

Cell, Volume 132

Supplemental Data

High-Resolution Mapping

and Characterization of Open Chromatin

across the Genome

Alan P. Boyle, Sean Davis, Hennady P. Shulha, Paul Meltzer, Elliott H. Margulies, Zhiping Weng, Terrence S. Furey, and Gregory E. Crawford

Supplemental Figures

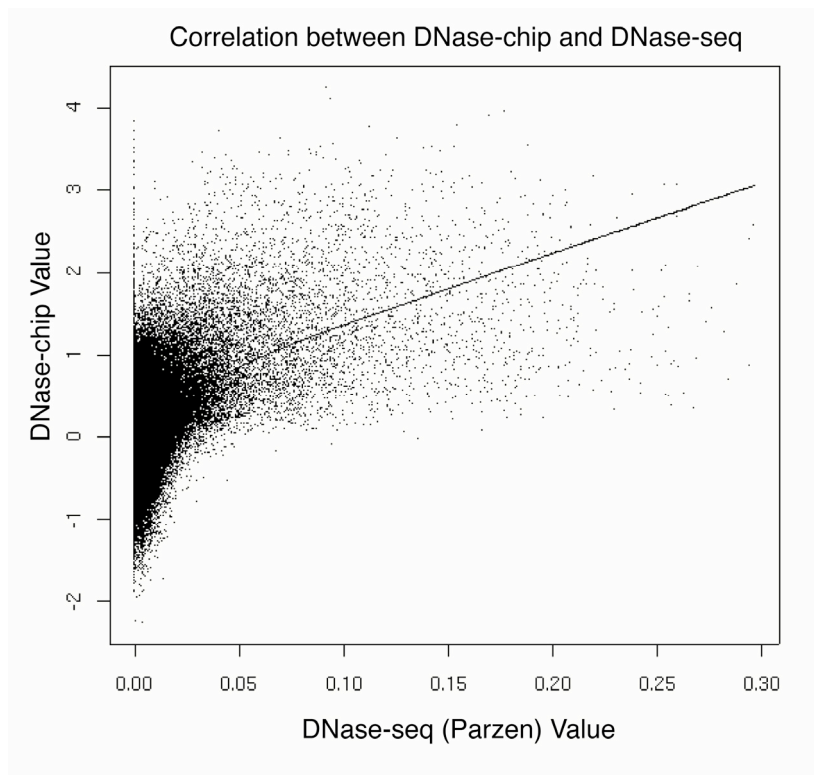


Figure S1: DNase-seq and DNase-chip data is strongly correlated (Pearson's $R=0.45$).

To calculate this correlation, we compared ACME p-values calculated for every chip probe to the maximum DNase-seq score within the tiled region. Probes are only 50bps

and are spaced approximately every 100bps. Therefore, this correlation is even more impressive given the lower resolution of microarray chips. The points in this figure with high DNase-chip values and no DNase-seq value ($x=0$) represent tiles contained within recent segmental duplications where sequences could not be uniquely aligned and were filtered out.

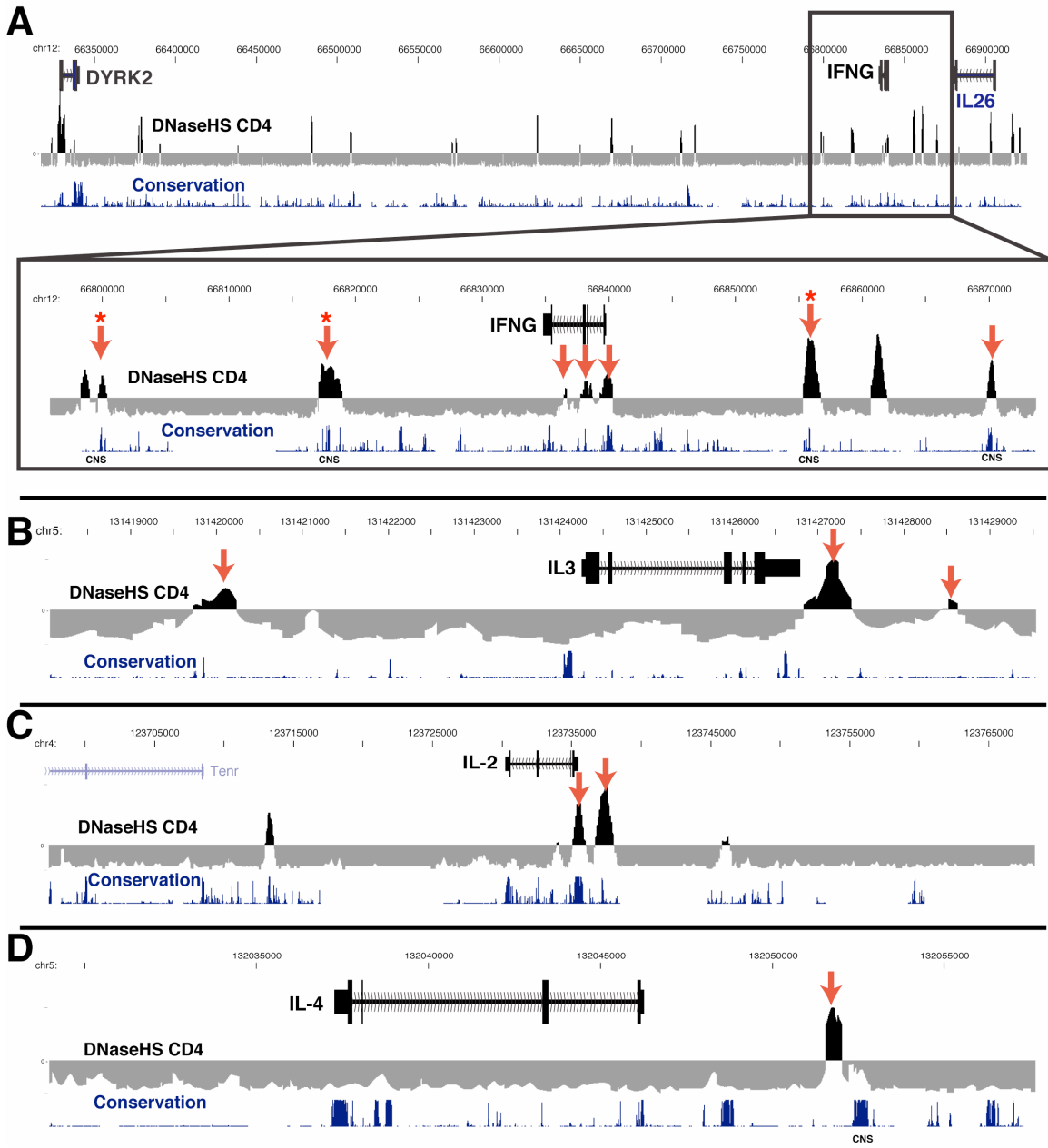


Figure S2: Red arrows indicate DNaseI hypersensitive (HS) sites previously identified using Southern blots while those without arrows indicate novel sites. (A) Zoomed out and close up view surrounding IFN γ gene. 9 sites that we find in the proximity of the IFN γ gene include 2 novel sites, 4 sites previously described in human T cells (Lee et al., 2004), and the remainder mapped in mouse (Schoenborn et al., 2007). Arrows with an asterisk represent HS sites that were identified in TH1 or TH2 mouse cells, but not naïve CD4 $^{+}$ T cells. (B) HS sites previously identified near Il-3 in the Jurkat cell line (Cockerill et al., 1993). Additional sites were found in Jurkat cells, but not found in our cell type (data not shown) (C) HS sites previously described upstream of Il-2 in resting CD4 $^{+}$ T cells (Siebenlist et al., 1986). (D) Il-4 downstream HS sites found in naïve CD4 $^{+}$ T cells (Agarwal and Rao, 1998).

