

Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the *ETS* gene family

Peter C. Hollenhorst,¹ Atul A. Shah,¹ Christopher Hopkins,² and Barbara J. Graves^{1,3}

¹Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112, USA;

²Agilent Technologies, Santa Clara, California 95051, USA

The conservation of in vitro DNA-binding properties within families of transcription factors presents a challenge for achieving in vivo specificity. To uncover the mechanisms regulating specificity within the *ETS* gene family, we have used chromatin immunoprecipitation coupled with genome-wide promoter microarrays to query the occupancy of three ETS proteins in a human T-cell line. Unexpectedly, redundant occupancy was frequently detected, while specific occupancy was less likely. Redundant binding correlated with housekeeping classes of genes, whereas specific binding examples represented more specialized genes. Bioinformatics approaches demonstrated that redundant binding correlated with consensus ETS-binding sequences near transcription start sites. In contrast, specific binding sites diverged dramatically from the consensus and were found further from transcription start sites. One route to specificity was found—a highly divergent binding site that facilitates ETS1 and RUNX1 cooperative DNA binding. The specific and redundant DNA-binding modes suggest two distinct roles for members of the ETS transcription factor family.

[*Keywords:* ETS; transcription; gene families; cooperative binding; promoter specificity; ChIP–chip]

Supplemental material is available at www.genesdev.org.

Received April 16, 2007; revised version accepted June 14, 2007.

Transcriptional activators and repressors bind regulatory sequences within promoters and enhancers by engaging sequence-specific DNA-binding sites. Many DNA-binding proteins with well-characterized DNA sequence preferences are now known. Furthermore, sequence motifs that match these preferences are found by bioinformatics-based queries of genomes. However, predicting the in vivo pattern of DNA–protein interactions based on in vitro-determined consensus and bioinformatics-based databases has proven difficult (Wasserman and Sandelin 2004). Genome-wide transcription factor occupancy techniques, such as chromatin immunoprecipitation (ChIP) coupled with DNA microarrays (termed ChIP–chip), can define transcription factor in vivo utilization of genomic sequences and inform bioinformatics approaches. However, most genome-wide occupancy studies in mammalian systems have failed to identify matches to in vitro consensus binding sequences in the majority of occupied regions (Weinmann et al. 2002; Li et al. 2003; Martone et al. 2003; Euskirchen et al. 2004; Krig et al. 2007). In addition, recognizable consensus sites can display no evidence of binding. Several speculations have been presented to explain

this lack of correlation. Protein partnerships could facilitate the use of nonconsensus binding sites by DNA-binding cooperativity. Chromatin structure could occlude some sequences that are predicted to be binding sites. A combination of experimental and bioinformatics approaches are necessary to test these hypotheses at a genomic level and answer the central question: What is the genetic basis for the control of gene expression?

Large gene families in mammalian genomes that encode transcription factors with highly related DNA-binding properties (e.g., ETS, GATA, HOX, or FOX proteins, and nuclear hormone receptors) (Messina et al. 2004) present a further challenge. The dilemma is how transcription factors with overlapping DNA sequence preferences direct distinct transcriptional responses in vivo. The problem itself is poorly characterized because many binding sites in promoters and enhancers have been assayed only for an arbitrary subset of family members. Furthermore, no mammalian ChIP–chip experiments have directly addressed this family conundrum. Thus, it is unresolved whether extensive genetic redundancy is a characteristic of these families or whether robust mechanisms operate that drive specificity.

The ETS family of transcription factors provides an excellent system to pursue these questions due to the extensive knowledge of the biological roles of *ETS* genes and the biochemical properties of ETS proteins (Shar-

³Corresponding author.

E-MAIL barbara.graves@hci.utah.edu; FAX (801) 585-1980.

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.1561707>.

rocks 2001; Oikawa and Yamada 2003). The family is defined by the conserved DNA-binding domain, termed the ETS domain, which bears a winged helix–turn–helix protein fold. Phylogenetic analysis of the 27 human ETS domains identifies subfamilies of more highly related members, termed clades (Fig. 1A). The DNA-binding properties of ETS proteins from all clades are remarkably similar due to the high conservation of amino acids within the ETS domain that are critical for DNA interaction. For example, in vitro site-selection studies performed on 10 ETS proteins each report preference for an invariant GGA core. In addition, five flanking positions also show conservation among these family members (Fig. 1B).

In spite of this high degree of conservation, experimental studies suggest that ETS proteins have unique biological functions. All 14 mouse *ETS* gene disruption strains, including seven targeting ubiquitously expressed *ETS* genes, show unique phenotypes (Hollenhorst et al. 2004; Zhou et al. 2005). In vivo transcription assays demonstrate functional differences with some of the ETS proteins being activators, whereas others are repressors (Kopp et al. 2004). Thus, the remarkable conservation of DNA binding is contrasted with expected diversity of biological function. Interestingly, this predicted specificity must exist in an environment in which multiple ETS proteins are present, because more than half of the 27 human *ETS* genes are expressed in any particular cell type (Galang et al. 2004; Hollenhorst et al. 2004). As with other gene families, the role of potential redundancy versus predicted specificity of the ETS family has not been rigorously tested.

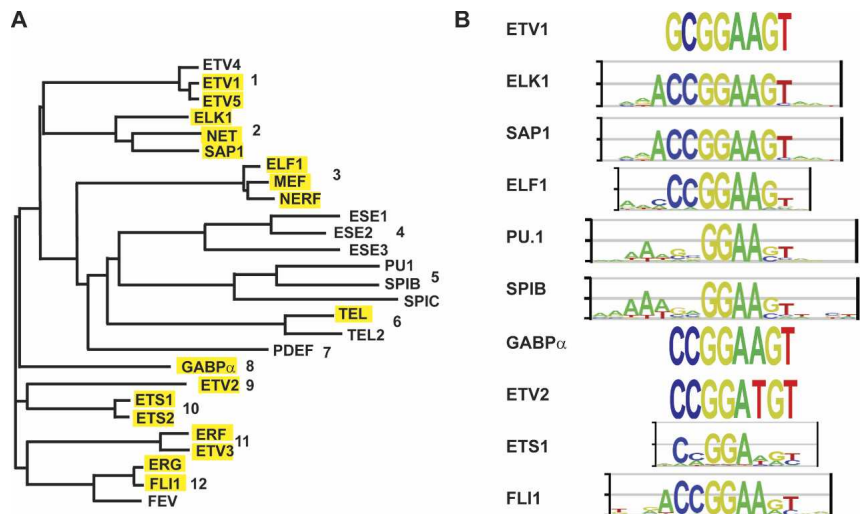
For individual ETS proteins, distinct functional domains that lie outside of the ETS domain could facilitate specificity. For example, one mechanism to enhance DNA-binding specificity is protein–protein interactions that mediate cooperative binding at distinct DNA se-

quences. The TCF clade (ELK1, SAP1, NET) functions with the DNA-binding factor, SRF, via a protein interaction domain (Price et al. 1995; Buchwalter et al. 2004). GABP α partners with GABP β , which mediates dimerization and formation of a GABP α / β hetero-tetramer that binds two ETS sites (de la Brousse et al. 1994). High-resolution molecular models of these complexes are available (Batchelor et al. 1998; Hassler and Richmond 2001; Mo et al. 2001); however, other partnerships are less well understood. For example, ETS1 could function with as many as nine different transcription factors (Li et al. 2000). Only RUNX1 (also known as AML1, CBF α 2, PEBP2) has been demonstrated to mediate DNA-binding cooperativity with ETS1 and, thus, potentially enhance specificity (Goetz et al. 2000; Gu et al. 2000). None of the potential ETS protein partnerships have been assayed by ChIP or shown to limit in vivo occupancy of other ETS proteins. Thus, the in vivo use of protein partnerships or any other specificity mechanism remains poorly characterized.

The unique biological function of ETS proteins predicts the selection of specific transcriptional targets. However, few target genes are linked definitively to individual ETS family members. Most of the >200 putative target genes for ETS proteins have been queried only by transcription effects that required overexpression in cell lines or by in vitro DNA binding, techniques that fail to identify the ETS protein(s) utilized in vivo (Semenchenko and Watson 2000). Furthermore, no genome-wide occupancy of an ETS protein has been reported.

Determining the genomic occupancy of ETS proteins by ChIP will provide an unprecedented view of in vivo DNA-binding specificity within a transcription factor family and allow us to test mechanisms regulating ETS protein targeting. By investigating the endogenous ETS proteins ETS1, ELF1, and GABP α in the Jurkat human

Figure 1. Conservation of mammalian ETS domains. (A) Phylogram tree of human ETS domain sequences. The amino acid sequence of all 27 human ETS domains were aligned by Clustal W (Thompson et al. 1994). The horizontal branch lengths relate to predicted evolutionary distance. (Longer branches are more divergent.) Nine clades with multiple highly similar domains and three additional singlet domains, indicated as numbers 1–12. *ETS* genes expressed in Jurkat T cells with mRNA levels above one copy per cell are highlighted in yellow (Hollenhorst et al. 2004). (B) ETS domain consensus binding sites. Illustrated sites were selected in vitro by SELEX method for the indicated mouse ETS proteins (Brown and McKnight 1992; Nye et al. 1992; Mao et al. 1994; Ray-Gallet et al. 1995; Shore and Sharrocks 1995; John et al. 1996). Sequence logos were created from PWM by enoLOGOS (Workman et al. 2005) with the height of each base related to reported frequency at that position. The selected ETV1, ETV2, and GABP α consensus sites were reported (Brown and McKnight 1992) without the necessary frequency distribution data to build a PWM.



