DNA Specificity Determinants Associate with Distinct Transcription Factor Functions

Peter C. Hollenhorst[1], Katherine J. Chandler[1], Rachel L. Poulsen[2], W. Evan Johnson[1,2], Nancy A. Speck[3], Barbara J. Graves[1]*

1 Huntsman Cancer Institute, Department of Oncological Sciences, University of Utah, Salt Lake City, Utah, United States of America, 2 Department of Statistics, Brigham Young University, Provo, Utah, United States of America, 3 Abramson Family Cancer Research Institute, Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

Abstract

To elucidate how genomic sequences build transcriptional control networks, we need to understand the connection between DNA sequence and transcription factor binding and function. Binding predictions based solely on consensus predictions are limited, because a single factor can use degenerate sequence motifs and because related transcription factors often prefer identical sequences. The ETS family transcription factor, ETS1, exemplifies these challenges. Unexpected, redundant occupancy of ETS1 and other ETS proteins is observed at promoters of housekeeping genes in T cells due to common sequence preferences and the presence of strong consensus motifs. However, ETS1 exhibits a specific function in T cell activation; thus, unique transcriptional targets are predicted. To uncover the sequence motifs that mediate specific functions of ETS1, a genome-wide approach, chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq), identified both promoter and enhancer binding events in Jurkat T cells. A comparison with DNase I sensitivity both validated the dataset and also improved accuracy. Redundant occupancy of ETS1 with the ETS protein GABPA occurred primarily in promoters of housekeeping genes, whereas ETS1 specific occupancy occurred in the enhancers of T cell–specific genes. Two routes to ETS1 specificity were identified: an intrinsic preference of ETS1 for a variant of the ETS family consensus sequence and the presence of a composite sequence that can support cooperative binding with a RUNX transcription factor. Genome-wide occupancy of RUNX factors corroborated the importance of this partnership. Furthermore, genome-wide occupancy of co-activator CBP indicated tight co-localization with ETS1 at specific enhancers, but not redundant promoters. The distinct sequences associated with redundant versus specific ETS1 occupancy were predictive of promoter or enhancer location and the ontology of nearby genes. These findings demonstrate that diversity of DNA binding motifs may enable variable transcription factor function at different genomic sites.

Introduction

Transcriptional regulation of gene expression is programmed through DNA sequence elements, termed promoters and enhancers. This genomic hard-wiring represents binding sites for transcription factors that have sequence specific DNA recognition and control development and homeostasis. Although the fundamental properties of protein-DNA recognition are well established, the advent of powerful technologies that provide genome-wide occupancy data has only recently allowed observation of these interactions *in vivo*. The emerging picture is that no single sequence motif fully explains all *in vivo* binding [1–3]. Furthermore, the *in vitro* derived consensus motifs are often present in only a minority of bound regions. These findings bring into question the purpose of binding site sequence variations. Possibilities are illustrated by experimental analysis of subsets of sites gathered from genomic data. For example, the PHA4/FOXO binding sites that program pharynx development in *C. elegans* differ in affinity, and thus carry

developmental programming information dictating time of expression [4]. In yeast, *PHO4* responsiveness to phosphate levels is regulated by alterative sequence motifs that affect affinity and program different roles for binding sites [5]. NF-κB and GR binding site variants can alter the repressing or activating transcriptional activity of the factor once it is bound [6,7]. The challenge of genomic databases is how to take full advantage of the vast number of binding sites, yet parse out functional consequences of variation. To realize their full potential, genomic approaches to transcriptional networks must go beyond a description of factor occupancy to include correlates of functionality.

We focus on the transcription factor ETS1 that provides a variety of contexts to address these central questions. ETS1 is a member of the ETS family of transcription factors that display similar DNA binding properties, including the recognition of a core GGA(A/T) motif. ETS family members are extensively co-expressed [8,9]. For example mRNA of 17 *ets* genes, including *ETS1*, is present in Jurkat T cells. Despite overlapping expression

Author Summary

Genomes contain sequences that encode both gene products and the instructions for where and when each gene is expressed. This gene expression code is critical for normal development and goes awry in disease processes such as cancer. The gene expression code is interpreted by proteins called transcription factors that bind to particular DNA sequences and carry instructions for gene activation or repression. This recognition code is challenged by the presence of highly-similar transcription factors that prefer almost identical DNA sequences. In addition, studies in living cells indicate that individual transcription factors have significant flexibility in sequence recognition. Here, we identify thousands of positions in the genome of human T cells that are bound by the transcription factor ETS1. These data, along with comparisons to other genomic datasets, allow us to identify DNA sequences that specify ETS1 binding while excluding binding of other related transcription factors. Furthermore, we discover that ETS1 binds more than one sequence and that these sequence variants can predict distinct biological functions of ETS1. Thus, this work contributes to our understanding of the gene expression code by addressing both how a transcription factor can bind unique genomic locations and why a transcription factor binds multiple DNA sequences.

and sequence preferences, experimental data indicate that individual ETS proteins have unique biological functions [10–13]. For ETS1, mouse deletion studies indicate a critical role in T cell activation [14]. This specific genetic function implies that ETS1 has a unique mechanism that allows it, but not other ETS proteins, to bind the promoters or enhancers of genes important for T cell activation. Finally, ETS1 functions, in part, by recruitment of the co-activator CBP to transcriptional control regions, presumably functioning to activate genes at which it binds [15–17]. We utilize this in depth understanding of ETS1 at both a biochemical and biological level to inform our genomic approach and facilitate functional analysis.

Initial genomic occupancy studies with ETS1 have provided insight into the genomic dilemma and led to unexpected observations. We previously identified two modes of ETS protein targeting to promoters in Jurkat T cells using chromatin immunoprecipitation and promoter microarrays (ChIP-chip). The surprising mode is redundant occupancy in which a sequence with the consensus CCGGAAGT is associated with occupancy of three different ETS transcription factors: ETS1, GABPA (GABPα), and ELF1. Because this sequence is consistent with the *in vitro* derived consensus sequences derived from multiple ETS family members, we concluded that it can alternately recruit various ETS transcription factors. This redundant mode of binding generally occurs in the promoters of housekeeping genes and may represent shared function of the ETS family in the maintenance of constitutive expression. The second ETS binding mode is specific occupancy (e.g. ETS1, but not GABPA or ELF1), which requires a GGA core motif, but is not associated with a close match to the consensus ETS sequence. We proposed that specific targets would mediate the specific biological functions of each ETS transcription factor. However, the promoter-limited approach did not identify a significant correlation between the specific targets of ETS1 and genes important for the role of ETS1 in T cell activation. Full investigation of this provocative dual role of the ETS family required an expansion to full genome analysis.

In this study we identified regions across the entire human genome occupied by ETS1, a DNA binding partner RUNX, and co-activator CBP in Jurkat T cells to decipher sequence determinants and investigate the biological significance of sequence diversity. We discovered a previously undescribed role for ETS1 at a large number of enhancers. Enhancer occupancy of ETS1 was associated with a unique variant of the ETS binding site and *in vitro* DNA binding assays illustrated how this variant sequence functions as an ETS1 specificity determinant. Enhancers co-occupied by ETS1 and RUNX contained a variant ETS sequence closely juxtaposed to a RUNX binding site – a composite sequence identical to that found in the T cell receptor enhancers. These distinct enhancer sequences contrasted with prior observations of sequences at ETS1 bound promoters. Importantly, ETS1 bound regions that contained the ETS/RUNX composite sequence were near genes important for T cell activation, thus establishing a tissue-specific, genomic dataset for a factor partnership. Furthermore, ETS1 was closely associated with CBP occupancy at ETS1 specific enhancers, but not at redundantly occupied promoters. By using genomic datasets for DNA binding factors, in addition to correlates of DNase I sensitive regions, histone marks, and co-factor binding, we decoded the functionality of *in vivo* binding sequences.

Results

ETS1 and GABPA co-occupy active promoters, but ETS1 specifically occupies T cell enhancers

High-throughput sequencing coupled with chromatin immuno-precipitation (ChIP-seq) facilitates genome-wide searches for transcription factor binding sites [1]. We detected 19,420 bound regions at an empirical false discovery rate of <0.01 for ETS1 in Jurkat T cells using this approach. This included almost all (94%) of the 1086 ETS1 bound promoters previously identified by ChIP-chip [18] plus an additional 6116 promoters, indicating a potential for higher sensitivity. ETS1 bound promoter regions centered within 500 bp of a transcription start site (TSS) (Figure S1); therefore, a 500-bp limit was used for promoter definition. A large number of regions, 12,283, were not in promoters. We sought to establish the validity of these potential enhancer regions by comparison to other types of genome-wide datasets. One powerful dataset from primary CD4$^+$ T cells (thus comparable to the CD4$^+$ Jurkat cell line) identifies DNase I accessible regions as mapped by high-throughput sequencing and ChIP-chip [19]. Based on the long history of linkage of DNase I sensitivity to enhancers, we screened ETS1 bound regions for overlap. 76% of ETS1 occupied regions overlapped with DNase I sensitivity. (Overlap was 98% for sites proximal to a TSS and 64% for distal sites.) This represents a significant enrichment over the mean 4% overlap with datasets randomly derived from control sequences ($P<0.001$, Figure 1A). Randomly selected DNase I sensitive, ETS1 bound regions were verified by quantitative PCR as ETS1 occupied (13 of 15 ETS1 bound, Figure S2), whereas regions that were not DNase I sensitive included many apparent false positives (0 of 8 ETS1 bound, Figure S2). This strong correlation not only helped validate the ETS1 data, but also suggested that DNase I sensitivity is a strong correlate of robust ChIP signals. We proposed that the 14,824 ETS1 bound regions that overlap DNase I sensitive regions represent functional regions, and only these were considered in further analysis.