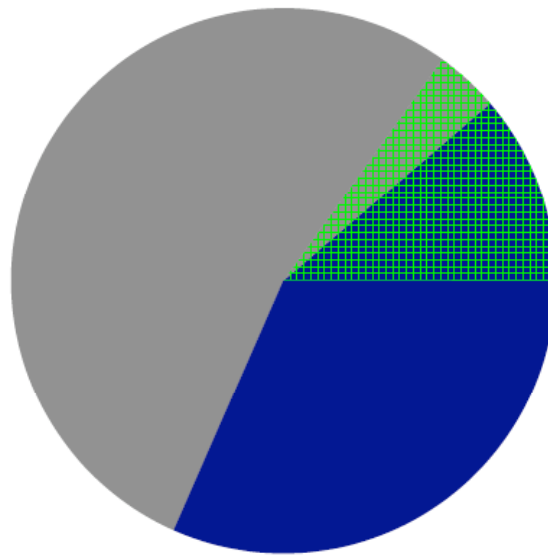
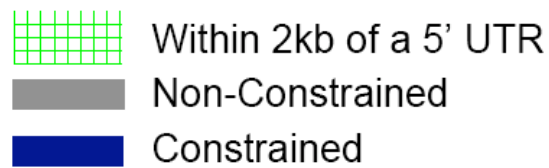


Figure S3: Pie graph illustrating the relationship between evolutionarily constrained regions and DNaseI hypersensitive (HS) sites. All represented overlaps are statistically significant ($P < 0.01$) using the GSC method as previously described (ENCODE, 2007; Margulies et al., 2003). Overall, 43% of the HS sites overlap at least one base of a constrained region (blue slice). When focusing specifically on those HS sites that lie within 2kb of a 5'UTR (15%; green hatch), 75% overlap constrained regions (blue slice overlapping green hatch).

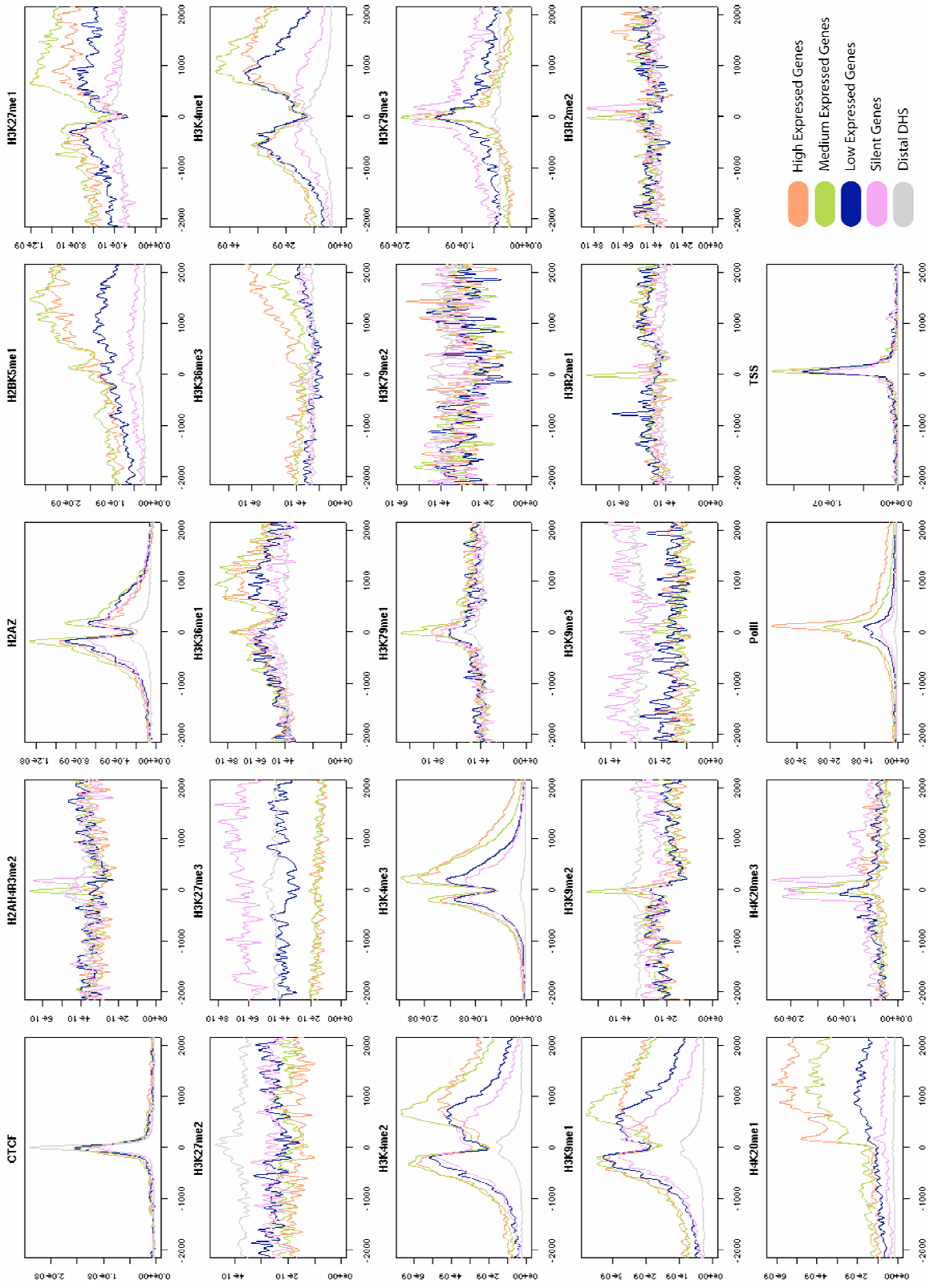


Figure S4: Distance and number of histone methylation, histone variant H2A.Z, CTCF, and RNA Polymerase II ChIP-seq sequences centered on peaks of different classes of DNaseI hypersensitive (HS) sites. Proximal HS sites less than 2 kb from a transcription start site (TSS) were subdivided into classes based on expression levels of associated genes as measured by an Affymetrix expression array. The distal class of HS sites are those that are greater than 2 kb from a TSS. The “high expression” class consists of the 570 genes with an expression level > 10 , the “medium expression” class consists of 791 expressed genes with an expression level between 8.2 and 8.9, the “low expression” class consists of 727 genes with an expression level between 5.5 and 5.8, and the “silent” class consists of 698 genes with an expression level < 4.5 . These levels were determined through analysis of the distribution of expression levels and attempts to keep approximately the same count in each bin. The distal class of HS sites consists of 6814 distal sites that are > 2 kb from a TSS and are not in a transcribed region. Values on y-axis are normalized sequence counts and the x-axis is the distance from the highest value of the HS site.