T-cell line, we discovered that these divergent family members frequently occupied the same genomic regions. This redundant occupancy correlated with a match to a strong consensus DNA-binding site and proximity to the transcriptional start site (TSS). Specific binding of ETS1 was also detected, but did not correlate with a strong match to a consensus site. A subset of ETS1-binding events correlated with an ETS–RUNX composite site that differed dramatically from ETS1 or RUNX1 consensus sites. The finding of two classes of ETS1 targets suggests a versatility of the ETS family, overlapping and specific DNA-binding modes that are mediated through distinct sequence motifs.

Results

ETS proteins display both specific and redundant occupancy

Three genes reported to be regulated by ETS1 illustrate the need for more robust *in vivo* approaches and attention to family issues. The T-cell receptor (*TCR*) α and β enhancers have been characterized *in vitro* as sites of cooperative binding between ETS1 and RUNX1. However, *in vivo* specificity is not clear, as transient expression assays indicate that multiple ETS proteins can activate via this binding site (Sun et al. 1995), and no ChIP has been reported. In contrast, the promoter of the protein kinase encoding gene, *CDC2L2*, is implicated as an ETS1 target by ChIP, but no tests for specificity were performed (Feng et al. 2004). To test ETS protein *in vivo* specificity, we investigated occupancy of the *CDC2L2* promoter and *TCR* α and *TCR* β enhancers by four distantly related ETS proteins—ETS1, GABP α , ELF1, and ELK1—in Jurkat T cells (Fig. 1A). (Based on steady-state mRNA levels, these *ETS* genes rank first, second, ninth, and 11th, respectively, of 17 *ETS* genes that are expressed in Jurkat T cells [Hollenhorst et al. 2004]). ETS1, ELF1, and GABP α , but not ELK1, redundantly occupied the

CDC2L2 promoter, whereas ETS1 specifically occupied the *TCR* α and *TCR* β enhancers (Fig. 2A). RUNX1 also occupied the *TCR* α and *TCR* β enhancers, supporting a role for RUNX1 in ETS1 specificity. These initial ChIP experiments detected the anticipated specific mode of binding for ETS proteins, but also found a surprising redundant mode.

Redundant occupancy by ETS proteins is widespread

To ascertain the biological significance of redundant occupancy and the relative importance of RUNX1 in specificity, we performed genome-wide promoter ChIP. The relative levels of specific and redundant binding of ETS proteins were assessed by a promoter microarray hybridized with ChIP DNA from the Jurkat human T-cell line. The promoter microarray represented the region from 5000 base pairs (bp) upstream of to 2000 bp downstream from the TSS of ~17,000 human genes with 60-mer oligonucleotides at an average spacing of 200 bp. Promoters were scored as “bound” by statistical methods that considered the enrichment of multiple neighboring probes and consistent occupancy in experimental repetitions. Promoters occupied by ETS1, ELF1, or GABP α were frequently bound by one or more of the other ETS proteins (Fig. 2B). A second, independent set of ChIP–chip experiments, which was performed with a second promoter microarray that covered only regions within 1000 bp of the TSS, also indicated a very strong correlation between ETS1 and ELF1 occupancy (Supplementary Fig. S1).

This extensive overlap in potential targets was unexpected, and therefore we considered several possible non-biological explanations. The overlap was not due to cross-reactivity of antibodies because the epitopes had no sequence similarity, and immunoprecipitation controls (Supplementary Fig. S2) as well as ChIP experiments (Fig. 2A) showed specificity. We considered a possible bias toward these genomic regions in the microarray design. ChIP–chip of E2F4, a transcription factor that

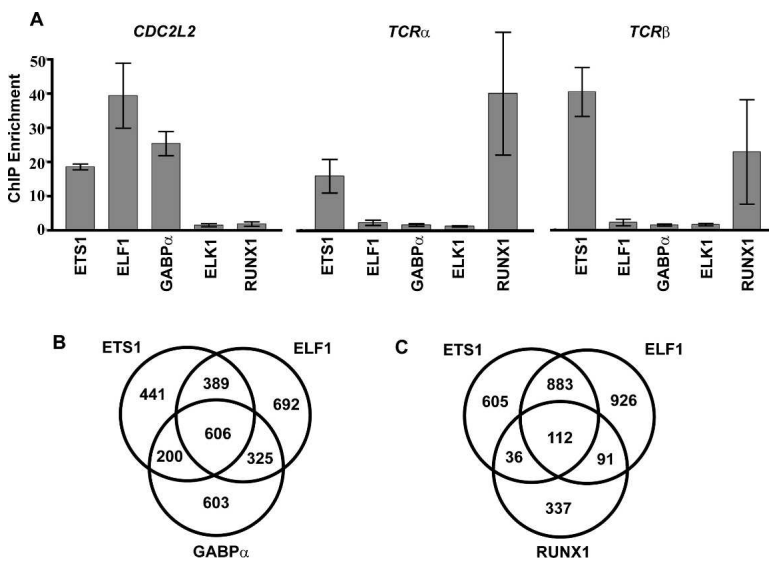


Figure 2. Specific and redundant promoter occupancy of ETS proteins. ChIP from the Jurkat human T-cell line with antibodies specific to the indicated ETS proteins. (A) Gene-specific region analysis. ChIP DNA was PCR-amplified with gene-specific primers. The ChIP enrichment is the ratio of the quantitative PCR signal of specific genomic regions over background genomic DNA (mean of two negative control genomic regions). Bars indicate the mean and standard error of the mean from three independent ChIP experiments. (B,C) Genome-wide occupancy analysis. ChIP DNA was amplified, labeled, and hybridized to a promoter microarray for 17,000 human genes representing sequences between -5 kb and $+2$ kb relative to the TSS. A bound promoter includes one or more peaks as defined by at least one probe with a $P(\bar{X})$ value of <0.001 . Data represent the average of two biologically independent replicates. Diagrams illustrate the number of promoter regions bound by ETS1, ELF1, GABP α , and RUNX1, and combinations thereof.

does not belong to the ETS family, served as a negative control; a set of targets distinct from those bound by ETS1, ELF1, and GABP α , but similar to E2F targets in other cell types, was identified (Table 1; Supplementary Fig. S1; Boyer et al. 2005). Another concern was the sensitivity necessary to detect specific sites. Quantitative PCR detected ETS1-specific occupancy at the *TCR α* and *TCR β* enhancers (Fig. 2A), but these sites were not near the TSS and, thus, were not on the promoter microarrays. To use this positive control we designed a third microarray that covered 20-kb regions surrounding these enhancers. ETS1-specific binding regions were detected and correlated with the known enhancers (data not shown). These controls indicated that the overlapping ChIP enrichments at ETS1, ELF1, and GABP α target promoters represent an accurate picture of genome-wide occupancy.

Redundant ETS binding correlates with a consensus ETS-binding site

We postulated that the redundant and specific classes of target genes may have different biological functions and that distinct mechanisms would dictate ETS protein re-

cruitment to each class. A more in-depth comparison of redundant and specific binding regions required a data set of each binding class that minimized false-positive results (albeit at the cost of increasing false-negative results). Therefore, data sets of segments bound by ETS1 and ELF1, ETS1 but not ELF1, or ELF1 but not ETS1 were created (Fig. 3A). ChIP and quantitative PCR with primers specific for each candidate segment showed 88% or greater concurrence with ChIP–chip data, thus validating ETS1 and ELF1 dual-bound as well as ETS1-specific data sets (Fig. 3B,C). Tests for occupancy by the ETS protein ELK1 yielded negative results. The ELF1-specific data set was less reliable (Fig. 3D). Thus, the ETS1 and ELF1 dual-bound and ETS1-specific data sets were used for further analyses.

We hypothesized that sequence elements would direct redundant versus specific binding. To search for such sequences we used the MEME algorithm, which identifies significantly overrepresented DNA sequences. The data output was position weight matrices (PWMs), which give frequency distributions of each base at each position (Bailey and Elkan 1994). The PWM for the most overrepresented sequence in the dual-bound data set (consensus: CCGGAAGT) (Fig. 4A) was strikingly similar to the derived in vitro-selected consensus sites for ETS1, ELF1, and GABP α (Fig. 1B). Greater than 70% of dual-bound segments had a sequence represented within this PWM. MEME did not identify any significantly overrepresented sequences in either the ETS1-specific data set or in any of 10 data sets randomly selected from the list of interrogated promoter regions (data not shown). To ensure that the distinction between dual and specific data sets was not due to the size of the data sets, randomly selected subsets of the dual-bound data set, similar in size to the specific data set, were tested; these smaller subsets also returned an ETS-like consensus sequence (data not shown). In conclusion, redundant binding by ETS proteins correlated with the presence of a strong consensus ETS-binding site.

