,36,74)</td>
</tr>
<tr>
<td>ETS2</td>
<td>U</td>
<td>None</td>
<td>Many tissues (75,76)</td>
<td>Embryonic lethal; placental defects (77)</td>
</tr>
<tr>
<td>GABPa</td>
<td>U</td>
<td>None</td>
<td>Many tissues (33)</td>
<td>Embryonic lethal (78)</td>
</tr>
<tr>
<td>PDEF</td>
<td>U</td>
<td>Prostate, colon, stomach, PC3</td>
<td>Prostate, ovary (29)</td>
<td>Embryonic lethal, no myeloid or lymphoid differentiation (80,81)</td>
</tr>
<tr>
<td>PU.1</td>
<td>U</td>
<td>HL-60, lung</td>
<td>Spleen, testis (30,31,79)</td>
<td>B-cell defects (82)</td>
</tr>
<tr>
<td>SPIB</td>
<td>U</td>
<td>Raji, Spleen</td>
<td>Lymphocytes-particularly mature B cells (31)</td>
<td></td>
</tr>
<tr>
<td>SPIC</td>
<td>U</td>
<td>Spleen</td>
<td>Spleen and lymph nodes (32)</td>
<td>Sterile males (38)</td>
</tr>
<tr>
<td>E1AF</td>
<td>U</td>
<td>HCT116, HeLa, HepG2, HMEC, PC3, HL-60</td>
<td>Tissues in branching morphogenesis (83); Brain, testis (84)</td>
<td>Lack some neuronal connections (39)</td>
</tr>
<tr>
<td>ERM</td>
<td>U</td>
<td>None</td>
<td>Tissues in branching morphogenesis (83); Many tissues (33)</td>
<td></td>
</tr>
<tr>
<td>TEL</td>
<td>U</td>
<td>None</td>
<td>Tissues in branching morphogenesis (83); Brain, placenta, lung, pancreas, heart (85)</td>
<td>Embryonic lethal; yolk sac angiogenesis defect, abnormal apoptosis (87)</td>
</tr>
<tr>
<td>TEL2</td>
<td>U</td>
<td>Lung</td>
<td>Many tissues—different isoforms show different specificities (88)</td>
<td></td>
</tr>
</tbody>
</table>

*U indicates whether expression is ubiquitous (+) based on presence in at least 22 of 23 cell samples.

Summary of tissue or cell line specificities. Tissues or cell lines are listed as specific if expression in that cell sample was at least 8-fold above the median expression of that gene across all 23 cell samples. For purposes of this calculation, asterisks were given a value of one. This is a rather stringent method of calculating cell type specificity in moderately-sized gene families.

**DISCUSSION**

We discovered that over two-thirds of the 27 human ets genes are expressed in most cell samples. This extensive co-expression, in combination with the conservation of the DNA binding domain, emphasizes the challenge of matching a particular ETS protein to a specific promoter. Our findings extend a more focused study on mouse mammary cells that detected mRNA for 24 ets genes in normal tissue and for 14–20 genes in cell lines (4). Furthermore, our conclusions may extend to other transcription factor families where understanding promoter selectivity is important. Examples include the hox and forkhead gene families, each with at least 39 members in humans (53,54). In concordance with the ets gene data, the co-expression of between 8 and 39 human hox genes is detected in 20 human tissues (53). Our real-time RT–PCR experimental design, which accurately measured the range of 1 to >500 copies of mRNA per cell, will facilitate the characterization of other gene families. Our study establishes the ets family as a model system for the study of specificity in moderately-sized gene families.

**Extrapolation of mRNA data to promoter occupancy**

The presence of 16–24 different ets mRNAs in a cell sample provides only a maximum number of ets genes that may regulate transcription in that cell. The ability of an ets mRNA to affect transcription requires translation to an active protein with the proper subcellular localization. New immunological reagents are required to survey ETS protein levels to compare with mRNA expression data. In one available example, there appears to be a convergence between protein and mRNA data. At least eight ETS proteins have been detected in T cells (ETS1, ETS2, ELF1, MEF, ELK1, SAP1, TEL and GABPα) (21,22,55–60). The corresponding mRNA for each of these proteins was present in both the Jurkat and thymus cell samples. The 10 additional genes that we detected as mRNAs in a cell sample contains a relatively high number of ETS proteins.

Ultimately, matching an ETS protein to a promoter in vivo requires an assay such as chromatin immunoprecipitation. To date, no specific promoter has been tested for in vivo association with each ETS protein present in a cell. However, more
limited analyses, such as the immunoprecipitation of the CD68 promoter by ELF1, FLI1 and PU.1 antibodies (61) and our CDC2L2 experiment, indicate that a positive signal from one ETS antibody does not preclude the relevance of other ETS proteins.

Extrapolation of expression profile data to other cells or tissues

Our analysis provided a framework for connecting a specific ETS protein to a target gene within the 23 cell samples tested. Can we extrapolate from these data to predict the likely ets family profiles of other cell samples? The 14 ets genes found to be ubiquitous, must be considered as candidates for binding to an ETS binding site in any cell. This group encompasses more than half of the ets genes. In addition, the most highly expressed ets genes encode good candidate proteins. A prediction can also be made regarding these genes. In a cell sample with an epithelial identity, we predict that ESE1 and ESE3 will be the most highly expressed ets genes. For myeloid cells, PU.1 likely represents the most abundant ets mRNA, whereas for other hematopoietic lineages a good candidate would be ETS1. Endothelial cells would be predicted to express ERG at high levels. For other cell types, the ubiquitously expressed ETS2 and GABPα are likely to be the most highly expressed ets genes.

In summary, expression profiling of all 27 human ets genes in the same 23 cell samples generated an unprecedented picture of the cell type specificity of an entire gene family. The most significant finding is the surprisingly high degree of overlapping expression of ets genes in each cell type. Despite this extensive co-expression, target gene specificity can be maintained, as illustrated by the promoter occupancy analysis of two potential targets. These findings demonstrate that multiple ETS proteins are candidates for any potential promoter target and provide a guide for the type of candidates to be considered in different cell types.

ACKNOWLEDGEMENTS

We thank Utah colleagues Tom McIntyre, Paul Shami, and Diana Stafforini for cell lines. Critical reading of manuscript by Don Ayer, Brad Cairns, and Mario Capecchi is gratefully acknowledged. We acknowledge statistical consultation with the Huntsman Cancer Institute Biostatistics Shared Resource, specifically Dr Aniko Szabo. This work was supported by a postdoctoral fellowship (PF-03-122-01-GMC) from the American Cancer Society to P.C.H., a grant to D.A.J. (RSG02-141-01-CNE) from the American Cancer Society, grants from the National Institutes of Health (T32 CA93247 for fellowship support to P.C.H., GM38663 awarded to B.J.G., and CA24014 awarded to the Huntsman Cancer Institute), and funds from the Huntsman Cancer Foundation.

REFERENCES