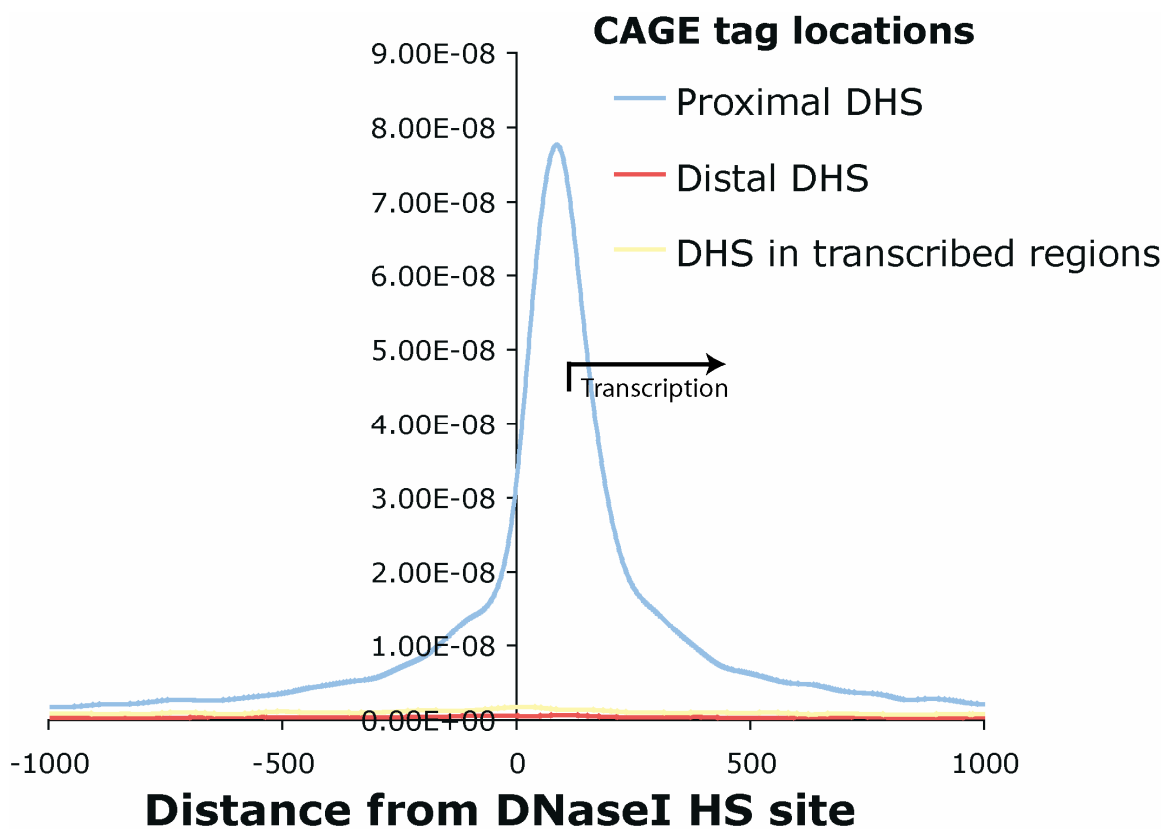


Figure S5: Counts of unclustered CAGE tags (Kawaji et al., 2006) are plotted based on their distance from the highest scoring portion of a DNaseI hypersensitive (DHS) site. Proximal DHS sites are defined as those within 2kb of an annotated transcription start site or RNA PolII signal. DHS in transcribed regions are non-proximal sites that overlap annotated transcribed regions as defined by all mRNAs and ESTs. Distal HS sites are those not in either of the previous two classes. Plotted on the y-axis is the normalized CAGE tag count and the x-axis is distance from the highest scoring point of the DHS.

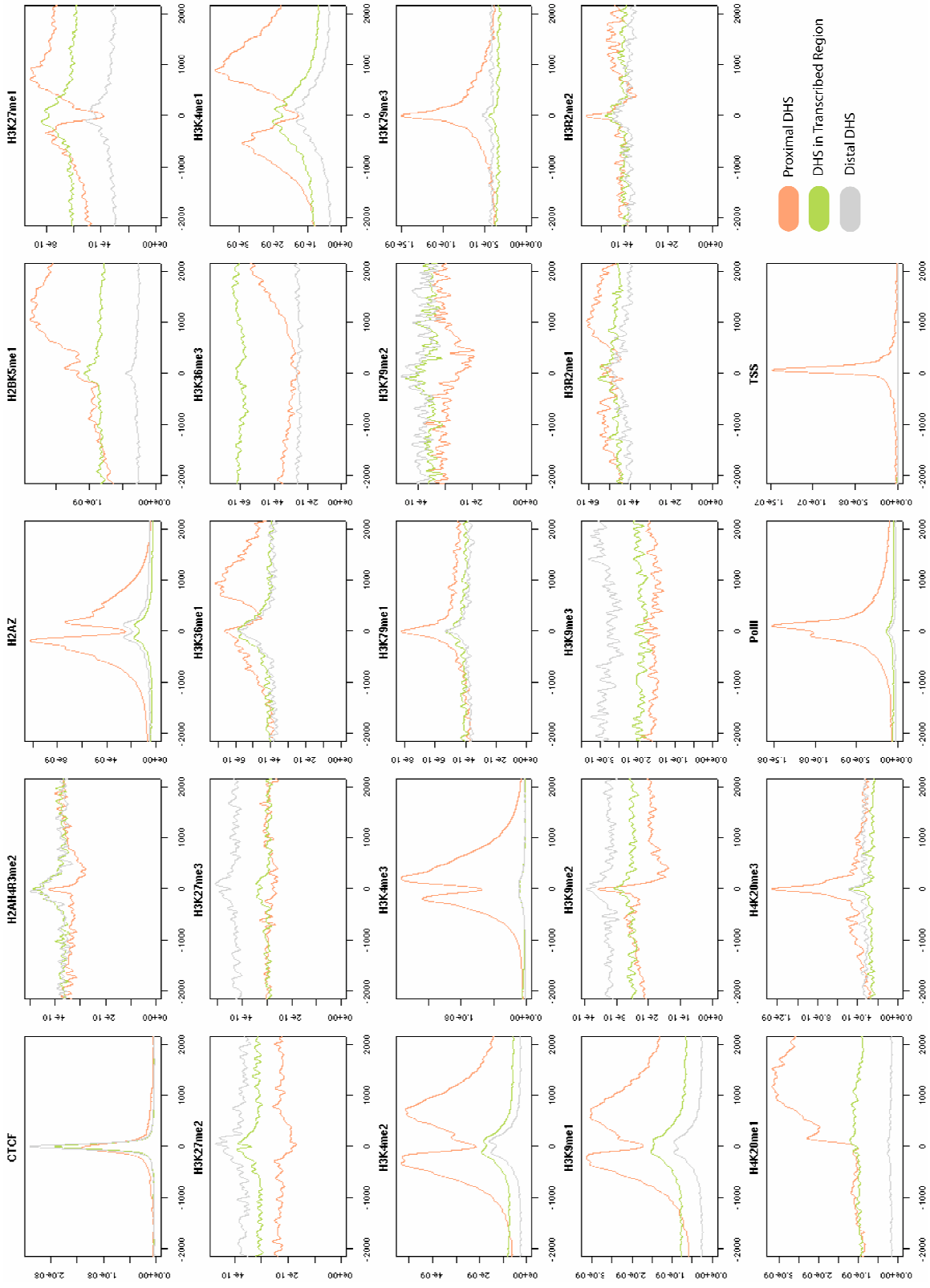


Figure S6: Distance and number of histone methylation, histone variant H2A.Z, CTCF, and RNA Polymerase II ChIP-seq sequences centered on peaks of different classes of DNaseI hypersensitive (HS) sites. These classes are based on the relationship between HS sites and genes. Proximal HS sites are within 2kb of a transcription start site (TSS) or a strong RNA PolII signal. DHS in transcribed regions are DHS that overlap an annotated mRNA or EST but are not in the proximal set. Distal HS sites are those not in either of the previous two classes. Plotted on the y-axis is the normalized sequence count and the x-axis is distance from the highest scoring point of the HS site.