To further investigate the presence of ETS consensus sites in dually occupied promoters, we searched in a biased manner for ETS-binding motifs. The PWM identified by MEME (Fig. 4A) was used as a query sequence for the pattern-recognition program PATSER (Hertz and Stormo 1999). The search was performed on the 3000 bp surrounding the TSS of genes closest to ETS1 specifically bound segments, as well as those bound dually by ETS1 and ELF1. A set of randomly selected genes also was searched. The matches near dual-bound genes scored significantly higher (mean PATSER score 9.5 for dual bound vs. 8.7 for both specific and random) than those near the ETS1-specific (*t*-test; $P < 0.0001$) or randomly selected genes ($P < 0.0001$). Scores for random genes and specific genes showed no significant difference ($P = 0.68$). In conclusion, two independent bioinformatics approaches indicated that a consensus ETS-binding site correlates with redundant ETS protein occupancy.

The finding of a nondiscriminating consensus sequence led us to investigate other sequence features that might accompany redundant occupancy. To test

Table 1. Overrepresented ontologies of genes near bound promoters

Data set	Ontology	<i>P</i> value ^a
ETS1/ELF1/GABP α in Jurkat (400) ^b	RNA processing	2.50E-08
	Ribosome	3.67E-05
	Cellular metabolism	8.24E-05
	Nucleic acid binding	1.61E-04
ETS1 specific in Jurkat (442)	Response to DNA damage stimulus	3.36E-07
	Protein biosynthesis	5.90E-06
	Cellular metabolism	5.90E-06
	Mitochondrion	1.98E-05
	Nucleic acid binding	2.67E-05
ETS1 in HT29 (400) ^b	Cellular metabolism	5.34E-12
	Ribonucleoprotein complex	4.97E-09
	Nucleic acid binding	2.10E-06
	Zinc ion binding	2.87E-05
E2F4 in Jurkat (400) ^b	Chromosome	4.15E-42
	Cell Cycle	2.79E-29
	Cell Division	9.08E-17
	Nucleotide metabolism	3.42E-15
	Replication	2.07E-14
	Response to DNA damage stimulus	2.79E-14
	DNA binding	6.99E-08
	Nucleotide binding	1.72E-05
	None	<0.001
Ten random gene sets (400 each) ^c	None	<0.001

^a*P* value from Gostat. Ontologies with *P* values >0.001 were not listed.

^bLists were rank ordered by $P(\bar{X})$ values or mean $P(\bar{X})$ values within each category, and the top 400 were selected to allow analysis of a similar number of genes (indicated by parenthesis) from each category.

^cGene sets were picked randomly from all genes with promoter regions represented on the microarrays.

Hollenhorst et al.

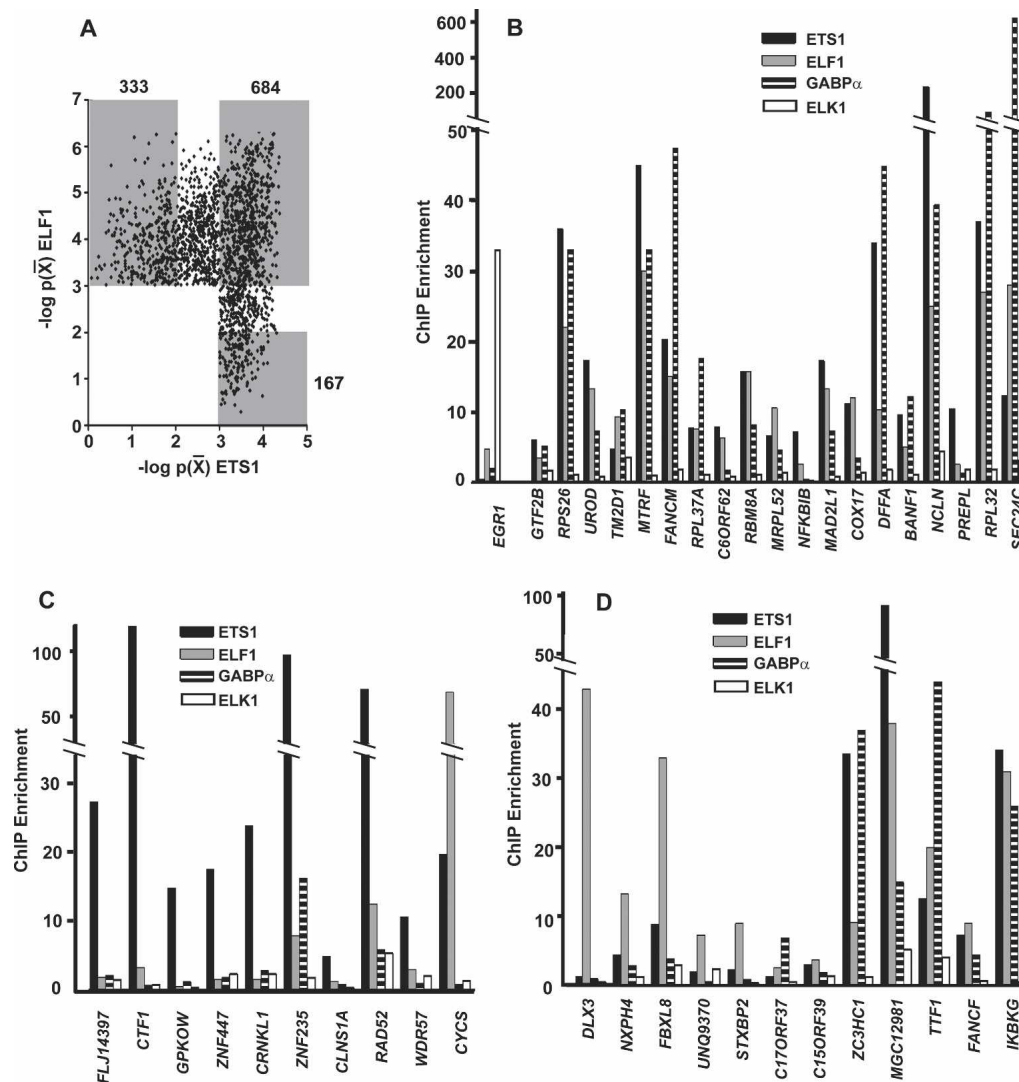


Figure 3. Validation of redundant and specific bound segments from genomic occupancy data sets. (A) Stringent classification of redundant versus specific data sets. Bound segments were identified regardless of promoter annotation. To identify specific binding events, the stringencies of an “unbound” score was reduced [minimum $P(\bar{X})$ value >0.01] to minimize false negatives. The probe with the lowest $P(\bar{X})$ value from each bound segment was taken to represent that segment for that ETS protein. To compare with occupancy, the value of this representative probe was plotted versus the highest $-\log P(\bar{X})$ value for the other ETS protein within 1 kb of this probe. Shaded areas define the three data sets used for the comparisons in the remainder of the figure (size indicated). ETS1-specific and dual-bound segments were identified by the ETS1 segment report. (Dual-bound segments identified by the ELF1 segment report were essentially the same.) ELF1-specific segments were identified by the ELF1 segment report. (B–D) Quantitative ChIP confirmed specific and redundant occupancy. ChIP DNA obtained with specific antibodies, as indicated, was analyzed by quantitative PCR, as in Figure 2A. The scale on the Y-axis was interrupted to show a broad range of values. Segments were assigned to a gene by the closest TSS. In B, the *EGR1* promoter, a well-characterized ELK1 target, served as an antibody positive control. The other 19 tested segments were randomly selected from the “dual-bound” data set from A. In C, 10 segments were randomly selected from the ETS1-specific data set from A. In D, 12 segments were selected from the ELF1-specific data set with preference for those with the least evidence of ETS1 occupancy [highest ETS1 $P(\bar{X})$ values].

whether there was a bias in the location of dual versus specific bound segments, the distance of each segment (measured from the highest-scoring oligonucleotide probe) to the TSS was determined for the ETS1 and ELF1 dual-bound and ETS1-specific data sets. Dual-bound segments clustered very strongly to a region within 200 bp of the TSS (Fig. 4B). (More detailed spacing conclusions are challenged by the limits of ChIP–chip resolution and

TSS annotation.) Segments bound specifically by ETS1 showed significantly less constraint on their location and frequently appeared more distally (Fig. 4B) (*t*-test of the mean distance; $P = 0.0002$). In an independent approach to detect potential location bias, we measured the distance between the TSS and the best PWM matches from the PASTER analysis. A subset of the randomly selected genes had their strongest PWM matches in