Datasets for histone marks in primary CD4$^+$ T cells [20] also provide a measure of the activity of promoters and enhancers and test for relevance of factor binding. For example, H3K4 tri-methylation correlates with active promoters [21]. 86% of ETS1
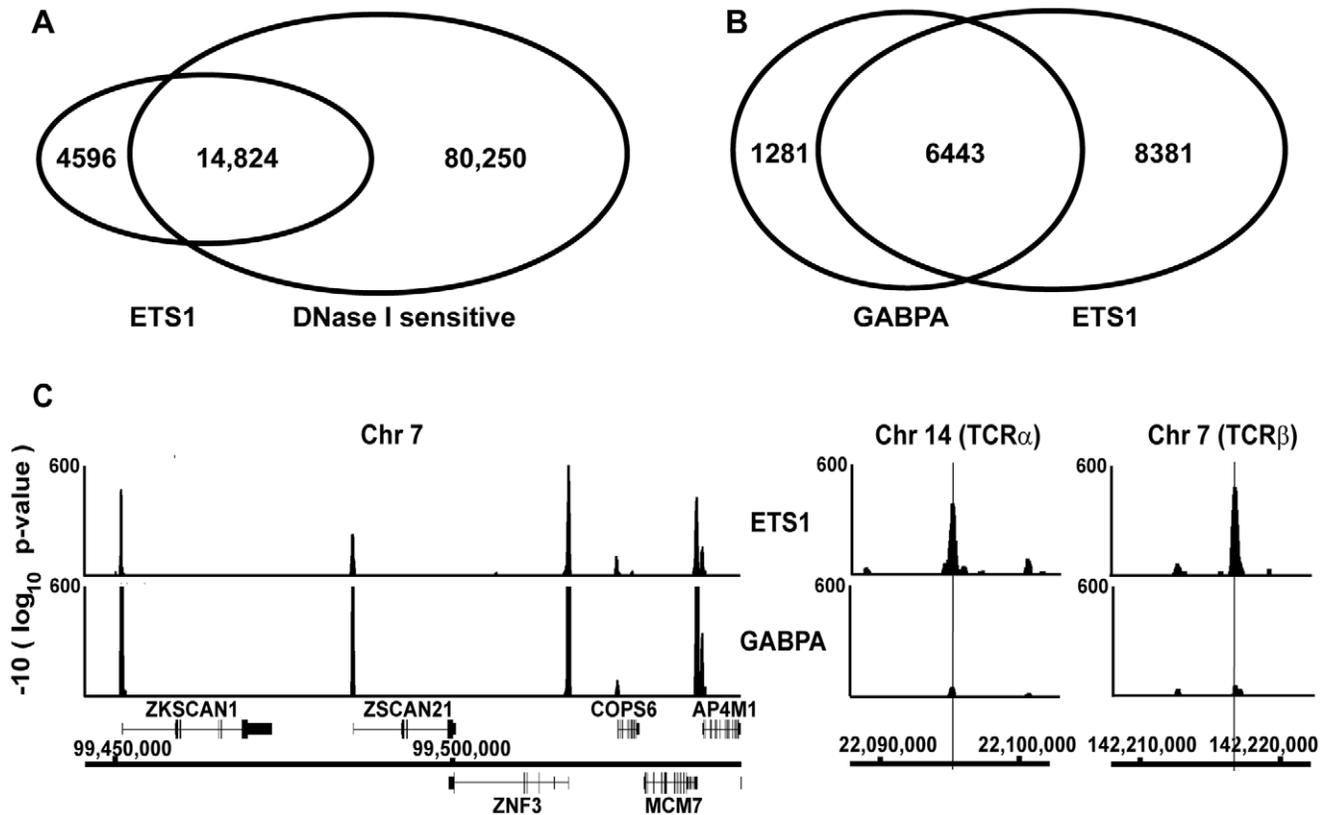
Figure 1. Genomic occupancy of ETS1 overlaps with DNase I sensitivity and GABPA occupancy. (A) Overlap of genomic regions bound by ETS1 in Jurkat T cells and regions found to be DNase I sensitive in CD4[+] T cells [19]. (B) Overlap of regions occupied by ETS1 and GABPA in Jurkat T cells [23]. Only regions that overlap with DNase I sensitivity were included. (C) Log transformed P values of ETS1 and GABPA occupancy in a scanning 250 bp window mapped using the Integrated Genome Browser (http://igb.bioviz.org/) to regions of the human chromosome (Chr) indicated by chromosomal coordinates (NCBI Build 36.1). Positions of Refseq genes are shown with genes transcribed from left to right above the nucleotide position bar and genes in the opposite orientation below. Vertical lines (right) indicate the ETS/RUNX binding sequences previously tested for function [27,44] in the *TCRα* enhancer (sequence GAGGATGTGGC) or the *TCRβ* enhancer (sequence CAGGATGTGGT).
doi:10.1371/journal.pgen.1000778.g001

bound promoters had H3K4 tri-methylation compared to 28% of promoters without ETS1 ($P<0.0001$) (Table 1). Likewise, H3K4 mono-methylation is enriched in enhancers [21,22]. For DNase I sensitive regions distal to a TSS, 52% with ETS1, but only 18% without ETS1, had an H3K4 mono-methyl mark ($P<0.0001$) (Table 1). Therefore, ETS1 occupancy is enriched at regions with histone marks that are indicative of enhancers and active promoters.

Contrary to the expectation that family members with unique genetic functions would have exclusive binding sites, we previously discovered the majority of proximal promoters bound by ETS1, are also occupied redundantly by other ETS proteins (e.g. GABPA and ELF1). The small number of ETS1 specific sites identified limited the robustness and fruitfulness of further analysis of this class of targets [18]. Using a much larger ChIP-seq dataset that

Table 1. Marks of enhancers and active promoters are associated with ETS1 occupancy.

Category[a]	Number[b]	H3K4me3[cd]	H3K4me1[cd]	CBP[d]
Promoters occupied by ETS1	8236	86%	16%	75%
Promoters lacking ETS1	14,445	28%	11%	9%
DNase I sensitive promoters lacking ETS1	5961	58%	19%	20%
Distal DNase I sensitive regions occupied by ETS1	7501	46%	52%	68%
Distal DNase I sensitive regions lacking ETS1	77,141	10%	18%	5%

a Promoters are defined as 500 bp upstream and downstream of a RefSeq TSS. Distal regions are regions with the center greater than 500 bp from a RefSeq TSS. ETS1 occupied regions contain the center of an ETS1 bound region within the promoter or distal region. Totals differ from the number of ETS1 bound regions (Figure 1) because in some cases multiple promoters or DNase I sensitive regions overlap with a single ETS1 bound region.
b The number of regions in the category.
c Positions of histone marks in CD4[+] T cells were determined by analyzing published ChIP-seq reads [20] with the Useq bioinformatics package.
d The percentage of regions in the category containing the indicated histone mark or CBP occupancy.
doi:10.1371/journal.pgen.1000778.t001

includes enhancer regions, we were poised to identify targets that mediate specific functions of ETS1. However, we first had to identify which of the 14,824 ETS1 bound regions were redundant sites, thus not strong candidates to mediate specific functions. We analyzed genome-wide occupancy data reported for GABPA in Jurkat cells [23] with the same methodology as our ETS1 analysis. There were 7724 GABPA bound regions, of which 6443 were redundantly bound by ETS1 (Figure 1B), illustrated graphically on a genome section in Figure 1C (left). The remaining 8381 ETS1 specific regions were exemplified by the T cell receptor (*TCR*) α and β enhancer loci (Figure 1C, right), which display ETS1, not GABPA, occupancy. 67% of GABPA and ETS1 co-occupied regions were proximal to a TSS consistent with previous findings that discovered this dominant class of promoter binding events [18]. In contrast, 68% of ETS1 specific regions were distal to the nearest TSS indicating enhancer regions and are candidates to mediate the specific functions of ETS1.