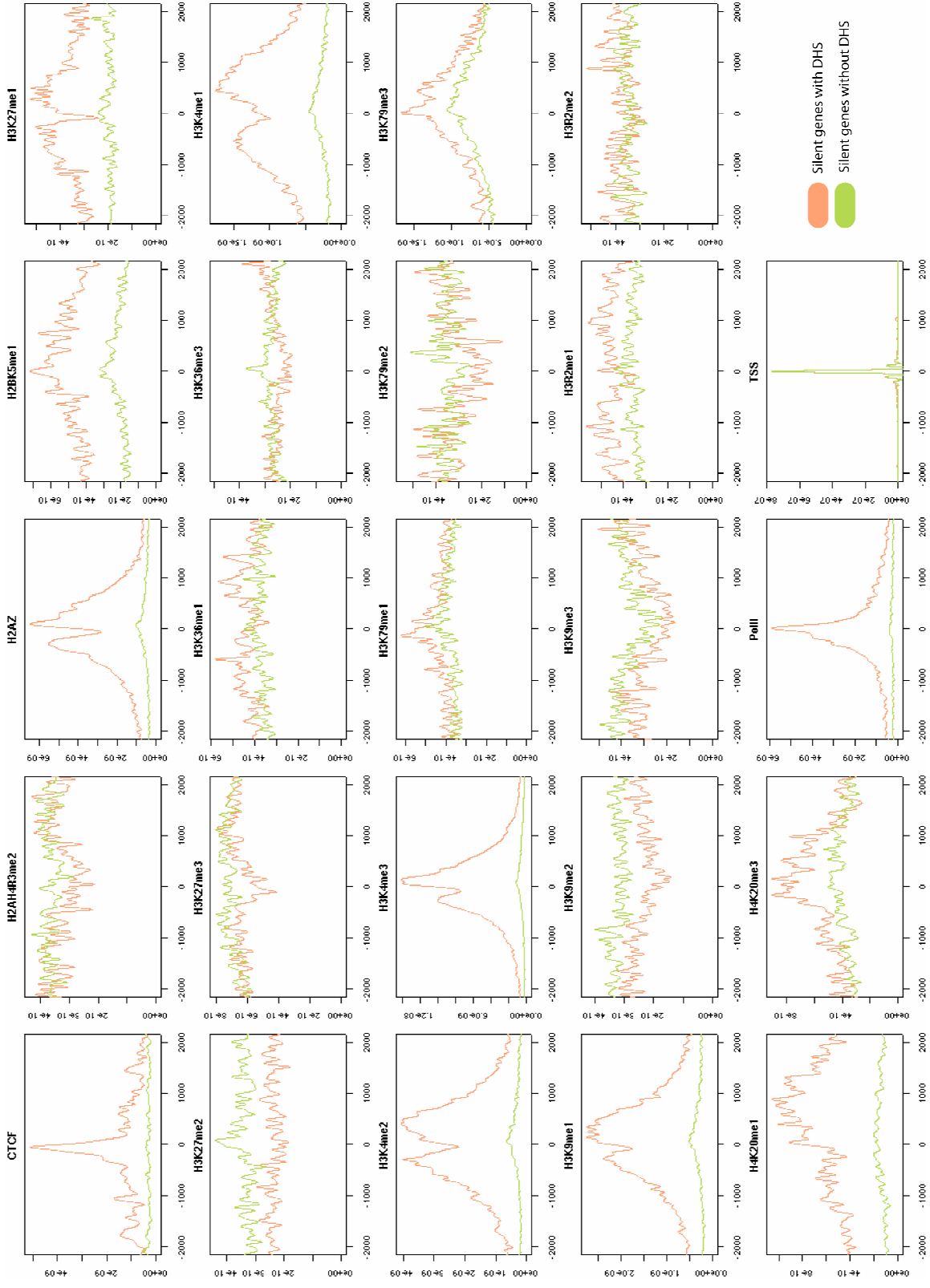


Figure S7: Distance and number of histone methylation, histone variant H2A.Z, CTCF, and RNA Polymerase II ChIP-seq sequences centered on the transcription start site (TSS) of transcriptionally silent genes that either have or do not have a DNaseI hypersensitive (HS) sites within 2Kb of their TSS. Silent genes are defined as those with expression < 4.5 based on Affymetrix expression array experiments. There are 903 silent genes with a DNaseI HS site and 1583 silent genes without a DNaseI HS site. The distribution of expression scores in each of the two sets is virtually identical. Plotted on the y-axis is the normalized sequence count and the x-axis is distance from the annotated TSS.

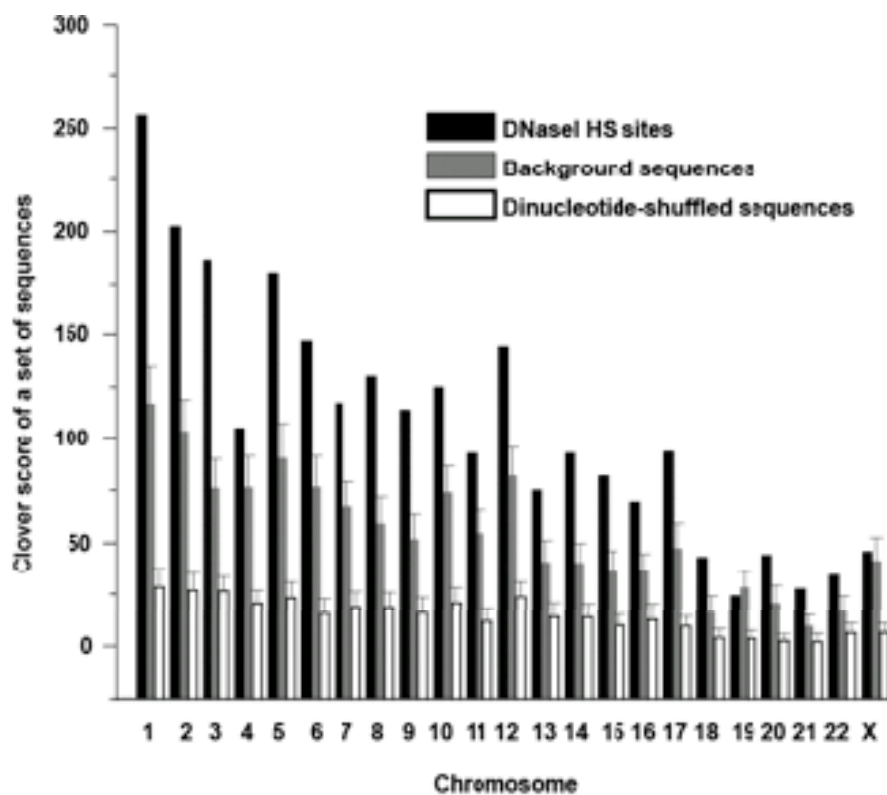


Figure S8: Clover scores of the AML motif within distal DNaseI HS sites located on each chromosome. Background model Clover scores were determined from equal-sized, randomly chosen background sequences and dinucleotide-shuffled DNaseI HS sites. Error bars indicate the standard deviation of 1000 sets of background sequences or shuffled sequences in each chromosome.