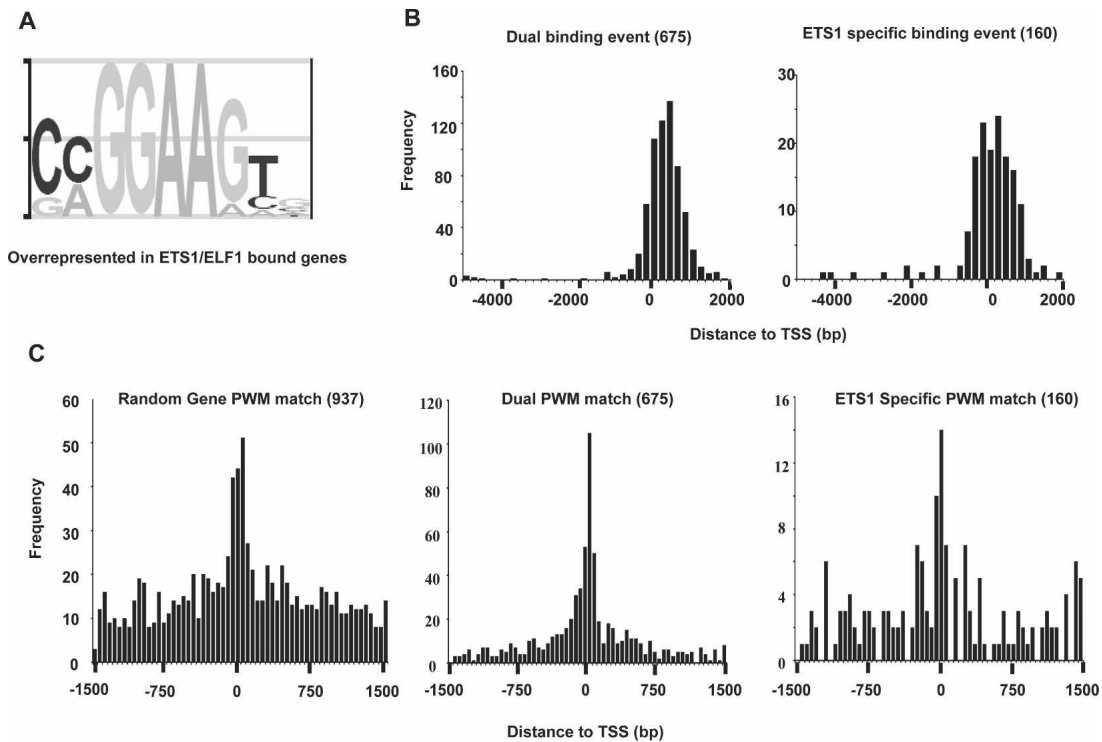


Figure 4. Sequence characteristics of redundantly and specifically bound segments. (A) ETS1 and ELF1 dual-bound segments correlate with a strong match to a consensus ETS-binding site. The most overrepresented complex PWM identified by MEME ($E = 1 \times 10^{-30}$) in dual-bound segments is illustrated by a sequence logo, as in Figure 1B. (B) Distances between TSS and bound segments (from Fig. 3) showed proximal bias of dual occupancy. The distance from the bound segment to the nearest TSS in the Ensembl database was measured by using the location of the oligonucleotide probe with the lowest $P(\bar{X})$ in each segment in the dual-bound or ETS1-specific data sets. Distances were binned and frequencies were plotted. The number of segments analyzed is indicated in parenthesis. (No nearby gene was identified for nine dual-bound segments and seven ETS1-specific segments). The mean distances from start were significantly different (362 bp for dual-bound; 532 bp for specific segments; t -test, $P = 0.0002$). (C) Distances between TSS and the best match to ETS-binding consensus showed proximal bias of dual occupancy. PATSER was used with the PWM (shown in A) to search for the best consensus match within 1500 bp of sequence upstream of and downstream from the TSS in the dual-bound or ETS1-specific data sets (from Fig. 3) as well as a set of 937 randomly selected genes. The position of the highest-scoring match to the PWM in each 3000-bp region surrounding the TSS was binned and plotted. The mean distance from the TSS to the best PWM match was significantly shorter for dual regions (386 bp) versus ETS1-specific (608 bp; t -test, $P < 0.0001$) or randomly selected regions (604 bp; t -test, $P < 0.0001$).

regions proximal to the TSS (Fig. 4C). However, the dual-bound genes were significantly enriched for this type of promoter, as indicated by significant difference between the mean distances (t -test; $P < 0.0001$). Notably, the mean distance of the ETS1-specific genes were not significantly different from that of the random genes (Fig. 4C). Only five ETS1-specific promoters (4%) had a perfect match to the PWM consensus within 200 bp of the TSS compared with 87 (14%) of the dual-bound promoters. In conclusion, two sequence properties correlated with redundant occupancy by ETS transcription factors—the presence of a consensus ETS-binding site and the tendency of this site to be located proximal to the TSS.

Housekeeping genes have redundantly occupied promoters

To investigate the biological role of the nonselective ETS binding at strong proximal ETS-binding sites, we asked

whether the dual-bound gene set represented a specific biological pathway. Overrepresented ontologies of the genes near the ETS1 and ELF1 dual-bound promoters were queried by Gostat (Beissbarth and Speed 2004). Housekeeping categories (e.g., RNA processing, ribosomal proteins, and cellular metabolism) had significant enrichment scores (Table 1). Random gene lists of similar size did not return any significant overrepresented categories. Additional informatics analyses supported this correlation between dual occupancy and housekeeping function. Eighty-five percent of the ETS1 and ELF1 dual-bound regions overlapped with CpG islands, a sequence feature consistent with the promoters of housekeeping genes (Bird 1986). In contrast, data sets built from promoter regions with matched GC content that were randomly selected from the extended promoter array regions only displayed an average of 44% overlap with CpG islands ($P < 0.01$). Next, we queried our data set against a human gene set annotated for sequence characteristics of housekeeping genes (De Ferrari and

Aitken 2006). The promoters of 16% of all genes surveyed had evidence of ETS1 and ELF1 dual occupancy (see Materials and Methods), whereas this proportion was 52% and 4% among genes classified as “housekeeping” and “nonhousekeeping,” respectively. Thus, three methods of classifying housekeeping genes each indicated an enrichment of redundant ETS occupancy at housekeeping promoters.

Co-occupancy of housekeeping promoters indicated a possible redundant function of ETS proteins at these promoters. We predicted that co-occupancy would not be cell-type specific, although different ETS protein combinations may be present in different cell types. To test this hypothesis, HT29 colon adenocarcinoma cell lines were used for ETS1 ChIP–chip. Promoters occupied by ETS1 were again overrepresented for housekeeping categories (Table 1). Therefore, ETS transcription factors appear to have a redundant role at the promoters of housekeeping genes, possibly in multiple cell types.

ETS1 and RUNX1 occupy promoters with a composite ETS1–RUNX-binding site

Specific occupancy of ETS1 in Jurkat T cells could be mediated through cooperative interactions only with RUNX1 or through a variety of cooperative partners. To differentiate between these two possibilities, a ChIP–chip experiment was performed with an antibody specific for RUNX1; 576 RUNX1-bound promoters were identified (Fig. 2C). However, only 36 of the 641 promot-

ers bound by ETS1, but not ELF1, were also occupied by RUNX1. Therefore, cooperative interactions with RUNX1 likely represent one of a number of mechanisms that can mediate ETS1 specificity in Jurkat T cells.

Although RUNX1 occupancy could not explain the majority of the ETS1-specific binding, eight of the promoters with the strongest ETS1-binding signals also had strong RUNX1 signals (Fig. 5A, circled). In ChIP coupled with quantitative PCR, some of these segments showed weak binding signals for ELF1 and GABP α , but in every case the ETS1 antibody gave a strikingly higher signal (cf. Figs. 3B and 5B), indicating that these segments strongly favored ETS1 binding.

Because RUNX1 and ETS1 co-occupancy at these eight segments correlated with a strong ETS1-binding signal in ChIP–chip, but not a strong ETS1 consensus binding site (mean PATSER score 8.3), we speculated that novel sequence determinants could be present in these regions. A MEME analysis provided a PWM for the most overrepresented sequences in the 87 DNA segments bound by both ETS1 and RUNX1 (consensus: CTGGGAATTG TAGTT) (Fig. 5C). Sequences in 40 segments were represented by the PWM, including all eight surveyed by quantitative PCR (Fig. 5B). This site was well conserved across multiple mammalian genomes in the majority of identified segments, suggesting functional importance (Supplementary Fig. S3). This PWM had some similarity to ETS1- and RUNX1-binding sites. However, the consensus deviated substantially from the *in vitro* selected consensus for ETS1 and RUNX1 (Fig. 5C; Nye et al.

