SUPPLEMENTARY MATERIALS & METHODS

Chromatin Immunoprecipitation

ChIP was performed as described.[1] Primer sequences used for ChIP-PCR are available upon request.

ChIP-Seq Binding Site Identification

For GATA2, GATA1, and ETS1 ChIP-Seq datasets, binding sites were identified with MACS[2] (ver. 1.3.7.1) using the full set of unique IP tags as the 'treatment' and an equally-sized, randomly sampled subset of unique input tags as the 'control,' with the default p-value (10^{-5}), and 'mfold' values of 10 (ETS1), 15 (GATA1), or 32 (GATA2). MACS was run three times (A, B, and C from **Table 1**) for each IP tag set, using a different random subset of input, and the base-wise intersection of the three runs was obtained using intersectBed from BEDTools.[3] For the H3K27me3 and H3K4me3 dataset, marked domains were identified three times (A, B, and C from **Table 1**) against three different equally sized random subsets of input using SICER[4] set to a window size of 200, a gap size of 600, and an FDR of 10^{-3}. Binding site locations relative to genes were determined using the ChIP-Seq Tool Set[5] and were assigned to the nearest RefSeq transcription start site within 50kb (for TF ChIP) or 2kb (for histone ChIP) using a custom Perl script. Binding site overlaps between factors were determined with BEDTools[3] and statistical significance was calculated by the genome structure correction test.[6,7] Pol II ChIP-Seq reads were obtained from the Gene Expression Omnibus (GSE27967).[8]

Gene expression data analysis

Processed array data for GATA2 knockdown (GSE16521) were downloaded from GEO. GATA2 knockdown and GATA1 restoration gene expression data were used to build

gene by gene linear models to quantify the significance of the treatment effect by empirical Bayes t-tests. Genes with FDR adjusted (Benjamini-Hochberg) p-values less than 0.05 were considered significantly differentially expressed. Pearson's Chi-squared tests with Yates' continuity correction were used to evaluate the significance of enrichment between and among the ChIP-Seq and gene expression datasets.

Motif analysis

Motif analysis was performed on 500bp sequences positioned on the center of the MACS binding site with DREME[9] using the default settings. Motifs were passed to the TOMTOM motif comparison tool[10] to identify nearest match motifs within transcription factor databases.

Analysis of tag density and generation of heatmaps

Matrices for the tag density heatmaps were generated using HOMER,[11] clustered using *k*-means clustering (*k*=10) in Cluster3,[12] and visualized in TreeView.[13] GATA1 and GATA2 binding heatmaps from Figure 6 were created in R from the MACS wiggle files.

SUPPLEMENTAL REFERENCES

1. Huang Z, Dore LC, Li Z, et al. GATA-2 Reinforces Megakaryocyte Development in the Absence of GATA-1. *Mol Cell Biol*. 2009;29(18):5168-5180.
2. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
3. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*. 2010;26(6):841-842.
4. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics (Oxford, England)*. 2009;25(15):1952-1958.
5. Blahnik KR, Dou L, O'Geen H, et al. Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic Acids Research*. 2010;38(3):e13.
6. Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799-816.

7. Bickel P, Boley N, Brown J, Huang H, Zhang N. Subsampling methods for genomic inference. *The Annals of Applied Statistics*. 2010;4(4):1660-1697.
8. Young MD, Willson TA, Wakefield MJ, et al. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Research*. 2011.
9. Bailey TL. DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics (Oxford, England)*. 2011.
10. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. *Genome Biology*. 2007;8(2):R24.
11. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*. 2010;38(4):576-589.
12. de Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics (Oxford, England)*. 2004;20(9):1453-1454.
13. Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics (Oxford, England)*. 2004;20(17):3246-3248.