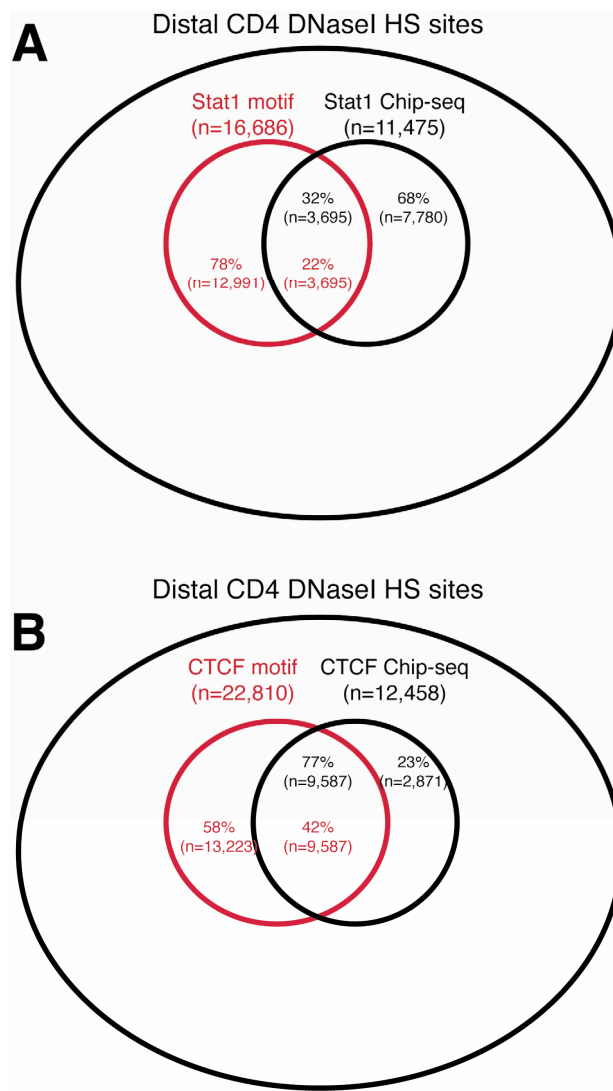


Figure S9: The overlaps of DNaseI hypersensitive (HS) site with motifs and ChIP-seq data. (A) Of all Stat1 ChIP-seq signals identified in HeLa cells (Robertson et al., 2007) that overlap distal HS sites, 32% overlap a Stat motif. (B) Of all CTCF ChIP-seq signals identified in CD4⁺ T cells (Barski et al., 2007) that overlap distal HS sites, 77% overlap a CTCF motif. Both are highly significant overlaps ($p < 0.0001$) based on 10,000 iterations of randomly assign motifs to distal HS sites and calculating the number of ChIP-seq overlaps.

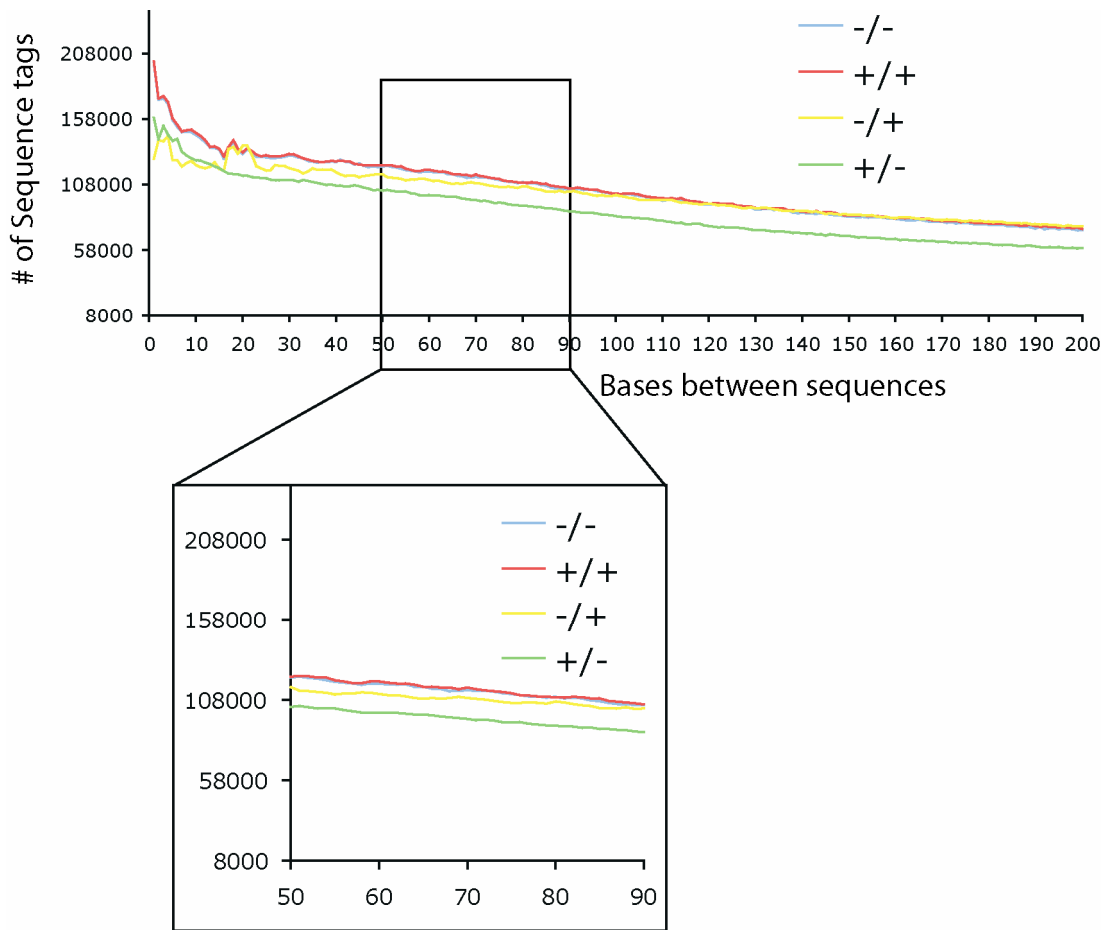


Figure S10: Oscillation patterns shown in Figure 5B are not detected within DNaseI hypersensitive (HS) sites as a whole. Distances between DNase-sequence tags were calculated only for sequences that map within DNaseI HS sites