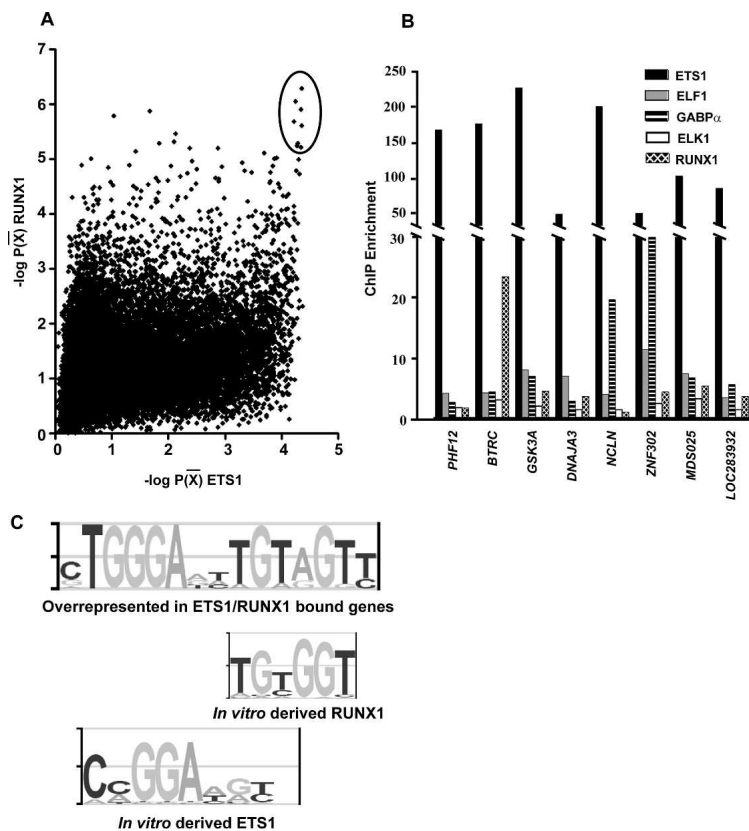


Figure 5. RUNX1 and ETS1 co-occupancy correlates with a sequence similar to ETS- and RUNX1-binding sites. (A) Dual occupancy criteria. The maximum $-\log P(\bar{X})$ value of ETS1 and RUNX1 for each promoter region represented on the array is plotted. Eight promoter regions had a $-\log P(\bar{X})$ value of >5 for RUNX1 and >4 for ETS1 (circled). (B) Quantitative ChIP validated ETS1 and RUNX1 co-occupancy. ChIP DNA was analyzed by quantitative PCR and gene-specific primers, as in Figure 3B, for each of these eight promoter regions. (C) A ETS1/RUNX1 composite site. (Top) The most overrepresented sequence identified by MEME analysis of 87 ETS1- and RUNX1-bound segments ($E = 4 \times 10^{-33}$). (Bottom) The *in vitro*-derived binding sites for RUNX1 (Meyers et al. 1993) and ETS1 (Nye et al. 1992).

1992; Meyers et al. 1993) and from the ETS1–RUNX1 cooperative binding sites in the *TCR α* and *TCR β* enhancers (Gottschalk and Leiden 1990). All 40 PWM matches displayed the canonical GGA. However, only 26 had the conserved (A/T) wobble 3' of the GGA, and 13 had an extremely divergent GGAG sequence. In addition, the other flanking nucleotides were dissimilar to those in the in vitro-derived ETS1 PWM, thus predicting extremely low affinity.

Based on the poor fit of this ETS1–RUNX1 composite site to consensus sites, we hypothesized that in vivo occupancy might require cooperative DNA binding between ETS1 and RUNX1. DNA-binding assays were performed to determine the relative affinity of ETS1 in the presence and absence of RUNX1 (Fig. 6). ETS1 bound weakly alone with an affinity 10- to 100-fold lower than that of a consensus site (Goetz et al. 2000). The affinity increased more than fivefold in the presence of RUNX1. Detection of ETS1 binding to GGAG sites required RUNX1. In conclusion, the ETS1–RUNX1 composite sites were marked by poor matches to ETS consensus sites and displayed low affinity that was improved by cooperative DNA binding. Remarkably, these extremely low-affinity sites displayed strong ChIP–chip signals (Fig. 5A) and extremely strong quantitative ChIP signals with gene-specific primers (Fig. 5B). We speculate that cooperative interactions can mediate extremely stable in vivo occupancy comparable to that of consensus ETS-binding sites.

The genes whose promoters were bound by ETS1 and RUNX1 did not fall into housekeeping categories. One alternative is that these genes are more specialized, representing differentiation-specific or tissue-specific targets. To consider candidates for ETS1 targets we note that an *ETS1* gene disruption in the mouse causes defects in T and B cells (Bories et al. 1995; Muthusamy et al. 1995). Furthermore, ETS1 and RUNX1 are most abundant in hematopoietic cell types. Therefore, we queried the literature for possible hematopoietic functions of the 40 genes that had the ETS1–RUNX1 composite binding sites. Indeed, four genes had known roles in hematopoietic cells. One example was the transcription factor LEF1, a hematopoietic-specific transcription factor that regulates the *TCR α* enhancer (Travis et al. 1991). In con-

clusion, in contrast to the redundant role of ETS proteins at housekeeping genes, we predict tissue-specific gene expression will require weaker ETS sites—and thus, cooperative partnerships—to use specific ETS proteins.

Discussion

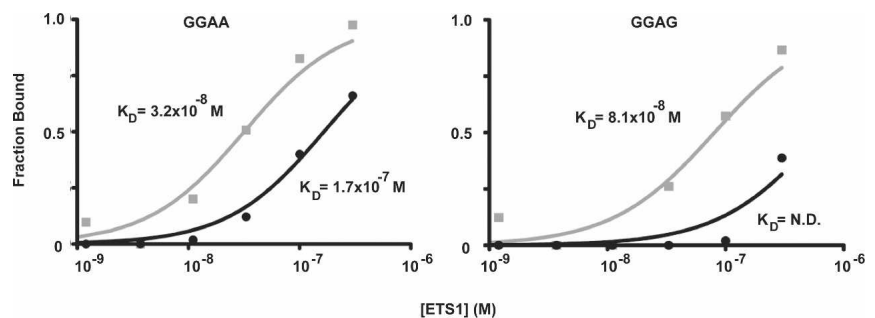
This genome-wide promoter occupancy study implicated the *ETS* family in a redundant role at the proximal promoters of housekeeping genes. We also detected specific binding, as defined by occupancy of ETS1, but not three other family members. Sequence motifs and spatial biases that correlated with the two modes of binding were identified by bioinformatics analyses. Most dramatically, unbiased searches for consensus sites correlated strong binding sites with redundant occupancy and weak binding sites with specific occupancy.

A redundant role for ETS proteins

Our survey of the in vivo occupancy of four ETS proteins from four different clades found that three of these proteins often occupy the same promoter regions. There are two possible models to explain the detection of multiple ETS proteins. This co-occupancy could represent a separate binding site for each factor or alternate occupancy of the same site in different cells, on different alleles, or at different times. Because the regions occupied redundantly by ETS family members correlated with a strong match to the consensus sites of multiple ETS proteins, we propose that the same binding site is bound alternatively by different ETS transcription factors.

Bioinformatics studies have identified hundreds of sequence motifs that are overrepresented in human promoters and ETS-like binding sites are always present on these lists (Bina et al. 2004; FitzGerald et al. 2004; Xie et al. 2005). Our results indicate that some of these sequence motifs are likely to be occupied in vivo by multiple ETS proteins. The ETS protein ELK1 did not co-occupy promoters with ETS1, ELF1, and GABP α . This is consistent with in vitro data suggesting that ELK1 has low affinity for a monomeric ETS site, but requires SRF

Figure 6. ETS1 and RUNX1 bind DNA cooperatively to the MEME-derived composite site. Equilibrium binding curves for ETS1 and a 35-bp region of the *MDS025* promoter (Fig. 5B) that displays a GGAA core (left) or with a single nucleotide change in the core to GGAG (right), which is also represented in the MEME-derived PWM. Experiments performed in the presence (gray squares) or absence (black circles) of 30 nM RUNX1 (1–302 fragment) and indicated ETS1. The K_D was derived by curve fitting by nonlinear least-squares analysis with fraction of DNA bound = $1/(1 + K_D/[ETS1])$. ETS1-only binding to the GGAG site was below the level necessary for quantification by this assay.



Hollenhorst et al.

and an adjoining SRF site for high-affinity binding (Price et al. 1995). Additional ChIP data will be required to uncover how many of the 23 remaining ETS transcription factors participate in redundant occupancy of strong ETS-binding sites.

The observation that 5%–15% of the 17,000 promoters are occupied by multiple ETS transcription factors suggests biological importance. In considering the potential significance we noted that many of the redundantly bound regions were associated with housekeeping genes. These findings are consistent with a bioinformatics report of a collection of ETS-type sequences as one of three sequence motifs found in proximal promoters of housekeeping genes (FitzGerald et al. 2004). Our discovery of redundant binding of ETS transcription factors at these genes suggests that this mode of binding could facilitate consistent regulation of ubiquitously expressed housekeeping genes. In this model, sustained expression would be independent of the varying levels or identity of individual ETS proteins in distinct cell types.