Genetic experiments have implicated ETS1 in T cell activation; therefore, we predicted that genomic sites that specifically bind ETS1 should be associated with genes necessary for T cell function. The genes nearest to distal ETS1 bound regions were assumed to be a reasonable estimate of ETS1 regulated genes. This gene set was compared to genes with 20-fold higher mRNA expression levels in CD4$^+$ T cells than the median expression in multiple cell types (287 genes) and, as a control, to 268 genes specific to pancreas (Figure 2). Compared to all genes, or pancreas specific genes, genes that displayed T cell–specific expression were more likely to be near one or more distal ETS1 bound regions. Furthermore, as the number of nearby distal ETS1 bound regions increased, the difference between the T cell–specific categories and the control categories became more apparent. Thus, distal ETS1 binding was associated with a tissue-specific role of ETS1 in T cells, further validating the functionality of ETS1 specific regions.
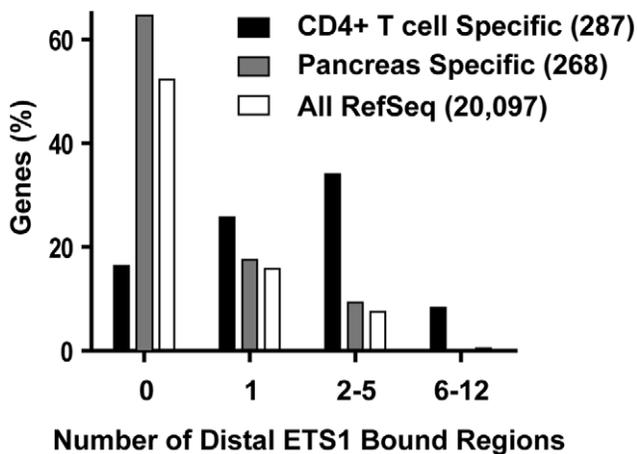

Figure 2. Distal ETS1 bound regions are found near T cell–specific genes. The frequency of neighboring, distal, ETS1 bound regions was compared for three categories of genes; all Refseq genes, CD4$^+$ T cell–specific genes, and pancreas specific genes. Tissue specific gene lists were derived from the GNF SymAtlas database [52] and were based on the level of mRNA in T cells or pancreas compared to the median in all surveyed tissues with cutoffs (20-fold higher for T cells, 5-fold higher for pancreas) that returned similar sized lists. The number of genes in each category is indicated in parenthesis. Each ETS1 bound region was matched to a single gene based on the nearest RefSeq TSS. The percent of genes in each category associated with one or more distal ETS1 bound regions (greater than 500 bp from the TSS) was plotted.
doi:10.1371/journal.pgen.1000778.g002

ETS1 occupancy of enhancers was associated with a distinct sequence

At this point we had a dataset of ETS1 bound regions that tracked with distal enhancer marks and T cell function. We sought to determine whether such bound regions displayed a unique sequence that would be responsible for ETS1 specific binding. Unbiased searching for overrepresented sequences was performed with the MEME algorithm [24]. ETS1 bound regions were grouped according to proximity to the nearest TSS (proximal versus distal) and specificity (redundant: overlap with GABPA; specific: no overlap with GABPA) to provide experimental and control datasets. The most over-represented sequence motifs in redundant, proximal regions (Motif 1) and specific, proximal regions (Motif 2) were identical to the motifs previously identified in redundant and specific promoter proximal regions [18]. In contrast, analysis of distal, specific ETS1 bound regions identified a third, distinct motif (Motif 3) as the most over-represented (Figure 3). The major differences from the ETS family consensus (CCGGAAGT) present in Motif 1 were the almost exclusive presence of an A at the second position and the inclusion, in some instances, of a T at the sixth position (CAGGA(A/T)GT). Therefore, specific ETS1 binding to enhancers is associated with a sequence distinct from those found at ETS1 bound promoters.

A more directed bioinformatics approach assessed the importance of single nucleotide changes from the ETS consensus for enhancer and promoter occupancy of ETS1. ETS1 bound regions were partitioned into equally-sized sets of ETS1/GABPA redundantly occupied promoters and ETS1 specific enhancers. The number of occurrences of 8-mer sequences was reported relative to the number of occurrences expected in a set of equally-sized random sequences (Table 2). The ETS consensus sequence (CCGGAAGT) was enriched in redundant promoters, but not in ETS1 specific enhancers. The enrichment of every possible single nucleotide change to the ETS consensus was then determined. Only the change of the C at the second position to an A (CaGGAAGT) resulted in a significant enrichment ($P<0.0001$) in ETS1 specific enhancers. However, this sequence was also enriched in redundant promoters. The A to T change at the sixth position (CCGGAtGT) reduced significantly the enrichment in redundant proximal regions ($P<0.0001$), but not distal regions ($P=0.3$). Furthermore, the combination of both nucleotide changes (CaGGAtGT) resulted in enrichment at ETS1 specific enhancers ($P<0.0001$), but not redundant promoters ($P=0.3$). Strikingly, a change of only two nucleotides inverted the enrichment pattern at redundant promoters and ETS1 specific enhancers. We concluded that the sequence CAGGATGT is a specificity element for ETS1.

In considering specificity within families of transcription factors the preference for a particular sequence may be due to intrinsic DNA binding properties of different family members. To test the ability of the two nucleotide changes to act alone or in combination to select for ETS1 we measured the relative binding affinity of ETS1 and second ETS factor, ELF1 which is also reported to be active in T cells [25]. Indeed, ELF1 is present at redundant, but not ETS1 specific promoters in Jurkat T cells in a similar manner to GABPA [18]. Binding affinity for an ETS consensus sequence, each single nucleotide variant, and the two nucleotide variant was interrogated *in vitro* with purified proteins by quantitative gel shift (Figure 4). The A to T change resulted in a loss of affinity for ELF1 (3.6-fold loss), but not for ETS1. The C to A change showed a modest effect on affinity and no discrimination between ELF1 and ETS1 (1.9-fold versus 1.5-fold loss). In contrast, the change of both nucleotides caused an 18.3-fold loss of affinity for ELF1, but only a 2.4-fold loss for ETS1. We
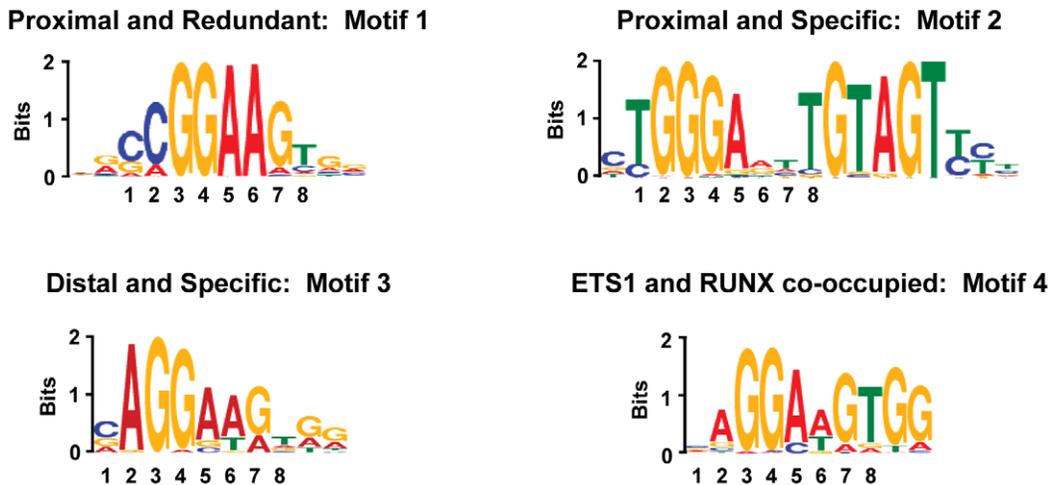
Proximal and Redundant: Motif 1

Proximal and Specific: Motif 2

Distal and Specific: Motif 3

ETS1 and RUNX co-occupied: Motif 4

Figure 3. Distinct sequence motifs are over-represented in different subsets of ETS1 bound regions. The indicated subsets of ETS1 bound regions were rank ordered by log transformed binomial P value and the top 250 regions were searched for over-represented sequences by MEME [24]. The most over-represented position weight matrix for each subset is represented (E-values: Motif 1, 9.9×10^{-281}; Motif 2, 2.8×10^{-287}; Motif 3, 8.8×10^{-105}; Motif 4, 1.8×10^{-142}). The height of each nucleotide indicates conservation at that position. Eight nucleotide positions in ETS binding sites are numbered for reference. ETS bound regions were classified either proximal (center of region within 500 bp of a TSS) or distal (center of region greater than 500 bp from a TSS). ETS1 bound regions were classified as redundant if they overlapped with a GABPA bound region and specific if they did not. ETS1 bound regions were classified as RUNX co-occupied if they overlapped with a RUNX bound region.
doi:10.1371/journal.pgen.1000778.g003

concluded that the two nucleotide variant sequence CaGGAtGT serves as a specificity determinant for ETS1 versus ELF1 due to an intrinsic binding property of ETS1.