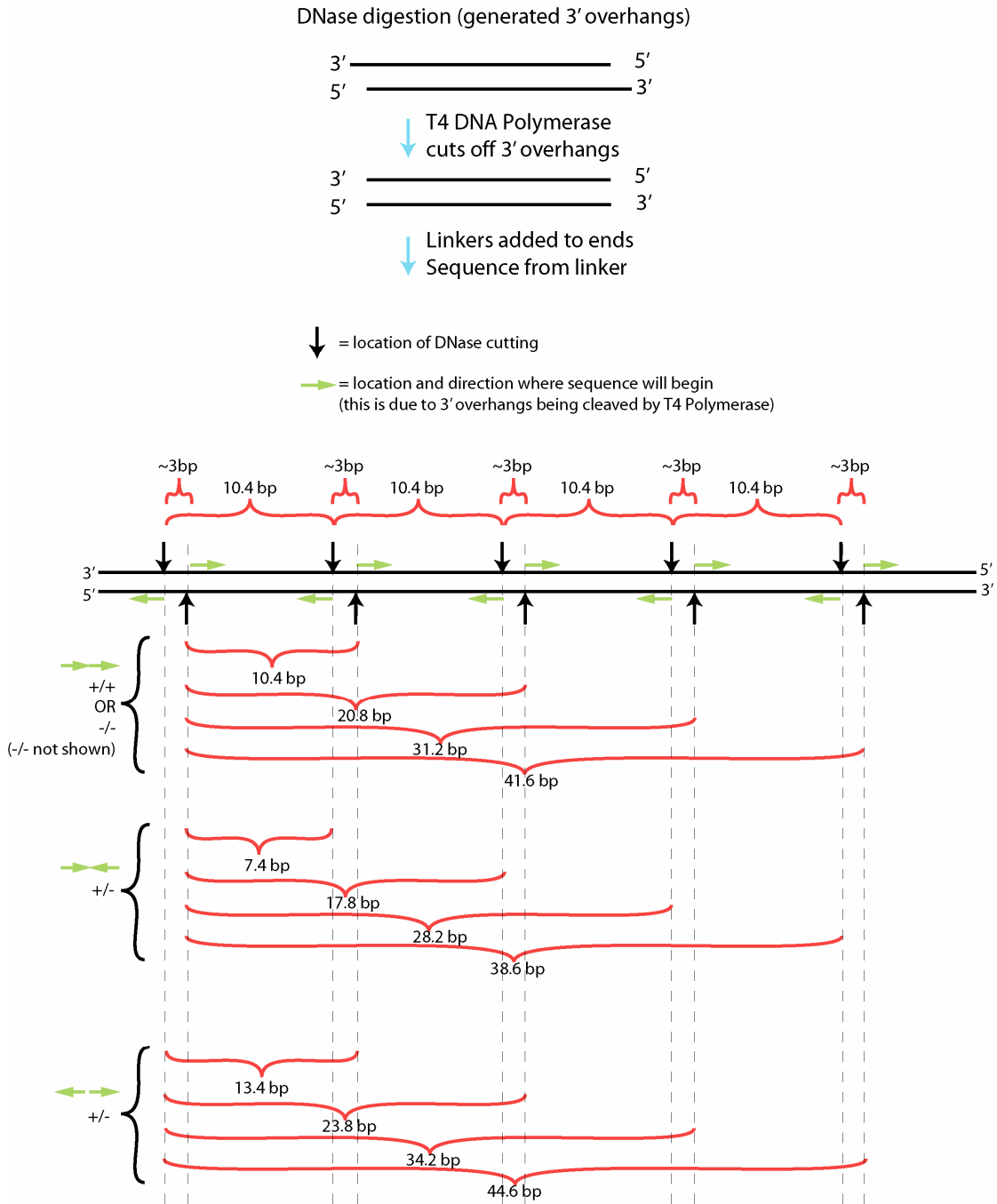


Figure S11: DNase I in the presence of Mg⁺ nicks each strand of DNA, leaving an average 3 basepair 3' overhang. T4 DNA Polymerase is used to trim off the overhang, which allows us to detect the overhang length (Figure 5C). This corresponds the minor groove that is exposed every 10.4 bases as it wraps around the nucleosome.

	All DNase data				Top 20%				Bottom 20%			
	Count	% of Sites	Avg Size (bp)	Avg Score	Count	% of Sites	Avg Size (bp)	Avg Score	Count	% of Sites	Avg Size (bp)	Avg Score
1st Exon	3157	3.3%	1162	2.15	1704	9.0%	1757	3.44	342	1.8%	80	0.23
2kb Upstream	12476	13.1%	1211	2.48	7320	38.6%	1810	3.83	1372	7.2%	84	0.23
1st Intron	15706	16.5%	532	0.80	1878	9.9%	1513	2.51	3568	18.8%	110	0.23
2kb Downstream	2586	2.7%	548	0.92	463	2.4%	1266	2.48	526	2.8%	111	0.23
Other Exons	2386	2.5%	595	0.89	358	1.9%	1387	2.55	475	2.5%	117	0.23
Other Introns	21205	22.3%	494	0.74	2276	12.0%	1268	2.25	4726	24.9%	125	0.23
Intergenic	37409	39.4%	521	0.83	4960	26.2%	1330	2.51	7950	41.9%	125	0.23
Total	94925	100.0%	632	1.07	18959	100.0%	1564	3.07	18959	100.0%	118	0.23

Table S1: The scores, size, and count of DNaseI HS sites calculated relative to the gene annotations track. Shown is the location of all DNaseI HS sites, the strongest scoring DNaseI HS sites (top 20%), and the weakest scoring DNaseI HS sites (bottom 20%).