More intriguingly, we speculate that the redundantly occupied regions identified in our study are targets for oncogenic ETS proteins. Preliminary support for this hypothesis comes from recent observations in human prostate cancer. It is now proposed that more than half of all cases correlate with a chromosomal rearrangement that leads to overexpression of one of three ETS proteins (Tomlins et al. 2005, 2006). A gene set expressed at higher levels in prostatic intraepithelial neoplasia (PIN) than in normal prostate tissue have two features that are similar to our dually occupied data set. An ETS-binding site is the most enriched site from the TRANSFAC database in the promoters of PIN-specific genes. By ontology analysis, genes involved in protein biosynthesis, including those encoding ribosomal proteins, are up-regulated in PINs (Tomlins et al. 2007). We speculate that some housekeeping genes, specifically those redundantly regulated by ETS proteins, could be a class of misregulated targets relevant to tumor progression.

Specific binding and protein partnerships

Specific ETS1 occupancy was uncovered by the ChIP-chip experiment, although this class of targets was less frequent than the dual-bound targets. Furthermore, unbiased searching for enriched sequences did not find a PWM closely related to the *in vitro*-derived PWM for ETS1 or the PWM derived from dual-bound targets. We propose that a high-affinity ETS1-binding site would preclude specific occupancy due to its concurrent high affinity for other ETS proteins. We speculated that ETS1 achieves sufficient affinity for specific sites only by DNA-binding cooperativity with additional transcription factor(s). Indeed, intersecting ETS1 and RUNX1 ChIP-chip data sets facilitated an informed, yet unbiased, search that discovered a composite site resembling both ETS- and RUNX1-binding sites, but not matching either consensus. As additional ChIP-chip data become available, a similar strategy may identify sequences important for the remaining ETS1-specific sites and for

specificity of other ETS proteins. ELK1-specific occupancy of the EGR1 promoter provides an exception to the trend of weak ETS sites for specific binding. This promoter has a strong ETS consensus juxtaposed to an SRF-binding site. This exception may have evolved because constitutive ELK1/SRF occupancy can occlude occupancy by other ETS proteins. Whether ELK1-specific binding generally occurs at weak or strong ETS sites awaits genome-wide occupancy data. In summary, the characterization of an ETS1/RUNX1 cooperative partnership by genome-wide occupancy data provides global evidence for combinatorial control of gene expression within the ETS family.

Limitations of undirected bioinformatics approaches

Mapping genome-wide transcription factor binding is envisioned to enable prediction of gene regulatory pathways. Bioinformatics approaches have attempted to use factor-binding predictions that are based on experimentally derived PWMs that favor the strongest binding sites *in vitro* (Aerts et al. 2003; Blanchette et al. 2006). Our observation that strong binding sites do not correlate with specificity challenges the accuracy of these *in vitro*-based approaches, especially for gene families. Other ChIP-chip assays have revealed that many transcription factors bind to regions of the genome that contain no strong matches to the consensus. Our data show how this phenomenon is related to specificity requirements for that particular transcription factor. Thus, bioinformatics approaches that can successfully identify transcription factor-binding sites *in silico* will likely require the integration of rules for cooperative DNA-binding partnerships.

General rules for gene families

An interesting question is whether these features of the ETS family are predictive of other transcription factor families. There are >20 identified families whose members have conserved DNA-binding domains and common binding properties (Messina et al. 2004). The only mammalian transcription factor family with more than one member assayed by ChIP-chip is the six-member E2F family (Ren et al. 2002; Weinmann et al. 2002; Oberley et al. 2003; Wells et al. 2003). The microarrays available at the time of these studies only assayed a small subset of promoters, and none of these studies identified differences between redundantly and specifically occupied regions. Thus, our study expanded the use of genome-wide occupancy techniques to survey a larger mammalian transcription factor family and revealed characteristics of specifically and redundantly occupied loci. We expect our discovery of extremely weak sites, as a feature of the ETS1–RUNX1 interaction, will be generally applicable to many other DNA-binding partnerships. On the other hand, the use of the ETS family in a redundant manner at proximal promoters may represent a biologically important feature unique to this family.

In conclusion, an *in vivo* genomic occupancy approach demonstrated both redundant and specific roles for ETS transcription factors. Interchangeable occupancy of diverse members of the family at proximal promoters suggests an unexpected overlapping function for proteins that share no similarity outside of their DNA-binding domains. Furthermore, this occupancy uncovers a strategy that could mediate stable expression of housekeeping genes by making them relatively resistant to changes in transcription factor concentration and regulatory modifications. In contrast, weaker binding sites provide opportunities to enhance affinity and add specificity for biological regulation. In conclusion, the distinct promoter occupancy patterns of ETS proteins demonstrate the versatile use of a transcription factor family.

Materials and methods

Genomic tiling microarray designs

Two promoter microarrays were used in our studies. The promoter microarray (Agilent Technologies, G4481A), which was used only for data in Supplementary Figure S1, consisted of two slides with a combined 88,000 60-mer oligonucleotide probes representing sequences from -1 kb to $+0.3$ kb relative to the TSS of $\sim 17,000$ best-defined human transcripts from University of California at Santa Cruz hg17/NCBI release 35 (May 2004). An average promoter region was represented by four to five probes. A second promoter microarray (Agilent Technologies, G4489A) consisted of two slides with a combined 488,000 60-mer oligonucleotides representing sequences from -5 kb to $+2$ kb relative to the same TSS as in the proximal microarray. A custom microarray was manufactured by Agilent Technologies to represent the region from 10 kb upstream of to 10 kb downstream from the previously identified *TCR α* and *TCR β* enhancers using probes from the Agilent genomic tiling database. The average spacing between probes within these regions was ~ 100 bp.

Cell culture

Cell lines were grown using standard tissue culture techniques. Jurkat cells were maintained in RPMI medium (GIBCO) plus 10% fetal bovine serum (FBS), 2 mM L-Glutamine, 10 mM Na Pyruvate, and 10 mM Hepes. HT29 cells were maintained in McCoy's 5A medium (GIBCO) plus 10% FBS and 2 mM L-Glutamine.

ChIP

Dynabeads (50 μ L) conjugated to sheep anti-rabbit IgG (DynaL Biotechnology) were mixed with 1 mL of Dilution Buffer (20 mM Tris at pH 7.9, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 4 mg/mL bovine serum albumin [BSA], mammalian protease inhibitors [Sigma, #P8340]) and rotated for 10 min at 4°C. Next, 5 μ L of polyclonal rabbit antibody (ETS1, sc-350; ELF1, sc-631; GABP α , sc-22810; ELK1, sc-355; and E2F4, sc-1082; Santa Cruz Biotechnology) or monoclonal mouse antibody (RUNX1, α 3.2.3.1; gift of Dr. Nancy Speck [Dartmouth Medical School, Hanover, NH]) was added and the slurry was rotated overnight at 4°C. Cross-linked and sheared chromatin extracts were prepared as described previously (Hollenhorst et al. 2004). The extract (100 μ L) was added to Dynabead/antibody slurry and rotated for 4 h at 4°C. Beads were washed four times for 5 min with immunoprecipitation wash buffer (20 mM Tris at pH

7.9, 2 mM EDTA, 250 mM NaCl, 0.25% NP-40, 0.05% SDS). Beads were resuspended in 100 μ L of 10 mM Tris (pH 8.0) and 100 μ g/mL RNase A and incubated for 30 min at 37°C. SDS concentration was brought to 1% and Proteinase K was added to 200 μ g/mL. Bead slurries were then incubated for 3 h at 55°C and 6 h at 65°C. ChIP DNA was purified from slurries by phenol/chloroform extraction and QiaQuick PCR Purification Kit (Qiagen).

ChIP DNA amplification, labeling, hybridization, and scanning

ChIP DNA was amplified by a whole-genome amplification kit (WGA2, Sigma) according to the manufacturer's instructions, except that the fragmentation step was skipped and the number of PCR cycles was increased to 20. Alternate random primed and linker-mediated PCR amplification protocols were compared with the WGA2 kit on the proximal promoter microarray and gave similar results for ETS1 occupancy (data not shown). Amplified DNA was treated with a QiaQuick PCR purification kit (Qiagen), then labeled, hybridized to the Agilent microarrays, and washed as previously described (Boyer et al. 2005). Hybridized microarrays were scanned using an Agilent G2565BA microarray scanner and raw image files were processed with Agilent Feature Extraction software (version 8.5). Two replicates of each ChIP-chip experiment from independent cell cultures were performed, with the exception of three repetitions of Jurkat ETS1 on the proximal promoter microarray and one repetition of HT29 ETS1 on the proximal promoter microarray. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE7449.

Classification of binding events

Data were analyzed with ChIP Analytics (version 1.3, Agilent Technologies) with the Whitehead Error Model. Normalization consisted of subtraction of median signals from negative control features, interarray median normalization, and dye-bias normalization. A weighted average was used for replicates. A value, X , was calculated for each probe; this value correlates with a log ratio, but includes a correction for low intensities (for equations, see ChIP Analytics 1.3 user's guide). A P value for each probe, $P(X)$, represented the probability of observing an X value as high as or higher than its own, given the normal distribution of the mean and standard deviation. To incorporate the data of neighboring probes (peaks), a value \bar{X} was calculated as the mean of the X value from that probe and the neighboring probe on either side. If the neighboring probe is >1 kb away, a value of 0 was substituted for X . A P value for \bar{X} , $P(\bar{X})$, was calculated just as for $P(X)$.