Table S1. Relationship between bivalent chromatin domains, gene promoters, and GATA switch sites

Genomic Region	Count	Percentage of Parent Class
RefSeq Promoters (mm9)	28542	--
H3K4me3 Domains	36277	--
H3K27me3 Domains	42631	--
Bivalent Domains	8505	23.4% (of H3K4me3) 20.0% (of H3K27me3)
• Overlapping Promoters	2866	33.7% (of Bivalent)
• Containing Switch Sites	505	5.94% (of Bivalent)
• Overlapping Promoters & Containing Switch Sites	257	3.02% (of Bivalent) 8.97% (of Bivalent Overlapping Promoters) 50.9% (of Bivalent Containing Switch Sites)
GATA Switch Sites	5451	--
• Overlapping Promoters	608	11.2% (of Switch Sites)
• Overlapping Bivalent Domains	529	9.70% (of Switch Sites)
• Overlapping Promoters & Bivalent Domains	37	0.68% (of Switch Sites) 6.09% (of Switch Sites in Promoters) 7.00% (of Switch Sites in Bivalent Domains)

Table S2. The 37 genes with bivalent switch promoters and their response to GATA1 restoration and GATA2 knockdown in G1ME cells

CHR	START	END	GENE	Fold Change	
				+GATA1	-GATA2
chr1	80209610	80210891	Fam124b	NA	NA
chr1	97563722	97564833	St8sia4	0.368	0.718
chr1	137904186	137904980	Tmem9	1.287	NA
chr1	187278072	187278946	Slc30a10	3.699	NA
chr2	34769835	34770748	Phf19	n.s.	n.s.
chr2	84877522	84878331	Ssrp1	n.s.	NA
chr2	91096098	91096929	Pacsin3	n.s.	NA
chr2	130665941	130666433	4930402H24Rik	n.s.	n.s.
chr3	90269909	90270684	Npr1	n.s.	n.s.
chr3	122322994	122323858	Fnbp1l	0.554	n.s.
chr4	109078587	109079248	Ttc39a/AK148461	NA	NA
chr4	154720243	154720895	Prkcz	n.s.	n.s.
chr4	155576456	155577232	AW011738	NA	NA
chr5	73127243	73128566	Txk	0.737	n.s.
chr5	114280126	114281044	Selplg	0.515	0.388
chr5	121098350	121099084	Rasal1	n.s.	NA
chr6	116623401	116624448	Rassf4	n.s.	n.s.
chr6	127059330	127060128	9630033F20Rik	0.763	n.s.
chr8	74035290	74036132	Plvap	1.991	n.s.
chr8	87324504	87325178	Nfix	n.s.	n.s.
chr8	119606160	119607400	Pkd1l2	n.s.	n.s.
chr9	78295665	78296360	Mto1	n.s.	n.s.
chr9	110959085	110960269	Ccrl2	n.s.	n.s.
chr11	82765459	82766210	Slfn5	n.s.	n.s.
chr11	95885394	95886291	Gip	n.s.	n.s.
chr11	96179492	96180283	Hoxb4	n.s.	n.s.
chr11	119215884	119217148	Sgsh	n.s.	n.s.
chr12	9036296	9037871	Ttc32	0.765	NA
chr12	109513940	109514913	Ccdc85c	NA	NA
chr13	54722400	54723913	Faf2	NA	NA
chr13	97840819	97841258	Fam169a	NA	NA
chr15	55139146	55139761	Col14a1	NS	n.s.
chr15	103197402	103198190	Itga5	n.s.	n.s.
chr17	12700495	12701379	Slc22a3	0.295	n.s.
chr17	37005962	37006868	Trim10	2.143	NA
chr19	19185605	19186232	Rorb	n.s.	NA
chrX	92221600	92222226	Spin4	n.s.	n.s.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. GATA1 restoration leads to megakaryocytic maturation of G1ME cells. **(A)** Cell surface expression of CD42 (glycoprotein 1b) 72 hours after MIGR1 or GATA1 transduction of G1ME cells. **(B)** Increased DNA content of G1ME cells after restoration with GATA1 compared to transduction with empty MIGR1 vector.