Motif	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	
AML1																								
AML-1a.M00271	0/0	0/0	0/0	/	0/0	0/0.004	0/0.001	0/0	0/0.004	0/0.008	/	0/0.003	/	0/0	0/0.001	0/0.006	0/0.005	0/0.001	/	/	/	/	/	
AML.M00769	0/0	0/0	0/0.001	/	0/0.009	0/0	/	0/0.001	0/0.004	/	/	/	/	0/0.002	0/0	0/0	0/0	/	/	/	0/0.003	/	/	
AML1.M00751	0/0	0/0	0/0	/	0/0	0/0	0/0	0/0	0/0	0/0	0/0.003	0/0	0/0.005	0/0	0/0	0/0.004	0/0	0/0.001	/	0/0.002	0/0.004	/	/	
AP-2rep.M00468	0/0	0/0.009	0/0.001	/	/	0/0	0/0.001	0/0.007	0/0	/	/	0/0.001	/	0/0	0/0.003	0/0.005	/	0/0	/	/	/	0/0	/	
core-binding factor.M00722	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	/	0/0	0/0	0/0	0/0	
Osf2.M00731	0/0	0/0	0/0	0/0	0/0	0/0	0/0.001	0/0	0/0	0/0.007	/	0/0	0/0.002	0/0	0/0	0/0.007	0/0	0/0.006	/	/	0/0.007	/	/	
C/EBP																								
C/EBP.M00159	0/0.001	0/0	0/0	0/0.003	0/0	0/0	0/0	0/0	0/0.001	0/0.006	0/0.003	0/0	0/0	0/0.003	/	/	/	0/0.002	/	/	/	/	0/0	
C/EBP.M00770	/	0/0	0/0	0.002/0	0/0	0/0	0/0	0/0	0/0	0/0.001	0/0	0/0	0/0	/	/	/	/	0/0.003	/	0/0.007	/	/	0/0	
ETS																								
ETS.M00771	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Ets.M00971	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.002	0/0	0/0	0/0	0/0	0/0
c-Ets-1 68.M00743	0/0	0/0	0/0	/	0/0	0/0	0/0	0/0.002	0/0	0/0	0/0	0/0.002	0/0.001	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	/	0/0	0/0.001
c-Ets-1.M00339	0/0	0/0	0/0	0/0.004	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
c-Ets-2.M00340	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	/	0/0	0/0	0/0	0/0	0/0
ELF-1.M00746	0/0	0/0	0/0	0/0.001	/	0/0	0/0	0/0.003	0/0.006	0/0	0/0	0/0	0/0.007	0/0	0/0	0/0.001	0/0	/	0/0.001	0/0	/	/	0/0.009	
EIk-1.M00007	0/0	0/0.001	0/0.001	0/0	/	0/0.006	0/0	0/0.001	0/0.001	0/0	0/0	0/0	0/0.001	0/0	0/0.003	0/0	0/0	/	0/0	/	/	0/0.002	0/0	
PU.1.M00658	0/0	0/0	0/0	/	0/0	0/0	0/0	0/0	0/0.006	0/0	0/0	0/0	/	0/0	0/0	0/0	0/0	/	0/0	0/0	/	0/0	0/0	
Tel-2.M00678	0/0	0/0	0/0	0/0.002	0/0.007	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	/	0/0	0/0.001	0/0	0/0	0/0.001	
NRF-2.M00108	0/0	/	0/0.004	/	/	0/0	/	/	/	0/0.003	/	0/0	/	/	0/0	0/0.001	0/0	/	0/0	0/0	0/0.003	0/0	/	
PEA3.M00655	0/0	0/0	0/0	/	/	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.005	0/0	0/0	0/0	0/0	0/0.002	0/0	0/0	0/0.001	0/0	0/0	
STAT																								
STAT.M00777	0/0	/	/	/	/	0/0.005	/	/	/	0/0.004	0/0	0/0	/	0/0	0/0	0/0	0/0.001	/	0/0.001	0/0.006	/	/	0/0.006	
STAT1.M00496	0/0	0/0	/	0/0	0/0.003	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	/	0/0	0/0.001	0/0	/	0/0	0/0	/	/	0/0	
STAT3.M00497	0/0	/	/	/	/	0/0	/	0/0.001	/	0/0	0/0	0/0	/	0/0.001	0/0.001	0/0	0/0.006	/	0/0	0/0	/	/	0/0	
STAT4.M00498	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.001	0/0	0/0	0/0	/	0/0.001	0/0	0/0	/	0/0	
STAT5A.M00493	0/0	0/0	0/0.009	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.006	0/0.001	/	/	0/0.001	0/0	/	/	/	0/0	
STAT5A.M00499	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.001	0/0	/	0/0.006	0/0	/	/	0/0	
STAT6.M00494	0/0	0/0	/	0/0.001	/	0/0	0/0	0/0	/	0/0	0/0.001	0/0	0/0	/	/	0/0.001	/	/	0/0	0/0	/	/	0/0	
STAT6.M00500	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	/	/	0/0	/	0/0	0/0	
NF-AT.M00302	0/0	0/0	0/0.006	0/0	0/0	0/0	0/0	/	0/0.002	0/0	0/0.004	0/0	0/0.001	/	0/0.002	/	/	0/0.001	/	0/0.007	0/0.007	/	0/0	
SRY.M00148	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.006	0/0	0/0	/	0/0	/	0/0	0/0.004	0/0.009	/	0/0	/	0/0	
HMG IY.M00750	0/0	0/0	0/0.008	0/0	0/0	0/0	0/0	0/0	/	0/0.001	/	0/0	0/0	/	/	/	/	/	/	0/0	/	/	0/0	
Pax-2.M00486	0/0	0/0	0/0.001	0/0	0/0	0/0	0/0	0/0.002	/	0/0	0/0.009	0/0	/	/	0/0.002	/	0/0	/	0/0	0/0.004	0/0.004	/	0/0	
IRF																								
IRF.M00772	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
IRF.M00972	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
IRF1.M00747	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.001	0/0	/	0/0	0/0	/	0/0	0/0	/	0/0	
ICSBP.M00699	0/0	0/0	0/0	/	/	0/0	0/0	/	0/0.009	0/0	0/0.002	0/0.002	0/0.001	0/0	0/0	/	0/0.003	0/0.003	/	0/0.001	0/0	0/0	/	
ISRE.M00258	0/0	0/0	0/0	/	/	0/0	0/0.006	0/0	0/0	0/0	0/0	/	0/0.001	0/0	0/0	/	0/0.001	0/0.001	/	/	/	/	0/0.004	
Nuclear hormone receptors (similar matrices to IRF/STAT)																								
AR.M00962	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0.001	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	/	/	0/0	0/0.008	/	0/0	
GR.M00192	0/0	0/0	/	/	/	/	0/0	/	0/0	/	0/0.005	0/0	0/0.001	0/0.001	0/0	/	0/0.001	/	/	/	/	0/0.009	0/0.003	
GR.M00921	0/0	0/0	0/0.004	/	0/0	0/0	0/0.003	0/0	0/0.002	0/0	0/0	0/0	0/0	0/0.001	0/0	0/0.005	/	/	/	/	0/0.004	0/0	0/0.007	
PR, GR.M00960	0/0	0/0.001	0/0.009	0/0.002	0/0.001	0/0.001	0/0	0/0.002	0/0.001	/	0/0.001	0/0	0/0	0/0	0/0	/	0/0.006	/	/	/	0/0	0/0.001	/	

SMAD.M00792	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	/	/
SOX.M01014	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
SREBP-1.M00221	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0.008/0.007	/	/	/	/	/	/	/	/	/
SRF.M00152	/	/	/	0/0	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
SRF.M00215	/	/	/	/	/	/	0/0.001	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
SRF.M00810	/	/	/	/	/	/	0/0.001	/	0.006/0.006	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Sp-1.M00931	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	0/0.002	/
Sp-1.M00932	0/0	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	0/0.002	/
Sp-1.M00933	0/0	0/0	/	/	0/0.005	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	0/0.009	/
Sp1.M00008	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	/	/
Sp1.M00196	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	/	/
Staf.M00262	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	/	/
T3R.M00963	0/0	0/0	/	/	0/0	/	/	/	/	/	/	0/0	/	/	/	/	/	/	/	/	/	/	/	/
TATA.M00216	/	/	/	/	/	/	0/0	/	/	/	/	/	/	/	/	0/0.006	/	/	0/0.001	/	/	0/0.007	/	
TATA.M00252	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	0/0	/	/	/	/	/
TBP.M00471	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	0/0.009	/	/	/	/
TBP.M00980	/	/	/	/	/	/	0/0.004	/	/	/	/	/	/	/	/	/	0/0.003	/	0/0	0/0.002	/	/	/	/
TCF11.M00285	/	0/0.004	0/0.004	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
TEF.M00672	/	/	/	0/0.009	/	0/0	/	/	/	/	/	/	/	/	/	/	/	/	0/0.007	0/0.009	/	/	/	/
TFII-I.M00706	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0.003	/	
TGIF.M00418	/	/	0.009/0	/	0.002/0	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
TTF-1.M00794	0/0	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Tax/CREB.M00114	0.007/0.007	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0.009/0.005	/	/	/	/
Tst-1.M00133	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	0/0.002	/	/	/	/
USF.M00121	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0.005	/	/	/	/	/	/	/	/	/
USF.M00122	/	/	0/0.003	/	/	0/0.008	0/0.002	/	/	/	0/0.003	/	/	/	0/0	/	/	/	0/0	0/0.005	0/0	/	/	/
USF.M00187	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	0/0.003	/	/	0/0	/	/	/	/
USF.M00217	/	/	/	/	/	/	/	/	/	/	0/0	/	/	/	0/0	0/0.001	0/0.002	/	/	/	/	/	/	/
USF.M00796	/	/	/	/	/	0/0	0/0.009	/	/	/	/	/	/	/	0/0	0/0	/	0/0.006	0/0.001	/	0/0	/	/	/
USF2.M00726	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0.001	/	/	/	/	/	/	0/0.006	/	/
VBP.M00228	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
XFD-1.M00267	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	0/0.001	/	0/0	/	/	/	/	/	/
XFD-2.M00268	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0.008	0/0.002	/	0/0	/	/	/	/	/	/
XFD-3.M00269	/	0/0.007	/	/	/	/	/	/	/	/	0/0.001	/	/	/	0/0	/	/	0/0	/	/	0.001/0.002	/	/	/
Xvent-1.M00445	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
YY1.M00059	/	0/0.003	/	/	/	/	0/0	/	0/0	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0.006	/
YY1.M00069	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/
YY1.M00793	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
ZF5.M00333	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
ZID.M00085	/	/	/	/	/	/	0/0.004	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
aMEF-2.M00403	/	/	/	/	/	/	/	/	/	/	0/0.006	/	/	/	0/0.004	/	/	0/0	0/0.003	/	/	/	/	/
alpha-CP1.M00687	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0	/	/	/	/	/	/
c-Myc:Max.M00118	/	/	/	/	/	0/0.003	/	/	/	/	/	/	/	0/0.005	0/0.003	/	/	/	0/0.003	/	/	/	/	/
c-Myc:Max.M00123	/	/	/	/	/	/	0/0.003	/	/	/	0/0.003	/	/	/	/	/	/	/	/	/	/	/	/	/
c-Myc:Max.M00615	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
c-Rel.M00053	/	/	/	/	/	/	/	/	/	/	/	/	/	0/0.009	/	/	/	/	/	/	/	/	/	/
myogenin/NF-1.M00056	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
myogenin.M00712	0/0	/	/	/	/	/	0/0.001	0/0	0/0.004	/	0/0.003	0/0	/	0/0.006	0/0	0/0	0/0	/	/	0/0	/	0/0	/	/
neural-restr-silencer-element.M00325	/	0.001/0	/	/	0/0	/	/	0/0.002	0/0	/	0/0	0.002/0.004	/	0/0	0/0	0.002/0.001	0/0	/	0/0	0.003/0.004	/	0/0	/	/
p53.M00034	/	/	/	/	/	/	/	/	/	/	0.007/0.002	/	/	/	/	/	/	0/0	/	/	/	/	/	/
PEBP.M00984	0/0	0/0	0/0.001	/	0/0.001	0/0.001	/	0/0	/	/	/	/	/	0/0	0/0.006	0/0	/	/	/	/	/	/	/	/
EIk-1.M00025	0/0.002	/	/	/	/	/	/	/	0.004/0.008	/	0/0.006	0.001/0.008	/	/	0/0.001	0/0	0/0.003	/	0/0	/	0/0.004	/	/	/
NERF1a.M00531	0/0.001	0/0	0/0.002	/	/	/	/	/	/	0/0	0/0.007	0/0	/	/	/	0/0.004	/	/	0/0.001	0/0.005	/	0/0.008	/	/