RUNX co-occupies enhancers with ETS1 through a distinct sequence motif

Unique cooperative DNA binding between closely apposed binding proteins can also drive specific occupancy of transcription factors. A well-characterized partnership for ETS1 is with the RUNX factors [26,27]. Interestingly, the consensus derived from the most frequent nucleotides at each position of Motif 3 (CAGGAAGTGG) is similar to the sequences at the *TCRβ* (CAGGATGTGG) and *TCRα* (GAGGATGTGG) enhancers that support cooperative binding of ETS1 with RUNX1 through an ETS/RUNX composite sequence (RUNX consensus YGYGGY). To test whether ETS1 bound enhancers were co-occupied by RUNX factors, genome-wide occupancy of RUNX1/3 (RUNX) was determined. Again, only regions that co-localized with DNase I sensitivity were considered bound. Strikingly, 64% of the 1075 RUNX bound regions were co-occupied by ETS1. In contrast, only 14% of RUNX bound regions were co-occupied by GABPA (Figure 5A). 77% of the ETS1/RUNX co-occupied regions were ETS1 specific and distal to a TSS (compared to 37% of ETS1 bound regions lacking RUNX), suggesting a role in T cell enhancer function. An unbiased search with MEME for overrepresented sequence motifs in regions co-occupied by ETS1 and RUNX identified a motif (Motif 4) similar to Motif 3, but with the RUNX binding site more strongly represented (Figure 3). Like many sequence identification algorithms, MEME is biased towards strongly preferred spacing distances between two binding sites. To test whether other spacings of ETS and RUNX sites were also over-represented in ETS1/RUNX bound regions, the distance from each ETS sequence to the nearest RUNX sequence was plotted (Figure 5B). This analysis indicated that only the spacing found by MEME was over-represented in these regions. Therefore, ETS1 and RUNX co-occupy enhancer regions in T cells through a composite ETS/RUNX binding site

similar to those found in the T cell receptor enhancers. These findings indicate that pairing with a neighboring DNA binding motif, in conjunction with intrinsic DNA binding properties, can drive specificity.

Distinct sequence motifs are correlated with unique modes of ETS1 occupancy

ETS1 occupancy of enhancers is associated with T cell–specific genes (Figure 2) and ETS1 specific motifs (Motifs 3 and 4), whereas promoter occupancy is associated with housekeeping genes [18] and shows enrichment for sequences (Motif 1) that cannot distinguish family members. The value of these motifs will be in their predictive accuracy. All ETS1 bound regions were searched for Motif 1 and Motif 4 with PATSER [28]. Regions containing Motif 1 were more likely to be found in promoters, and regions containing Motif 4 were more likely to be found in enhancers (Table 3, Table S1). Associated genes, as determined by nearest TSS, were searched for over-represented ontologies with the GoMiner program [29]. Genes with Motif 4 were associated with T cell activation categories, whereas those with Motif 1 were associated with housekeeping ontologies (Table 3). Therefore, each motif is predictive of the type of transcriptional control element and class of ETS1 target gene.

CBP/p300 co-localizes with ETS1 at enhancers, but not at promoters

The emerging differences for ETS1 at promoters versus enhancers opened the possibility of distinct functions of ETS1 at these loci. One mechanism of transcriptional activation by ETS1 is the recruitment of the co-activators CBP and p300 [16]. Identification of p300 occupancy within the 30 mb ENCODE region of the human genome revealed a greater proportion at distal sites than at promoters [21], and p300 has been shown to mark tissue specific enhancers in mice [30]. Thus, we proposed that ETS1 would recruit CBP/p300 to T cell–specific enhancers, but not promoters. Genome-wide occupancy for CBP detected 14,374 CBP bound/DNase I sensitive regions in Jurkat T cells.

Table 2. ETS1 occupancy of redundant promoters and specific enhancers is associated with distinct sequences.

Sequence[a]	Enrichment in redundant and proximal regions[b]	Enrichment in specific and distal regions[b]
CCGGAAGT	33	1
gCGGAAGT	15	1
aCGGAAGT	4	1
tCGGAAGT	2	0
CgGGAAGT	2	1
CaGGAAGT	10	14
CtGGAAGT	2	2
CCcGAAGT	1	0
CCaGAAGT	1	2
CCtGAAGT	1	1
CCGcAAGT	0	1
CCGaAAGT	0	0
CCGtAAGT	1	0
CCGGgAGT	2	0
CCGGcAGT	2	0
CCGGtAGT	1	0
CCGGAgGT	2	1
CCGGAcGT	2	0
CCGGAtGT	4	1
CCGGAAcT	4	0
CCGGAAaT	5	1
CCGGAAtT	1	0
CCGGAAGg	9	1
CCGGAAGc	23	1
CCGGAAGa	8	1
CaGGAtGT	1	10

a The ETS consensus from Motif 1 (Figure 3), every possible single nucleotide change, and one double nucleotide change are shown.
b The number of occurrences of each octamer in equally-sized data sets of ETS1 bound, redundant, proximal regions and ETS1 bound, specific, distal regions were counted. Enrichment was determined by dividing this number by the expected number of occurrences of an octamer in random sequence space of the same size and rounding to the nearest integer. All enrichment values greater than two are significant, $P<0.0001$, Fisher's exact test.
doi:10.1371/journal.pgen.1000778.t002

CBP bound regions showed a surprisingly high overlap with ETS1 bound regions at both redundant promoters (75%, $P<0.0001$) and ETS1 occupied enhancers (68%, $P<0.0001$) compared to regions not bound by ETS1 (Table 1). The strong presence of CBP corroborated the general functionality of ETS1 binding sites.

Due to the unexpected equivalence of CBP overlap at both enhancers and promoters we investigated the connection between CBP binding and ETS1 function by a more detailed mapping method that presented the two types of sites as a class average (Figure 6A). At redundantly occupied promoters, the location of ETS1, CBP, GABPA, H3K4 tri-methylation, and Motif 1 were plotted relative to the TSS. At ETS1 occupied enhancers the location of CBP, RUNX, H3K4 mono-methylation, and Motif 4 were plotted relative to the center of the ETS1 bound region. At promoters ETS1 and GABPA binding were coincident with the consensus ETS binding site at a position 25–30 bp upstream of the transcription start site. This extremely TSS proximal location and

the location of histone H3K4 tri-methylation on either side of the ETS1 bound region indicated that redundant ETS binding occurs in the nucleosome-free region [31]. CBP occupancy was co-incident with the downstream H3K4 tri-methyl, but not ETS1 and GABPA binding, suggesting that CBP is not directly bound by ETS factors at promoters. In contrast, at enhancers, CBP, ETS1, and RUNX binding overlapped, suggesting that ETS1 and/or RUNX may directly bind CBP. Again, ETS1 occupied a region between histone marks, in this case H3K4 mono-methyl, indicating that ETS1 binds between nucleosomes.

To test directly whether ETS1 was necessary for occupancy of RUNX and CBP, ETS1 protein levels were knocked-down by two independent shRNAs and occupancy was monitored by ChIP (Figure 6B and 6C). Decreased ETS1 protein levels correlated with a loss of ETS1, CBP, and RUNX ChIP enrichment at the *TCRβ* enhancer (containing Motif 4). We concluded that ETS1 is critical for recruitment or stable binding of CBP in enhancers important for T cell activation. In contrast, reduction of ETS1 occupancy in a redundantly occupied promoter (containing Motif 1) did not affect CBP enrichment. Thus, distinct sequence motifs at ETS1 binding sites correlate with not only different types of regulatory elements, but also distinct histone marks and co-activator binding. We conclude that these sequences mediate unique functions of ETS1.

Discussion

Comprehensive identification of ETS1 binding sites in Jurkat T cells revealed that ETS1 was present at a large number of enhancers in a context distinct from that previously observed at promoters. Enhancers differed from promoters in the sequence elements that bind ETS1, in the compliment of neighboring proteins and histone marks, and in the ontology of nearby genes. Specifically, the tissue specific function of ETS1 correlated with enhancer binding via an ETS/RUNX composite sequence. Furthermore, our data indicated that this T cell–specific enhancer function, but not the housekeeping promoter function of ETS1, is associated with co-localization of the co-activator CBP. Therefore, the sequences motifs identified here are associated with specific enhancer occupancy of ETS1 and define a different function than the sequence associated with redundant ETS1 occupancy at promoters.

Overlap with DNase I sensitivity improves the accuracy of a ChIP-seq dataset

Selecting regions at which ChIP-seq enrichment coincided with DNase I sensitivity improved the accuracy of a dataset of transcription factor bound regions. The fraction of regions removed (24% of ETS1 bound regions, 39% of CBP bound regions and 70% of RUNX regions) may reflect the quality of the antibodies used for ChIP-seq. Removed regions had ETS1 binding properties (presence of binding motifs, correlations with other factors and histone marks), but at markedly lower levels than retained regions. Thus, we propose that the use of DNase I sensitivity screening improves the quality of a ChIP-seq dataset and may be particularly useful for the interpretation of data generated with suboptimal antibody reagents.