Figure S2. GATA1 and GATA2 binding sites are close to transcription start sites. GATA1 and GATA2 binding sites were binned according to their location relative to the nearest annotated transcription start site and depicted as a histogram with a resolution (bin size) of 50bp. Lines in gold and tan are the result of analyses performed on background sets, as in (Figure 1C).

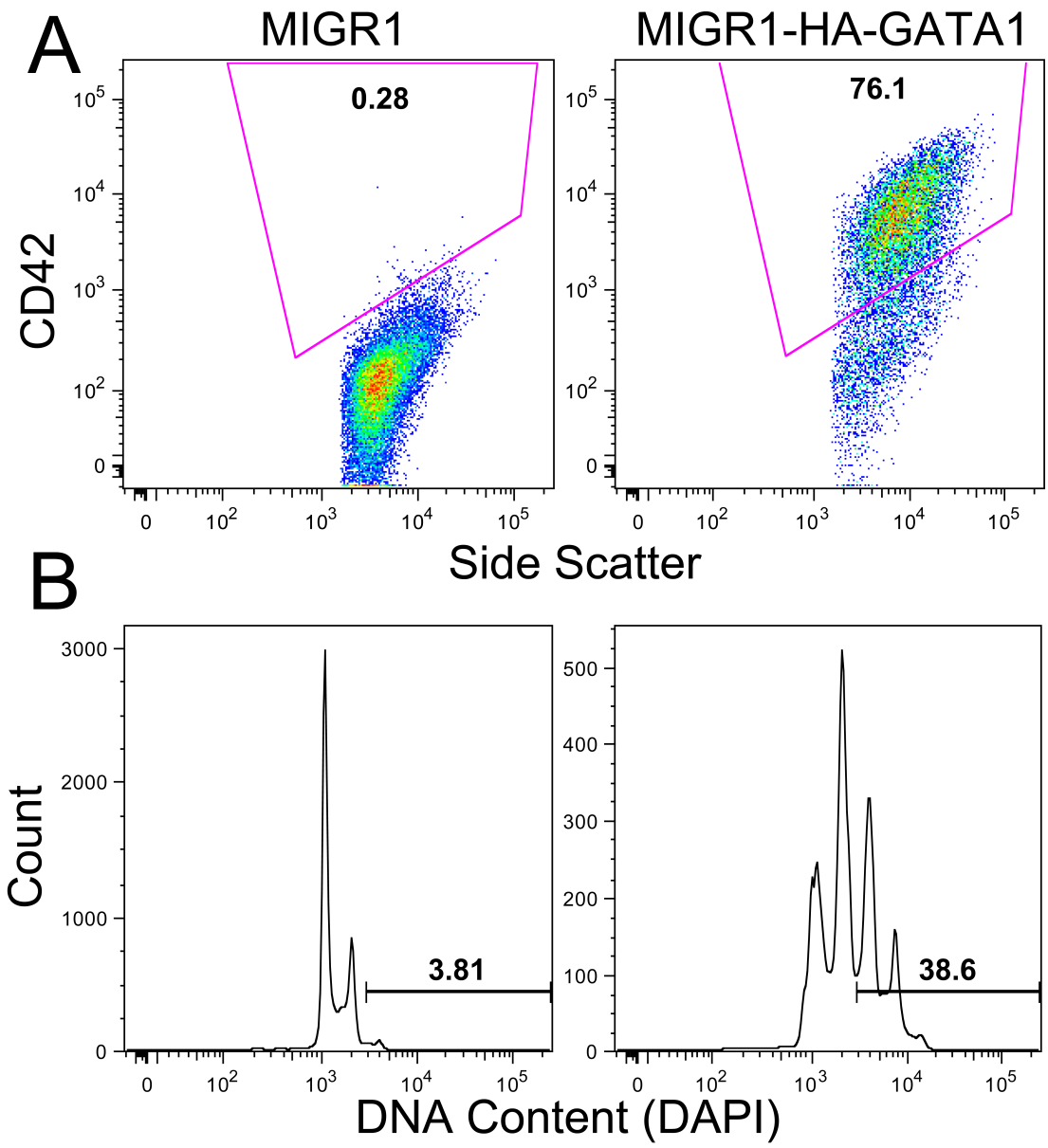
Figure S3. (A) Genome browser image showing GATA1 (green) and GATA2 (blue) occupancy at the *Sfpi1* locus, which encodes the master myeloid transcription factor, PU.1. Height of signal along y-axis correlates with ChIP-Seq tag count (signal intensity). **(B)** Genome browser image of GATA1 and GATA2 binding sites at the *Gata2* locus. **(C)** GATA1 and GATA2 ChIP-PCR was performed in G1ME cells infected with a virus that expresses a cell surface marker, $H2K^k$, or a bicistronic vector expressing GATA1 and $H2K^k$. Two days after infection, the cells were sorted using magnetic anti-$H2K^k$ beads and purified cells were subjected to our standard ChIP protocol. Y-axis shows relative occupancy as determined by comparison to a standard curve of input dilutions using the appropriate primers.