Table S2: The full table of motifs and their corresponding enrichment scores. Motifs that correspond to CTCF (Kim et al., 2007) and 517 vertebrate transcription factors from the TRANSFAC database (Wingender et al., 1996) were tested for their enrichment in the distal DNaseI HS sites in each chromosome. Distal DNaseI HS sites defined as those more than 2kb away from any transcription start site. Two background sequence sets were used for testing the enrichment: the union set of all ChIP-chip hits generated by the ENCODE Transcription regulation group at the 5% false discovery rate cutoff (ENCODE, 2007) and random dinucleotide shuffling of the input sequence set (DNaseI HS sites in a chromosome). The table lists the 50 motifs that showed significant enrichment (p -value < 0.01) in more than 11 chromosomes as well as those which are only enriched in fewer than 12 chromosomes or not enriched at all. The first column lists motif names followed by their TRANSFAC accession numbers, and the subsequent columns are p -values of the motifs in each of the chromosomes in the two background sequences sets, separated by “/”. Missing values indicate that the motif is not enriched in the chromosome. We further grouped the top 50 motifs that have similar matrices, because a motif can appear to be enriched because its matrix is similar to another motif that is truly enriched. Two motifs were considered to be similar if the alignment score between them according to the Malign program (Haverty et al., 2004) was greater than 0.05. Some motifs with Malign score greater than 0.05 were separated into different groups if it was known that some motifs play essential role in regulating either hematopoietic or immune systems. These include (a) AML1 (Nucifora and Rowley, 1995) and C/EBP (Scott et al., 1992), and (b) STAT, IRF, and nuclear hormone receptors

(NHR) (Meraro et al., 1999; Yao et al., 2006). The other two top groupings of TAL1 (Wadman et al., 1997) and ETS (Ho et al., 1990; Tenen et al., 1997) have also been shown to regulate cells in the hematopoietic system.

Supplemental References

Agarwal, S., and Rao, A. (1998). Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. *Immunity* 9, 765-775.

Cockerill, P.N., Shannon, M.F., Bert, A.G., Ryan, G.R., and Vadas, M.A. (1993). The granulocyte-macrophage colony-stimulating factor/interleukin 3 locus is regulated by an inducible cyclosporin A-sensitive enhancer. *Proc Natl Acad Sci U S A* 90, 2466-2470.

ENCODE (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799-816.

Haverty, P.M., Frith, M.C., and Weng, Z. (2004). CARRIE web service: automated transcriptional regulatory network inference and interactive analysis. *Nucleic Acids Res* 32, W213-216.

Ho, I.C., Bhat, N.K., Gottschalk, L.R., Lindsten, T., Thompson, C.B., Papas, T.S., and Leiden, J.M. (1990). Sequence-specific binding of human Ets-1 to the T cell receptor alpha gene enhancer. *Science* 250, 814-818.

Kawaji, H., Kasukawa, T., Fukuda, S., Katayama, S., Kai, C., Kawai, J., Carninci, P., and Hayashizaki, Y. (2006). CAGE Basic/Analysis Databases: the CAGE resource for comprehensive promoter analysis. *Nucleic Acids Res* 34, D632-636.

Lee, D.U., Avni, O., Chen, L., and Rao, A. (2004). A distal enhancer in the interferon-gamma (IFN-gamma) locus revealed by genome sequence comparison. *J Biol Chem* 279, 4802-4810.

Margulies, E.H., Blanchette, M., Haussler, D., and Green, E.D. (2003). Identification and characterization of multi-species conserved sequences. *Genome Res* 13, 2507-2518.

Meraro, D., Hashmueli, S., Koren, B., Azriel, A., Oumard, A., Kirchhoff, S., Hauser, H., Nagulapalli, S., Atchison, M.L., and Levi, B.Z. (1999). Protein-protein and DNA-protein interactions affect the activity of lymphoid-specific IFN regulatory factors. *J Immunol* 163, 6468-6478.

Nucifora, G., and Rowley, J.D. (1995). AML1 and the 8;21 and 3;21 translocations in acute and chronic myeloid leukemia. *Blood* 86, 1-14.

Schoenborn, J.R., Dorschner, M.O., Sekimata, M., Santer, D.M., Shnyreva, M., Fitzpatrick, D.R., Stamatoyannopoulos, J.A., and Wilson, C.B. (2007). Comprehensive

epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. *Nat Immunol* 8, 732-742.

Scott, L.M., Civin, C.I., Rorth, P., and Friedman, A.D. (1992). A novel temporal expression pattern of three C/EBP family members in differentiating myelomonocytic cells. *Blood* 80, 1725-1735.

Siebenlist, U., Durand, D.B., Bressler, P., Holbrook, N.J., Norris, C.A., Kamoun, M., Kant, J.A., and Crabtree, G.R. (1986). Promoter region of interleukin-2 gene undergoes chromatin structure changes and confers inducibility on chloramphenicol acetyltransferase gene during activation of T cells. *Mol Cell Biol* 6, 3042-3049.

Tenen, D.G., Hromas, R., Licht, J.D., and Zhang, D.E. (1997). Transcription factors, normal myeloid development, and leukemia. *Blood* 90, 489-519.

Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *Embo J* 16, 3145-3157.

Wingender, E., Dietze, P., Karas, H., and Knuppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 24, 238-241.

Yao, Z., Cui, Y., Watford, W.T., Bream, J.H., Yamaoka, K., Hissong, B.D., Li, D., Durum, S.K., Jiang, Q., Bhandoola, A., *et al.* (2006). Stat5a/b are essential for normal lymphoid development and differentiation. *Proc Natl Acad Sci U S A* *103*, 1000-1005.