ETS1 binding determinants vary at specific enhancers and redundant promoters

The genome-wide set of ETS1 binding sites showed sequence variants that distinguish enhancer versus promoter binding events. Specific ETS1 occupancy of enhancers was associated with a sequence that varies by two nucleotides from the ETS consensus sequence used for redundant binding at promoters. Our bioinfor-
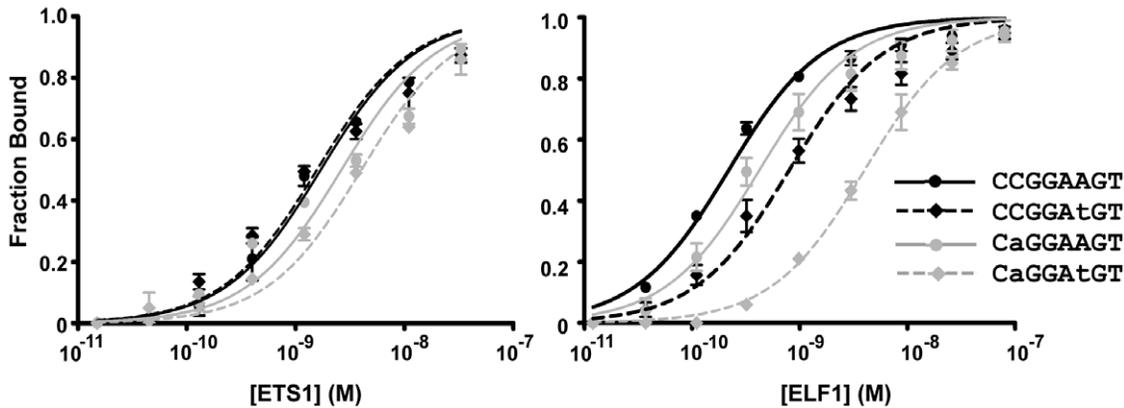
Figure 4. Intrinsic DNA binding affinity differences for ETS1 and ELF1. Full-length recombinant versions of the human ETS proteins ETS1 and ELF1 were purified from bacteria and assayed for affinity to radiolabeled oligonucleotides by gel-shift analysis. Datapoints represent the mean and standard error of the mean of two replicates for ETS1 and three replicates for ELF1. Each K_D was derived by curve fitting by nonlinear least-squares analysis of equilibrium binding curves with fraction of DNA bound = $1/(1+ K_D/[ETS1])$. The K_D of ETS1 and ELF1 for the ETS consensus sequence was 1.7×10^{-9} and 1.7×10^{-10} M, respectively. The fold decrease in affinity due to the A to T change, the C to A change, and the combination of both changes were 1.0, 1.5, and 2.4 for ETS1 and 3.6, 1.9, and 18.3 for ELF1, respectively.
doi:10.1371/journal.pgen.1000778.g004

matics analysis indicated that these two nucleotide changes are not equivalent (Table 2). The C to A change at the second position appeared to be required for specific ETS1 binding to enhancers, but also occurs at redundant promoters. In contrast, the A to T change at the sixth position appeared to restrict occupancy of redundant promoters, but not specific enhancers. The A to T change has previously been shown to provide specificity for ETS1 versus the ETS protein ELF1 *in vitro* [32]. Our *in vitro* comparison confirmed the role of this single nucleotide change and identified a dramatic specificity difference between ETS1 and ELF1 when both nucleotides were changed (Figure 4). However, the *in vitro* data did not explain why the C to A change alone appeared necessary for genomic enrichment in ETS1 specific enhancers (Table 2). Therefore, the nucleotide preferences at these ETS

binding sequences are likely due to a combination of the intrinsic differences in DNA binding attributes of ETS proteins and other *in vivo* factors.

A striking difference between the ETS consensus sequence, CCGGAAGT, and the C to A variant, CaGGAAGT, is the susceptibility to DNA methylation. Indeed, methylation of this CpG dinucleotide within the consensus has been shown to block the binding of ETS proteins [33,34]. We have previously observed a very strong correlation between redundant ETS occupancy of promoters, the sequence CCGGAAGT and CpG islands [18]. The CpG islands at housekeeping promoters are generally hypomethylated, whereas CpG dinucleotides distributed in lower density throughout the genome are likely to be methylated [35,36]. Thus, ETS sites may be shielded from methylation at CpG island-
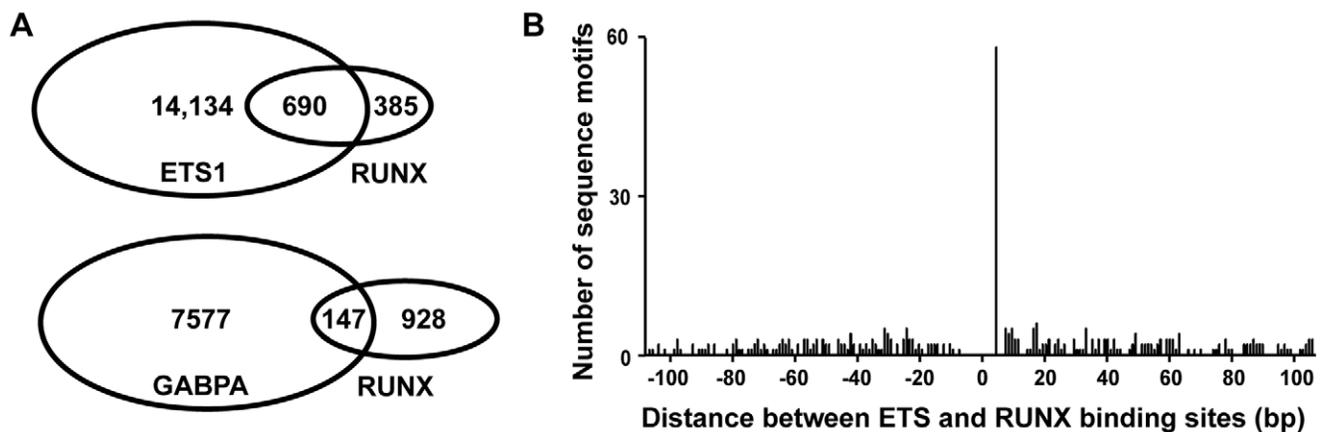

Figure 5. ETS1/RUNX co-occupancy correlates with specific spacing of ETS and RUNX binding sites. (A) Overlap of ETS1/RUNX and GABPA/RUNX bound regions. The RUNX antibody was raised against the conserved DNA binding domain and does not differentiate between the homologous RUNX1 and RUNX3 proteins present in T cells (N. Speck unpublished observation). (B) Spacing of ETS and RUNX binding sites in ETS1/RUNX co-occupied regions. The 690 ETS1 bound regions that were co-occupied by RUNX were scanned for matches to the *in vitro* derived position weight matrixes for ETS1 (M00032) and RUNX (M00271) from the Transfac database (http://www.biobase-international.com/index.php?id=transfac). For each ETS1 sequence found, the distance to all forward oriented RUNX sequences in the same region were determined such that a RUNX site 5′ to an ETS site in the orientation CCGGAAGT was negative and 3′ was positive. A similar mapping of RUNX sites in the reverse orientation returned no spacing frequencies higher than five. The prominent peak at a spacing of five bp correlates with the spacing and orientation in the composite sequence CAGGATGTGGT, from Motif 4 (Figure 3).
doi:10.1371/journal.pgen.1000778.g005

Table 3. ETS1 bound regions containing Motif 1 and Motif 4 have different characteristics.

Motif	Regions[a]	Promoter[b]	Most overrepresented ontologies[c]	P value[d]
4	1364	24%	Leukocyte activation	7.6×10^{-10}
			Phosphate metabolic process	9.9×10^{-9}
			Cell activation	1.4×10^{-8}
			Lymphocyte activation	3.1×10^{-8}
			Protein amino acid phosphorylation	3.7×10^{-8}
			T cell activation	1.2×10^{-7}
			Positive regulation of lymphocyte activation	1.3×10^{-7}
			Immune system process	1.9×10^{-7}
1	4492	59%	Macromolecule metabolic process	4.9×10^{-29}
			Biopolymer metabolic process	1.5×10^{-26}
			Primary metabolic process	1.1×10^{-20}
			Cellular metabolic process	1.7×10^{-20}
			Gene Expression	2.3×10^{-19}
			Nucleotide and nucleic acid metabolism	2.8×10^{-18}
			Metabolic process	3.7×10^{-17}
			RNA processing	7.3×10^{-16}

a Number of ETS1 bound regions containing a match to the PWM for Motif 1 and 4 as shown in Figure 3. PWM nucleotide frequencies and cutoffs are detailed in the Materials and Methods and Table S2.
b Percentage of regions with the center within 500 bp of a RefSeq TSS.
c Regions were mapped to the nearest RefSeq gene and gene lists were analyzed by GoMiner. Overrepresented ontologies are listed in the order returned by GoMiner with no editing.
d P value for each ontology category from GoMiner.
doi:10.1371/journal.pgen.1000778.t003

containing promoters, but not at enhancers. Therefore, the C to A change in the ETS binding sequences at enhancers may have evolved to protect these sites from the repressive effects of DNA methylation. Because other transcription factors whose binding sites bear a CpG dinucleotide (NRF-1, BoxA, SP1, CRE, and E-box) are also enriched in housekeeping promoters [37], the use of an alternate sequence in tissue-specific enhancers may also extend to these transcription factor families.