For the data shown in Figures 2 and 5A and Table 1, a "bound" promoter designation required that one or more probes within a promoter region have a $P(\bar{X})$ of <0.001 as determined by the "gene report" output of ChIP Analytics. The $P(\bar{X})$ values used in Table 1 and Figure 5A were the lowest $P(\bar{X})$ for any probe within that promoter region. Genes in the De Ferrari data set were considered to have evidence of dual occupancy if the corresponding promoters had minimum $P(\bar{X})$ values of <0.01 for both ETS1 and ELF1.

The segments shown in Figure 3A were derived from the "segment report" of the ChIP Analytics, which classifies a genomic region as a "bound" segment if there is a series of "bound" probes with <1 kb gaps. A "bound" probe must satisfy a significance heuristic of $P(\bar{X}) < 0.001$ and either $P(X) < 0.001$ and one

Hollenhorst et al.

neighboring probe with $P(X) < 0.01$,] or $[P(X) < 0.005$ for that probe and a neighbor, or $P(X) < 0.005$ for both neighbors]. For each bound segment, the probe in the segment with the lowest $P(\bar{X})$ was considered the center of that segment and that $P(\bar{X})$ value was used as the $P(\bar{X})$ value for the segment.

Dual-bound ETS1 and ELF1 or ETS1 and RUNX1, or specifically occupied ETS1 but not ELF1 were segregated by examining the $P(\bar{X})$ value of all probes from the ELF1 or RUNX1 probe report that lie within 1 kb from the center of the ETS1 segment (ChIP Analytics). Segments were considered dual bound if one of these probes had a $P(\bar{X})$ value < 0.001 and specific if none of these probes had a $P(\bar{X})$ value < 0.01 . Dual-bound and specific segments were matched to a specific gene by identifying the nearest TSS from the Ensembl database.

Analysis of bound segments by MEME, PATSER, and GStat

Bound segments were shortened by utilizing only the region spanning the central probe [lowest $P(\bar{X})$ value] and ending at probes that displayed a 100-fold increase in $P(\bar{X})$ value. If the edge of a segment (gap of at least 1000 bp) was encountered first, the segment was extended by 200 bp and ended. Segments were analyzed by MEME (<http://meme.sdsc.edu/meme/meme.html>) (Bailey and Elkan 1994) to identify sequences of variable length that occur more often than expected. Such sequences are reported as PWMs and are given an E value (expect value) describing the number of times that PWM would be identified by chance in a set of sequences of that size. MEME was run with default settings, except that the maximum motif length was set at 15 nucleotides. All sequences shown had the lowest E value of all complex sequences returned. (Runs of a single nucleotide were not considered complex.)

Three kilobases of sequence surrounding the TSS of each gene matched to ETS1/ELF1 dual-bound or ETS1 specifically bound segments and 937 randomly selected genes (Ensembl) were analyzed by PATSER (<http://rsat.ulb.ac.be/rsat>) (Hertz and Stormo 1999). The PATSER program moved a window equal to the length of the ETS PWM (Fig. 4A) along both strand sequences and assigned a score to each position. The position of the highest-scoring PWM match for each sequence relative to the TSS was recorded.

The genes with the 400 lowest $P(\bar{X})$ values for each category (except for ETS1-specific genes, where all 437 genes were used) were analyzed for overrepresented ontologies using GStat (<http://gostat.wehi.edu.au>) (Beissbarth and Speed 2004). [For ETS1/ELF1/GABP α co-occupied genes, each $P(\bar{X})$ value was < 0.001 and the mean $P(\bar{X})$ value was used]. A list of the ~17,000 genes represented on the microarray was used as a background gene list. Random gene lists (400 each) were generated from this background gene list. The maximal P value for returned categories was set to 0.001. Redundant gene categories (differing by less than three genes) were collapsed to one category and uninformative gene categories were not recorded.

Real-time PCR

Real-time PCR was performed as described previously (Hollenhorst et al. 2004). Serial dilutions of Jurkat ChIP input DNA were used as a standard curve for real-time PCR. Primers designed for the specified genomic regions were found to amplify a single product from genomic input DNA based on a single melting peak. Each ChIP DNA sample was assayed for the levels of two negative control regions, the 3' ends of the *albumin*, and *BCL-XL* genes. In all cases, the absolute levels of these control regions varied by less than twofold. The mean level of the control regions was considered the background level of ge-

nomeric DNA. ChIP enrichments are reported as a ratio of the absolute measurement of each genomic locus to the background level of genomic DNA in the same sample. Primers used to assay genomic loci are listed in Supplementary Table S3.

Protein expression and purification

Human ETS1 (p51) was cloned into bacterial expression vector pET28A (Novagen) at a site that introduces a 6 \times HIS tag at the N terminus. ETS1 protein was expressed and purified as described previously (Jonsen et al. 1996). Protein concentration was determined by comparison to BSA standards by Coomassie brilliant blue-stained SDS-PAGE gels, and activity was determined by binding to the high-affinity ETS1-binding duplex 5'-TCGACGGCCAAGCCGGAAGTGAGTGCC-3' (Nye et al. 1992).

A fragment of the RUNX1 protein that includes amino acids 1–302 was a gift of Nancy Speck. This fragment retains all of the regions necessary for cooperative DNA binding with ETS1 and was purified from baculovirus-infected SF9 cells (Gu et al. 2000).

DNA-binding assays

Quantitative electrophoretic mobility shift assays were performed as described previously (Jonsen et al. 1996). In brief, indicated concentrations of ETS1 protein were mixed with 32 P-labeled double-stranded oligonucleotides at 1×10^{-11} M, then incubated for 1 h on ice. Duplexes, designated either GGAA or GGAG were composed of the following sequences, respectively: CACAGGAATGCTGGGAATTGTAGTTTTTCGCTCTGT; CACAGGAATGCTGGGAGTTGTAGTTTTTCGCTCTGT. Reaction mixtures with RUNX1 (amino acids 1–302) at 3×10^{-8} M were incubated on ice for 30 min before addition of ETS1 protein. Aliquots of binding mixtures were run on a 6% polyacrylamide gel and relative radioactivity in bound or unbound DNA bands was quantified by PhosphorImager (Molecular Dynamics). K_{DS} were calculated as described previously (Goetz et al. 2000) using least squares curve fit of fraction of DNA bound = $1/(1 + K_{DS}/[ETS1])$.

Acknowledgments

We thank University of Utah colleagues Brian Dalley from the Microarray Core and Cody Haroldsen from the Informatics Core for assistance. This work was supported by the National Institutes of Health (B.J.G., R01GM38663; P.C.H. fellowship support, T32CA93247; P01CA24014 for Huntsman Cancer Institute) and by the American Cancer Society (P.C.H. post-doctoral fellowship, #PF-03-122-01-GMC). We also acknowledge support from the Huntsman Cancer Institute and the U.S. Department of Energy.

References

- Aerts, S., Van Loo, P., Thijs, G., Moreau, Y., and De Moor, B. 2003. Computational detection of cis-regulatory modules. *Bioinformatics* **19** (Suppl. 2): II5–II14.
- Bailey, T.L. and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the second international conference on intelligent systems for molecular biology* (eds. R. Altman et al.), pp. 28–36. AAAI Press, Menlo Park, CA.
- Batchelor, A.H., Piper, D.E., de la Brousse, F.C., McKnight, S.L., and Wolberger, C. 1998. The structure of GABP α/β : An ETS