Figure S4. ETS1, but not ERG or ETS2, occupies many GATA2 binding sites in G1ME cells. ChIP-PCR for ERG (blue), ETS1 (red), and ETS2 (green) shows enrichment for ETS1 at GATA2 binding sites associated with several critical hematopoietic genes.

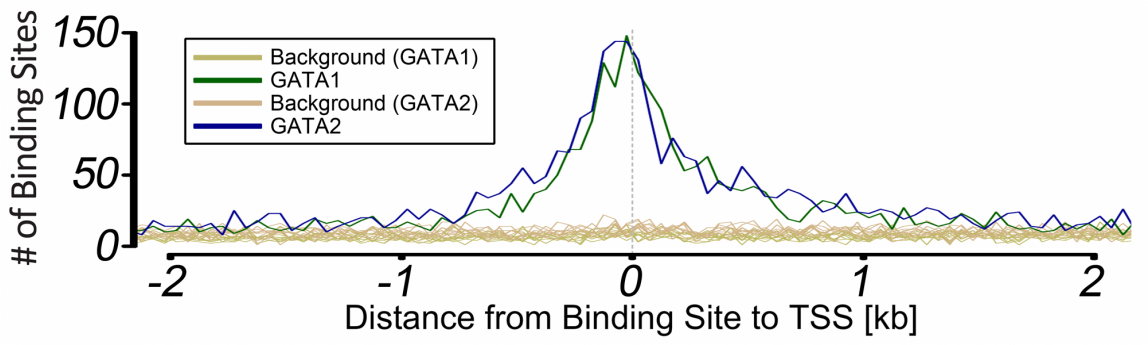
Figure S5. ETS1 binding sites overlap with small portions of GATA factor binding sites in G1ME cells. (A) Venn diagram showing the overlap of ETS1 binding sites with GATA2 binding sites in untransduced G1ME cells. **(B)** Venn diagram showing the overlap of ETS1 binding sites with genomic regions characterized as GATA switch sites in untransduced G1ME cells. **(C)** Motifs that were identified as enriched in the ETS1 binding sites include several ETS family motifs.

Figure S6. Distribution of sizes of genomic binding sites and marked domains. Box-and-whisker plots show the distribution of genomic sizes identified as binding sites for each of the transcription factors (left) and marked domains for each histone methylation mark (right).