Another factor that might influence the ETS1 binding sequence observed *in vivo* is the presence of closely juxtaposed binding sites for other transcription factors. A subset of the ETS1 specific enhancers were co-occupied by RUNX and had a composite ETS/RUNX binding sequence (Motif 4). In the context of this sequence, the A to T change at the sixth position of the ETS sequence allows the RUNX sequence to be a closer match to the RUNX consensus (YGYGGT, sixth position underlined). This factor could contribute to the enrichment of the A to T change in specific enhancers. However, a T at the sixth position was no more likely in regions co-occupied by ETS1 and RUNX (Motif 4) than at ETS1 specific enhancers in general (Motif 3). This indicates that either an A or a T at this position can support ETS1 and RUNX co-occupancy. Furthermore, only 55% of the ETS1 bound regions that had the sequence CAGGATGT, had the full ETS/RUNX composite CAG-GATGTGG. We propose that the remaining 45% of regions recruit ETS1 either only through the ETS binding site, or in cooperation with other unidentified transcription factors. In conclusion, the sequences associated with ETS1 specific occupancy of enhancers reflect intrinsic differences in DNA binding or interactions with other factors and may be influenced by susceptibility to DNA methylation.

ETS1 and RUNX co-occupancy predict tissue-specific enhancers

Mice with an *ETS1* gene disruption have reduced numbers of NK and NKT cells and show defects in T cell activation [14,38,39]. RUNX genes are essential for NK, NKT, and T cell differentiation [40–42]. However, the role of ETS1 and RUNX in these immune functions has not been fully understood on the level of individual target genes. Our data suggest that ETS1 and RUNX regulate genes important for T cell activation pathways by direct occupancy of nearby enhancers via a particular ETS/RUNX binding site. The gene categories presented in Table 3 suggest that the primary role of these transcription factors is not the direct activation of genes downstream of T cell receptor signaling, but rather the control of expression of the signaling machinery. This conclusion is consistent with the finding that *ETS1* null T cells are defective in activation upon receptor stimulation, but respond normally to pharmacological stimulation, which bypasses membrane proximal signaling events [39].

The strongest determinant of ETS1 specificity, Motif 4, fixed the sequence of both ETS1 and RUNX binding sites as well as spacing and relative orientation of the two sites (Figure 3 and Figure 5). This strict conservation was somewhat surprising because ETS1 and RUNX1 bind to DNA cooperatively *in vitro* at a variety of other spacings and orientations [26,43]. Alternate spacing can also function in transcription activation *in vivo*. For example, the MMLV enhancer is activated by ETS1 and RUNX1 at a sequence in which the ETS and RUNX sites are four nucleotides further apart than in Motif 4 [44]. Furthermore, Motif 2 can also support ETS1 and RUNX1 cooperativity *in vitro* [18]. Motif 2 utilizes much more divergent ETS and RUNX sequences set two nucleotides further apart than in Motif 4. (Our analysis in Figure 5 does not identify this spacing because these sequences are too divergent from the Transfac ETS1 and RUNX motifs.) However, only Motif 4, not Motif 2 or the MMLV enhancer motif, was associated with ontologies aligned with T cell–specific functions (Table 3, and data not shown). We speculate that this spacing could have a function in addition to the simple
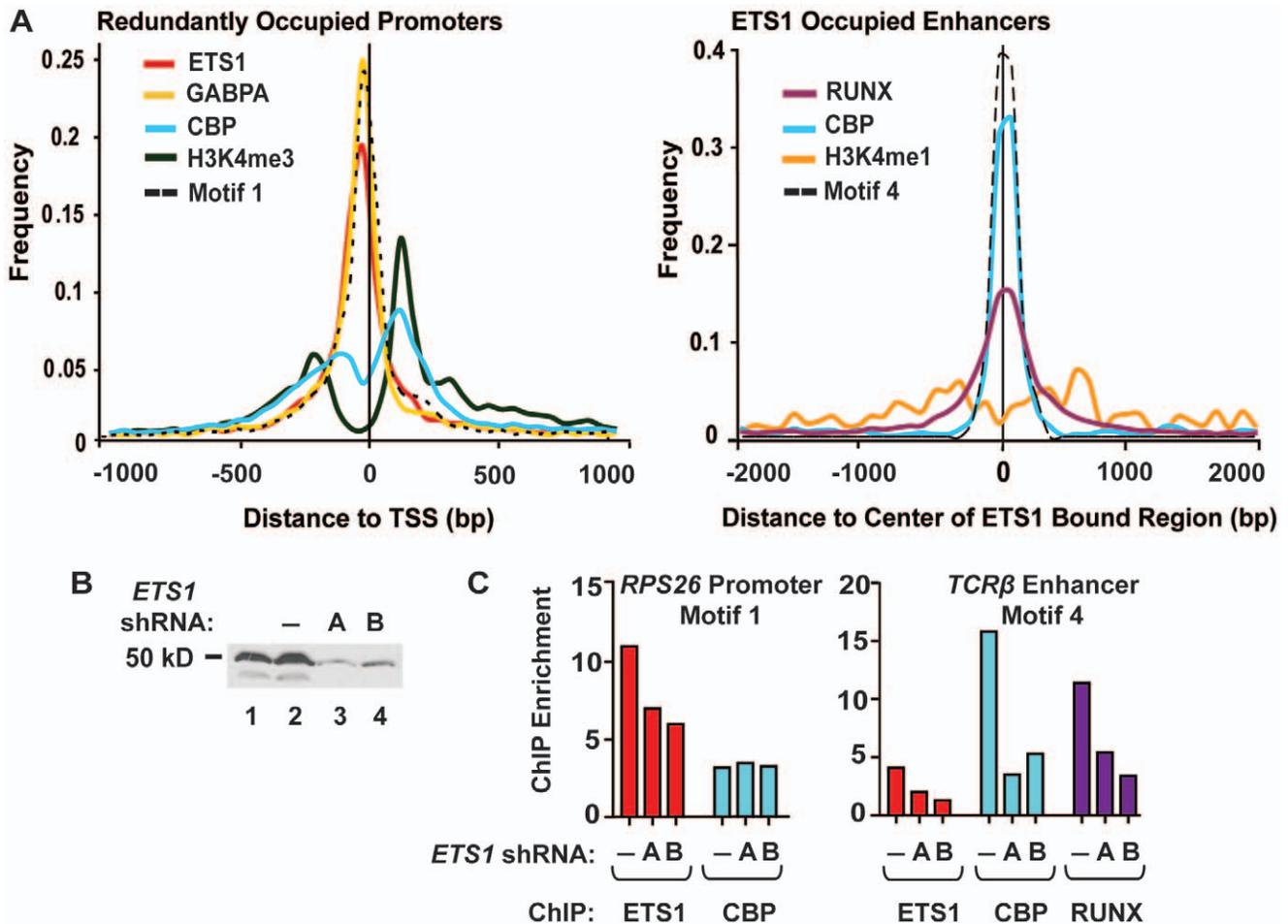
Figure 6. Distinct properties of promoters and enhancers occupied by ETS1. (A) Factor or histone modification positions were plotted as a class average across redundant promoters (left) or ETS1 occupied enhancers (right). At redundant promoters, the occupancy profiles of ETS1, GABPA, CBP, H3K4 tri-methyl, and Motif 1 were plotted from the center of each occupied region to the nearest RefSeq TSS. At ETS1-occupied enhancers, the occupancy profiles of RUNX1, CBP, H3K4 mono-methyl, and Motif 4 were plotted with respect to the center of the ETS1 bound region. For each factor, histone modification, or motif, a histogram of 30 bp bins was generated to represent the frequency of occupancy for each distance. The number of occurrences at each distance was normalized to the total number of regions with an occurrence of each factor, histone modification, or motif. A vertical line indicates the zero position of each chart (TSS or center of ETS1 bound region). (B) Protein immunoblot of Jurkat whole cell extracts with the ETS1 antibody. Lane 1, no shRNA; Lane 2, negative control shRNA targeting luciferase; Lanes 3 and 4, two independent shRNAs (A and B) targeting ETS1. The two bands apparent in Lanes 1 and 2 are consistent with the 51 and 42 kDa splicing isoforms of ETS1. (C) ETS1, CBP, and RUNX ChIP enrichment at the *TCRβ* enhancer and the *RPS26* promoter. The shRNAs were expressed in Jurkat T cells prior to ChIP, as indicated. Two independent biological replicates provided similar patterns, but different maximum levels of enrichment. A representative experiment is shown. Primer sequences used are provided in Table S3.
doi:10.1371/journal.pgen.1000778.g006

recruitment of ETS1 and RUNX. This may reflect a requirement for a specific conformation of ETS1 and RUNX for the transcriptional activation function of these enhancers. For example, because both ETS1 and RUNX can bind the co-activator CBP [15,16,45], and CBP occupies the same position as ETS1 and RUNX at enhancers (Figure 6A), cooperative CBP recruitment may require this particular configuration of ETS1 and RUNX.