- domain-ankyrin repeat heterodimer bound to DNA. *Science* **279**: 1037–1041.
- Beissbarth, T. and Speed, T.P. 2004. Gostat: Find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**: 1464–1465.
- Bina, M., Wyss, P., Ren, W., Szpankowski, W., Thomas, E., Randhawa, R., Reddy, S., John, P.M., Pares-Matos, E.I., Stein, A., et al. 2004. Exploring the characteristics of sequence elements in proximal promoters of human genes. *Genomics* **84**: 929–940.
- Bird, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209–213.
- Blanchette, M., Bataille, A.R., Chen, X., Poitras, C., Laganier, J., Lefebvre, C., Deblois, G., Giguere, V., Ferretti, V., Bergeron, D., et al. 2006. Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Res.* **16**: 656–668.
- Bories, J.-C., Willerford, D.M., Grevin, D., Davidson, L., Camus, A., Martin, P., Stehelin, D., and Alt, F.W. 1995. Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the *Ets-1* proto-oncogene. *Nature* **377**: 635–638.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947–956.
- Brown, T.A. and McKnight, S.L. 1992. Specificities of protein-protein and protein-DNA interaction of GABP α and two newly defined ets-related proteins. *Genes & Dev.* **6**: 2502–2512.
- Buchwalter, G., Gross, C., and Wasylyk, B. 2004. Ets ternary complex transcription factors. *Gene* **324**: 1–14.
- De Ferrari, L. and Aitken, S. 2006. Mining housekeeping genes with a Naive Bayes classifier. *BMC Genomics* **7**: 277. doi: 10.1186/1471-2164-7-277.
- de la Brousse, F.C., Birkenmeier, E.H., King, D.S., Rowe, L.B., and McKnight, S.L. 1994. Molecular and genetic characterization of GABP β . *Genes & Dev.* **8**: 1853–1865.
- Euskirchen, G., Royce, T.E., Bertone, P., Martone, R., Rinn, J.L., Nelson, F.K., Sayward, F., Luscombe, N.M., Miller, P., Gerstein, M., et al. 2004. CREB binds to multiple loci on human chromosome 22. *Mol. Cell. Biol.* **24**: 3804–3814.
- Feng, Y., Goulet, A.C., and Nelson, M.A. 2004. Identification and characterization of the human *Cdc2l2* gene promoter. *Gene* **330**: 75–84.
- FitzGerald, P.C., Shlyakhtenko, A., Mir, A.A., and Vinson, C. 2004. Clustering of DNA sequences in human promoters. *Genome Res.* **14**: 1562–1574.
- Galang, C.K., Muller, W.J., Foos, G., Oshima, R.G., and Hauser, C.A. 2004. Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. *J. Biol. Chem.* **279**: 11281–11292.
- Goetz, T.L., Gu, T.L., Speck, N.A., and Graves, B.J. 2000. Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor $\alpha 2$. *Mol. Cell. Biol.* **20**: 81–90.
- Gottschalk, L.R. and Leiden, J.M. 1990. Identification and functional characterization of the human T-cell receptor β gene transcriptional enhancer: Common nuclear proteins interact with the transcriptional regulatory elements of the T-cell receptor α and β genes. *Mol. Cell. Biol.* **10**: 5486–5495.
- Gu, T.L., Goetz, T.L., Graves, B.J., and Speck, N.A. 2000. Auto-inhibition and partner proteins, core-binding factor β (CBF β) and Ets-1, modulate DNA binding by CBF $\alpha 2$ (AML1). *Mol. Cell. Biol.* **20**: 91–103.
- Hassler, M. and Richmond, T.J. 2001. The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. *EMBO J.* **20**: 3018–3028.
- Hertz, G.Z. and Stormo, G.D. 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**: 563–577.
- Hollenhorst, P.C., Jones, D.A., and Graves, B.J. 2004. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Res.* **32**: 5693–5702.
- John, S., Marais, R., Child, R., Light, Y., and Leonard, W.J. 1996. Importance of low affinity Elf-1 sites in the regulation of lymphoid-specific inducible gene expression. *J. Exp. Med.* **183**: 743–750.
- Jonsen, M.D., Petersen, J.M., Xu, Q., and Graves, B.J. 1996. Characterization of the cooperative function of inhibitory sequences of Ets-1. *Mol. Cell. Biol.* **16**: 2065–2073.
- Kopp, J.L., Wilder, P.J., Desler, M., Kim, J.H., Hou, J., Nowling, T., and Rizzino, A. 2004. Unique and selective effects of five Ets family members, Elf3, Ets1, Ets2, PEA3, and PU.1, on the promoter of the type II transforming growth factor- β receptor gene. *J. Biol. Chem.* **279**: 19407–19420.
- Krig, S.R., Jin, V.X., Bieda, M.C., O'Geen, H., Yaswen, P., Green, R., and Farnham, P.J. 2007. Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays. *J. Biol. Chem.* **282**: 9703–9712.
- Li, R., Pei, H., and Watson, D.K. 2000. Regulation of Ets function by protein-protein interactions. *Oncogene* **19**: 6514–6523.
- Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. 2003. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc. Natl. Acad. Sci.* **100**: 8164–8169.
- Mao, X., Miesfeldt, S., Yang, H., Leiden, J.M., and Thompson, C.B. 1994. The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *J. Biol. Chem.* **269**: 18216–18222.
- Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., et al. 2003. Distribution of NF- κ B-binding sites across human chromosome 22. *Proc. Natl. Acad. Sci.* **100**: 12247–12252.
- Messina, D.N., Glasscock, J., Gish, W., and Lovett, M. 2004. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res.* **14**: 2041–2047.
- Meyers, S., Downing, J.R., and Hiebert, S.W. 1993. Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: The runt homology domain is required for DNA binding and protein-protein interactions. *Mol. Cell. Biol.* **13**: 6336–6345.
- Mo, Y., Ho, W., Johnston, K., and Marmorstein, R. 2001. Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. *J. Mol. Biol.* **314**: 495–506.
- Muthusamy, N., Barton, K., and Leiden, J.M. 1995. Defective activation and survival of T-cells lacking the Ets-1 transcription factor. *Nature* **377**: 639–642.
- Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D., and Graves, B.J. 1992. Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes & Dev.* **6**: 975–990.
- Oberley, M.J., Inman, D.R., and Farnham, P.J. 2003. E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9. *J. Biol. Chem.* **278**:

Hollenhorst et al.

- 42466–42476.
- Oikawa, T. and Yamada, T. 2003. Molecular biology of the Ets family of transcription factors. *Gene* **303**: 11–34.
- Price, M.A., Rogers, A.E., and Treisman, R. 1995. Comparative analysis of the ternary complex factors Elk-1, SAP-1a and SAP-2 (ERP/NET). *EMBO J.* **14**: 2589–2601.
- Ray-Gallet, D., Mao, C., Tavitian, A., and Moreau-Gachelin, F. 1995. DNA binding specificities of Spi-1/PU.1 and Spi-B transcription factors and identification of a Spi-1/Spi-B binding site in the *c-fes/c-fps* promoter. *Oncogene* **11**: 303–313.
- Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes & Dev.* **16**: 245–256.
- Sementchenko, V.I. and Watson, D.K. 2000. Ets target genes: Past, present and future. *Oncogene* **19**: 6533–6548.
- Sharrocks, A.D. 2001. The ETS-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.* **2**: 827–837.
- Shore, P. and Sharrocks, A.D. 1995. The ETS domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. *Nucleic Acids Res.* **23**: 4698–4706.
- Sun, W., Graves, B.J., and Speck, N.A. 1995. Transactivation of the Moloney murine leukemia virus and T-cell receptor β -chain enhancers by *cbf* and *ets* requires intact binding sites for both proteins. *J. Virol.* **69**: 4941–4949.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**: 644–648.
- Tomlins, S.A., Mehra, R., Rhodes, D.R., Smith, L.R., Roulston, D., Helgeson, B.E., Cao, X., Wei, J.T., Rubin, M.A., Shah, R.B., et al. 2006. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res.* **66**: 3396–3400.
- Tomlins, S.A., Mehra, R., Rhodes, D.R., Cao, X., Wang, L., Dhanasekaran, S.M., Kalyana-Sundaram, S., Wei, J.T., Rubin, M.A., Pienta, K.J., et al. 2007. Integrative molecular concept modeling of prostate cancer progression. *Nat. Genet.* **39**: 41–51.
- Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. 1991. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function [corrected]. *Genes & Dev.* **5**: 880–894.
- Wasserman, W.W. and Sandelin, A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* **5**: 276–287.
- Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes & Dev.* **16**: 235–244.
- Wells, J., Yan, P.S., Cechvala, M., Huang, T., and Farnham, P.J. 2003. Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene* **22**: 1445–1460.
- Workman, C.T., Yin, Y., Corcoran, D.L., Ideker, T., Stormo, G.D., and Benos, P.V. 2005. enoLOGOS: A versatile web tool for energy normalized sequence logos. *Nucleic Acids Res* **33** (Web Server issue): W389–W392 doi: 10.1093/nar/gki439.
- Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**: 338–345.
- Zhou, J., Chehab, R., Tkalcovic, J., Naylor, M.J., Harris, J., Wilson, T.J., Tsao, S., Tellis, I., Zavarsek, S., Xu, D., et al. 2005. Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation. *EMBO J.* **24**: 635–644.



Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the *ETS* gene family

Peter C. Hollenhorst, Atul A. Shah, Christopher Hopkins, et al.

Genes Dev. 2007, **21**: originally published online July 24, 2007
Access the most recent version at doi:[10.1101/gad.1561707](https://doi.org/10.1101/gad.1561707)

Supplemental Material

<http://genesdev.cshlp.org/content/suppl/2007/07/23/gad.1561707.DC1>

References

This article cites 55 articles, 29 of which can be accessed free at:
<http://genesdev.cshlp.org/content/21/15/1882.full.html#ref-list-1>

License

Freely available online through the Genes & Development Open Access option.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

horizon
a PerkinElmer company

Streamline your research with
Horizon Discovery's ASO tool

The advertisement features a dark blue background with a glowing DNA double helix structure on the left. The 'horizon' logo is in white, with 'a PerkinElmer company' in smaller white text below it. To the right, the text 'Streamline your research with Horizon Discovery's ASO tool' is displayed in white, with 'Horizon Discovery's ASO tool' in a larger, bold font.