A differential pattern of CBP enrichment at enhancers and promoters

This first picture of genome-wide occupancy of a transcription factor in combination with the co-factor CBP presented surprising diversity. In spite of the general picture that CBP occupancy was strongly correlated with ETS1 binding at both tissue specific enhancers and at active promoters in Jurkat T cells (Table 1), fine

mapping uncovered a more complex picture (Figure 6A). The coincident binding observed at enhancers and the sensitivity of CBP occupancy at the *TCRβ* enhancer to an ETS1 knockdown (Figure 6C) is consistent with direct recruitment of CBP by ETS1 and/or RUNX. These data are consistent with reports that CBP/p300 has a strong correlation with enhancers [21,30,46]. This not only supports the functionality of ETS1 bound distal enhancers, but also strongly demonstrates the role of DNA factors in CBP recruitment. In contrast, the lack of concordance of ETS1 and CBP binding events at promoters suggests that CBP is associated with other factors at these sites. Potential CBP recruitment mechanisms include interaction with general transcription factors [47] and enhancer-promoter looping [48,49]. Either of these mechanisms could contribute to the location of CBP at ETS1 bound promoters. We note, however, that the CpG island-containing promoters of housekeeping genes, at which we observe

redundant ETS occupancy, are thought to lack enhancers. Thus, we suggest that CBP is brought to these promoters by enhancer-independent interactions with the transcriptional machinery. One possibility is that ETS1 participates in recruitment, but maintenance at these constitutively active sites relies on cooperation with basal machinery or modified histones.

Binding site sequence variation guides diverse roles for transcription factors

Like many cellular proteins, transcription factors can have multiple roles that vary based on cell type and condition. Transcription factor function can also vary based on the context of other proteins present at each genomic locus. Here we show that the type of genes that are near ETS1 binding events, and the location of the co-activator CBP differ based on the sequence that recruits ETS1 to DNA. Thus, two different roles of ETS1 in T cells – a role at housekeeping promoters, and one at tissue specific enhancers – can be defined by distinct sequence motifs. The sequence variation for different functions of a transcription factor provides an explanation for the lack of a single binding sequence in many genome-wide occupancy studies. Our investigation provides a route to sort genome-wide binding data by the presence of such sequence motifs and other correlative data to define the distinct functions of a transcription factor.

Materials and Methods

ChIP

ChIP from Jurkat T cells was performed as described previously [18]. In brief, 5×10^7 cells were crosslinked with 1% formaldehyde and sheared chromatin extract was prepared. Dynabeads (Invitrogen) coupled to the appropriate secondary antibody were used to immunoprecipitate extracts treated with one of the following antibodies; polyclonal ETS1, sc-355; polyclonal CBP, A-22; (Santa Cruz Biotechnology), or monoclonal RUNX, α3.2.3.1. Crosslinks were reversed by heating and DNA was purified. Input controls were prepared in parallel, but with no immunoprecipitation step. qPCR analysis of ChIP DNA was performed as described previously [18]. In brief, the level of each region was determined by comparison to a standard curve of ChIP input DNA. Enrichments are a ratio of the level of the target region in each sample over the mean of the level of two negative control genomic regions.

Computational methods

The software used for ChIP-seq analysis is open source and available from the Useq project website (http://useq.sourceforge.net/). Human annotation and sequence were obtained from the UCSC Genome Browser (March 2006, NCBI Build 36.1, HG18).

ChIP-seq analysis

ChIP and Input DNA was prepared for sequencing using Illumina's ChIP-seq kit. Each ChIP DNA sample was pooled from three independent replicates. 36 bp reads were generated using Illumina's Genome Analyzer II and standard pipeline software.

The following software from the Useq package [50] was used to identify regions enriched by ChIP compared to input control. ElandParser mapped sequence reads to the genome with an alignment score of >13 ($-10 \log_{10} (0.05)$). FilterPointData was used to remove reads mapping to repeat regions included in the satellite repeat track from the UCSC genome browser (http://genome.ucsc.edu/). The number of non-repeat reads that mapped to the human genome for each sequencing sample was 6,683,411 for ETS1, 9,509,960 for RUNX, 8,525,775 for CBP, and

13,825,035 for input. PeakShiftFinder was used to measure the shift in the peak location between each DNA strand. ScanSeqs used a sliding window of 250 bp to score for enrichment across the genome and adjusted reads from opposite strands by 150 bp (ETS1, RUNX), or 125 bp (CBP), to remove the peak shift bias. EnrichedRegionMaker identified enriched regions. Significance was determined by calculating a binomial P value for each 250 bp window and controlled for multiple testing by calculating an empirical false discovery rate. The "Best Window" in each enriched region with an empirical false discovery rate of <0.01 were called as "bound regions" and had a median size of 250 bp.

Bound regions were overlapped using the IntersectRegions tool from Useq with no gap between regions except for overlaps reported in Table 1 in which a gap up to 300 bp was allowed. Enriched regions for ETS1, GABPA, and RUNX were screened for intersection with DNase I sensitive regions [19] before further analysis. This screening reduced the number of ETS1 bound regions from 19,420 to 14,824, the number of GABPA bound regions from 9214 to 7724, the number of CBP bound regions from 23,757 to 14,374, and the number of RUNX bound regions from 3632 to 1075. The P value for the overlap between ETS1 and DNase I sensitive sites was determined using IntersectRegions and comparing the ETS1 overlap to the overlap of 1000 iterations of a random regions of equivalent size derived from input point data. The P values for overlaps shown in Table 1 were derived by Fisher's exact test. The nearest Refseq TSS was determined using the FindNeighboringGenes tool. All ETS1 bound regions that intersect with DNase I sensitive regions are provided in Table S1. Table S1 also annotates the nearest gene, the presence or absence of Motif 1, 2 or 4 (Figure 3) and overlapping GABPA, RUNX, and CBP bound regions.

The ChIP-seq datasets and peak files are available for download from NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), accession number GSE17954.

Protein purification and DNA binding assays

Protein purification and DNA binding assays were performed as described previously [18]. In brief, full-length human cDNAs of *ETS1* (p51) and *ELF1* were cloned into pet28a (Novagen) with an N-terminal 6x HIS tag, and expressed in bacteria. Proteins were isolated from inclusion bodies, resuspended in urea buffer (10 mM Tris (pH 7.9), 4 M urea, 500 mM NaCl, 15 mM immidazole), and bound to Ni-sepharose beads. After washing with urea buffer, protein was eluted with urea buffer with 750 mM immidazole and dialyzed overnight into 10 mM Tris (pH 7.9), 0.5 mM EDTA, 50 mM KCl, 1 mM DTT, and 10% glycerol. Protein concentration was determined by comparison to BSA standards on Coomassie brilliant blue stained SDS-PAGE gels. Three-fold serial dilutions of each protein were incubated with ^{32}P-labeled double stranded oligonucleotides (DNA concentration 1×10^{-11}) for 30 min on ice and then run on a 6% polyacrylamide gel. Oligonucleotide probes had the following sequences (5′–3′): GGCCAAGCCGGAAGTGTGTGG-TAAACACTTT, GGCCAAGCCGGATGTGTGTGGTAAACA-CTTT,

GGCCAAGCAGGAAGTGTGTGGTAAACACTTT,

GGCCAAGCAGGATGTGTGTGGTAAACACTTT. K_Ds were calculated by measuring the radioactivity in unbound bands by Phosphorimager (Molecular Dynamics) and using least squares curve fit analysis with fraction of DNA bound $= 1/(1 + K_D/[Protein])$.

Bioinformatic analysis by MEME, PATSER, and GOMINER

MEME (http://meme.sdsc.edu/meme4_1_1/cgi-bin/meme.cgi) was run with default settings except the minimum motif

length was set to 8 and the maximum motif length was set to 15. For each set of regions, the 250 with the highest log-transformed binomial *P* value for ETS1 were analyzed. Only the motif with the lowest E-value was reported.

The position weight matrixes for each motif used in the PATSER program (part of Regulatory Sequence Analysis Tools: http://rsat.ulb.ac.be/rsat/) to identify matches in each ETS1 bound region are listed in Table S2. PATSER score cutoffs used were Motif 1:9; Motif 2:12; Motif 4:10.3. Matches are listed in Table S1.

GoMiner (http://discover.nci.nih.gov/gominer/) was used to identify over-represented gene ontologies. All Refseq genes were used as the "total" and each subset of Refseq genes was used as the "change" file. Default settings were used except "Evidence level 4" and "All/gene ontology" were selected.

ETS1 knockdown

Two distinct small-hairpin RNAs (shRNAs) targeting ETS1, or a negative control shRNA targeting luciferase were cloned into pMK0.1p [51] and introduced to Jurkat T cells by MMLV based retroviruses. Jurkat T cells expressing the shRNA were selected by puromycin resistance. The sequences targeted were: luciferase, CTTACGCTGAGTACTTCGA; ETS1 A, AGGTGTAGACTTCCAGAAG; ETS1 B, CTGATGTAAGGCAATTAAT.

Supporting Information

Figure S1 ETS1 bound regions have a high density within 500 bp of a TSS. The distance from the center of each of the 19,420 ETS1 bound regions to the nearest Refseq TSS was recorded. Distances were binned in 50 bp bins and frequencies plotted for bins from 0 to 5,000. Distances greater than 5,000 were discarded. 7,137 ETS1 bound regions had centers within 500 bp of a TSS.
Found at: doi:10.1371/journal.pgen.1000778.s001 (2.64 MB TIF)

Figure S2 Validation of ChIP-seq results by Q-PCR analysis of ChIP DNA. Enrichment was tested using primer sets specific for each region (Table S1). The level of each region was determined by comparison to a standard curve of ChIP input DNA. Enrichments are a ratio of the level of the target region in each sample over the mean of the level of two negative control genomic regions. The enrichments shown are the mean and standard error of the mean of two independent ChIP experiments. Regions were

considered bound that had an mean enrichment of equal to or greater than 2. (A) 15 randomly selected ETS1 bound regions that overlapped with a DNase I sensitive site in CD4+ T cells (1–15), 8 randomly selected ETS1 bound regions that did not overlap with a DNase I sensitive site (16–23), and 9 randomly selected DNase I sensitive regions that were not scored as ETS1 bound (24–32) were tested for ETS1 ChIP enrichment in Jurkat T cells. Significantly more (13 of 15) ETS1 bound/DNase I sensitive regions were verified compared the other two categories to (0 of 8 and 0 of 9; P<0.0001, Fisher's exact test). (B) 10 of 10 randomly selected CBP bound/DNase I sensitive regions were verified for CBP enrichment. (C) 8 of 8 randomly selected RUNX bound/DNase I sensitive regions were verified for RUNX enrichment.
Found at: doi:10.1371/journal.pgen.1000778.s002 (5.85 MB TIF)

Table S1 Properties of ETS1 bound regions. All chromosomal regions identified as ETS1 bound that also overlapped with DNase I sensitivity are listed. Regions that also overlapped with GABPA, RUNX, or CBP are indicated. Regions that contained Motif 1, 2, or 4 are indicated. The nearest RefSeq mRNA and gene, and the distance to the TSS is shown. The binomial P value and empirical false discovery rate reflect the significance of the ETS1 bound region.
Found at: doi:10.1371/journal.pgen.1000778.s003 (3.54 MB XLS)

Table S2 Position weight matrixes for Motif 1, 2, and 4.
Found at: doi:10.1371/journal.pgen.1000778.s004 (0.12 MB DOC)

Table S3 Oligonucleotide primers used in this study.
Found at: doi:10.1371/journal.pgen.1000778.s005 (0.07 MB DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: PCH KJC BJG. Performed the experiments: PCH KJC. Analyzed the data: PCH KJC RLP WEJ BJG. Contributed reagents/materials/analysis tools: NAS. Wrote the paper: PCH KJC NAS BJG.

References

1. Massie CE, Mills IG (2008) ChIPping away at gene regulation. EMBO Rep 9: 337–343.
2. Ji H, Vokes SA, Wong WH (2006) A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. Nucleic Acids Res 34: e146.
3. Rabinovich A, Jin VX, Rabinovich R, Xu X, Farnham PJ (2008) E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. Genome Res 18: 1763–1777.
4. Gaudet J, Mango SE (2002) Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science 295: 821–825.
5. Lam FH, Steger DJ, O'Shea EK (2008) Chromatin decouples promoter threshold from dynamic range. Nature 453: 246–250.
6. Leung TH, Hoffmann A, Baltimore D (2004) One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. Cell 118: 453–464.
7. Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, et al. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. Science 324: 407–410.
8. Galang CK, Muller WJ, Foos G, Oshima RG, Hauser CA (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.
9. Hollenhorst PC, Jones DA, Graves BJ (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic Acids Res 32: 5693–5702.
10. Sharrocks AD (2001) The ETS-domain transcription factor family. Nat Rev Mol Cell Biol 2: 827–837.
11. Oikawa T, Yamada T (2003) Molecular biology of the Ets family of transcription factors. Gene 303: 11–34.
12. De Val S, Chi NC, Meadows SM, Minovitsky S, Anderson JP, et al. (2008) Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. Cell 135: 1053–1064.
13. Chen C, Ouyang W, Grigura V, Zhou Q, Carnes K, et al. (2005) ERM is required for transcriptional control of the spermatogonial stem cell niche. Nature 436: 1030–1034.
14. Muthusamy N, Barton K, Leiden JM (1995) Defective activation and survival of T cells lacking the Ets-1 transcription factor. Nature 377: 639–642.
15. Yang C, Shapiro LH, Rivera M, Kumar A, Brindle PK (1998) A role for CREB binding protein and p300 transcriptional coactivators in Ets-1 transactivation functions. Mol Cell Biol 18: 2218–2229.
16. Jayaraman G, Srinivas R, Duggan C, Ferreira E, Swaminathan S, et al. (1999) p300/cAMP-responsive element-binding protein interactions with ets-1 and ets-2 in the transcriptional activation of the human stromelysin promoter. J Biol Chem 274: 17342–17352.
17. Foulds CE, Nelson ML, Blaszczak A, Graves BJ (2004) MAPK phosphorylation activates Ets-1 and Ets-2 by CBP/p300 recruitment. Mol Cell Biol 24: 10954–10964.
18. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev 21: 1882–1894.
19. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132: 311–322.

20. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823–837.

21. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311–318.

22. Robertson AG, Bilenky M, Tam A, Zhao Y, Zeng T, et al. (2008) Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. Genome Res 18: 1906–1917.

23. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, et al. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 5: 829–834.

24. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intellegent Systems for Molecular Biology. Menlo Park, CA: AAAI Press. pp 28–36.

25. Wang CY, Bassuk AG, Boise LH, Thompson CB, Bravo R, et al. (1994) Activation of the granulocyte-macrophage colony-stimulating factor promoter in T cells requires cooperative binding of Elf-1 and AP-1 transcription factors. Mol Cell Biol 14: 1153–1159.

26. Goetz TL, Gu TL, Speck NA, Graves BJ (2000) Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. Mol Cell Biol 20: 81–90.

27. Giese K, Kingsley C, Kirshner JR, Grosschedl R (1995) Assembly and function of a TCRα enhancer complex is dependent on LEF-1 induced DNA bending and multiple protein-protein interactions. Genes Dev 9: 995–1008.

28. Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics 15: 563–577.

29. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, et al. (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 4: R28.

30. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457: 854–858.

31. Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10: 161–172.

32. Bosselut R, Levin J, Adjadj E, Ghysdael J (1993) A single amino-acid substitution in the ETS domain alters core DNA binding specificity of Ets-1 to that of the related transcription factors Elf-1 and E74. Nucleic Acids Res 21: 5184–5191.

33. Yokomori N, Kobayashi R, Moore R, Sueyoshi T, Negishi M (1995) A DNA methylation site in the male-specific P450 (Cyp 2d-9) promoter and binding of the heteromeric transcription factor GABP. Mol Cell Biol 15: 5355–5362.

34. Gaston K, Fried M (1995) CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. Nucleic Acids Res 23: 901–909.

35. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6–21.

36. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.

37. Rozenberg JM, Shlyakhtenko A, Glass K, Rishi V, Myakishev MV, et al. (2008) All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. BMC Genomics 9: 67.

38. Bories J-C, Willerford DM, Grevin D, Davidson L, Camus A, et al. (1995) Increased T cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene. Nature 377: 635–638.

39. Barton K, Muthusamy N, Fischer C, Ting C-N, Walunas T, et al. (1998) The Ets-1 transcription factor is required for the development of natural killer cells in mice. Immunity 9: 555–563.

40. Guo Y, Maillard I, Chakraborti S, Rothenberg EV, Speck NA (2008) Core binding factors are necessary for natural killer cell development and cooperate with Notch signaling during T cell–specification. Blood 112: 480–492.

41. Egawa T, Eberl G, Taniuchi I, Benlagha K, Geissmann F, et al. (2005) Genetic evidence supporting selection of the Valpha14i NKT cell lineage from double-positive thymocyte precursors. Immunity 22: 705–716.

42. Ohno S, Sato T, Kohu K, Takeda K, Okumura K, et al. (2008) Runx proteins are involved in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation. Int Immunol 20: 71–79.

43. Wotton D, Ghysdael J, Wang S, Speck NA, Owen MJ (1994) Cooperative binding of Ets-1 and Core Binding Factor to DNA. Mol Cell Biol 14: 840–850.

44. Sun W, Graves BJ, Speck NA (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.

45. Kitabayashi I, Yokoyama A, Shimizu K, Ohki M (1998) Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. EMBO J 17: 2994–3004.

46. Wang Z, Zang C, Cui K, Schones DE, Barski A, et al. (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell 138: 1019–1031.

47. Black JC, Choi JE, Lombardo SR, Carey M (2006) A mechanism for coordinating chromatin modification and preinitiation complex assembly. Mol Cell 23: 809–818.

48. Resendes KK, Rosmarin AG (2006) GA-binding protein and p300 are essential components of a retinoic acid-induced enhanceosome in myeloid cells. Mol Cell Biol 26: 3060–3070.

49. Teferedegne B, Green MR, Guo Z, Boss JM (2006) Mechanism of action of a distal NF-kappaB-dependent enhancer. Mol Cell Biol 26: 5759–5770.

50. Nix DA, Courdy SJ, Boucher KM (2008) Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. BMC Bioinformatics 9: 523.

51. Masutomi K, Yu EY, Khurts S, Ben-Porath I, Currier JL, et al. (2003) Telomerase maintains telomere structure in normal human cells. Cell 114: 241–253.

52. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, et al. (2002) Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A 99: 4465–4470.