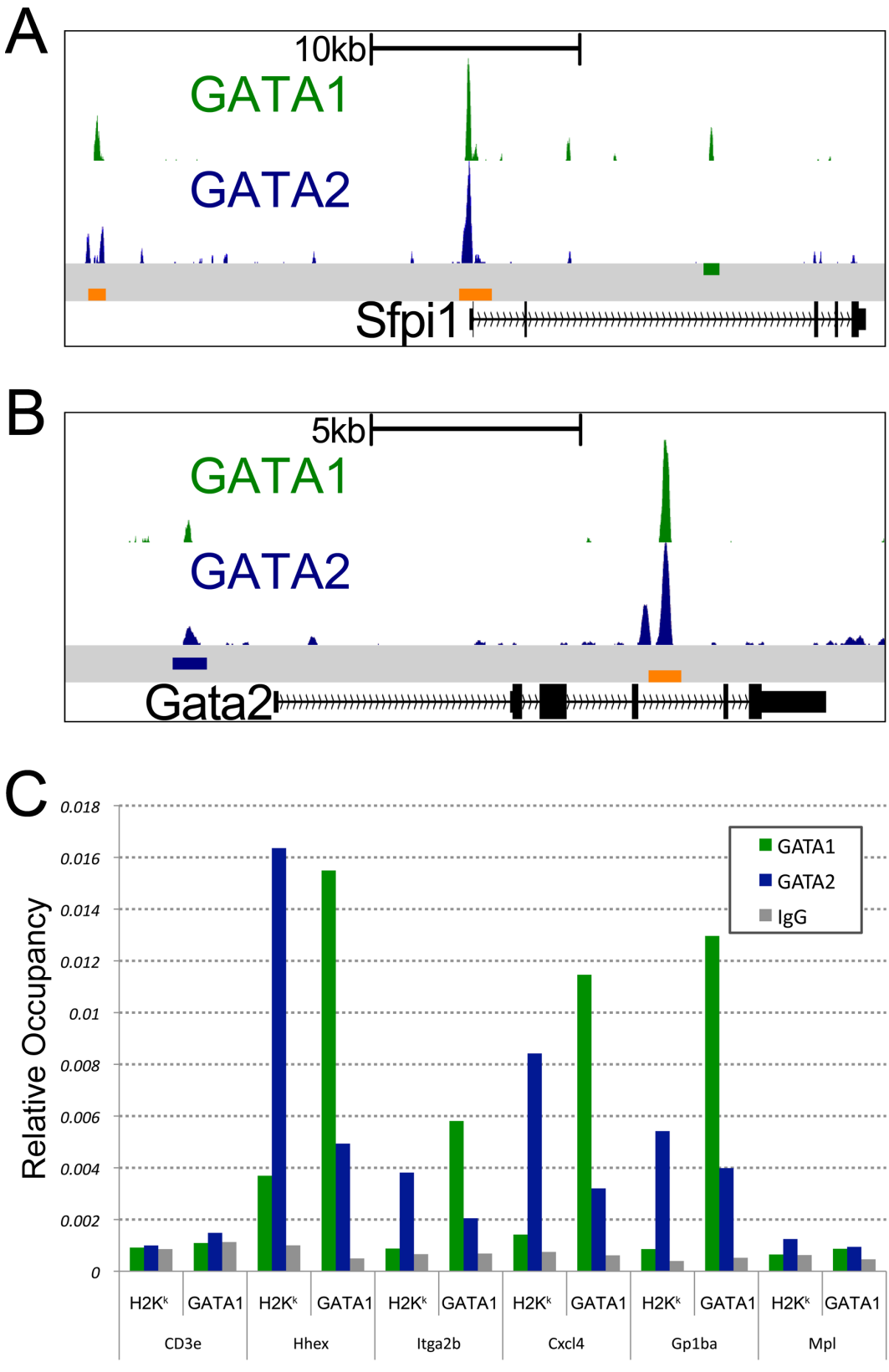
Figure S7. Regions bound by both GATA2 and ETS1 have higher H3K4me3 and PolII tag counts. Heatmaps depicting the tag counts for each ChIP-Seq dataset across the genomic regions characterized as GATA2/ETS1 shared sites (top), GATA2-selective sites (middle), and ETS1-selective sites, as in Figures 5A & 6A.



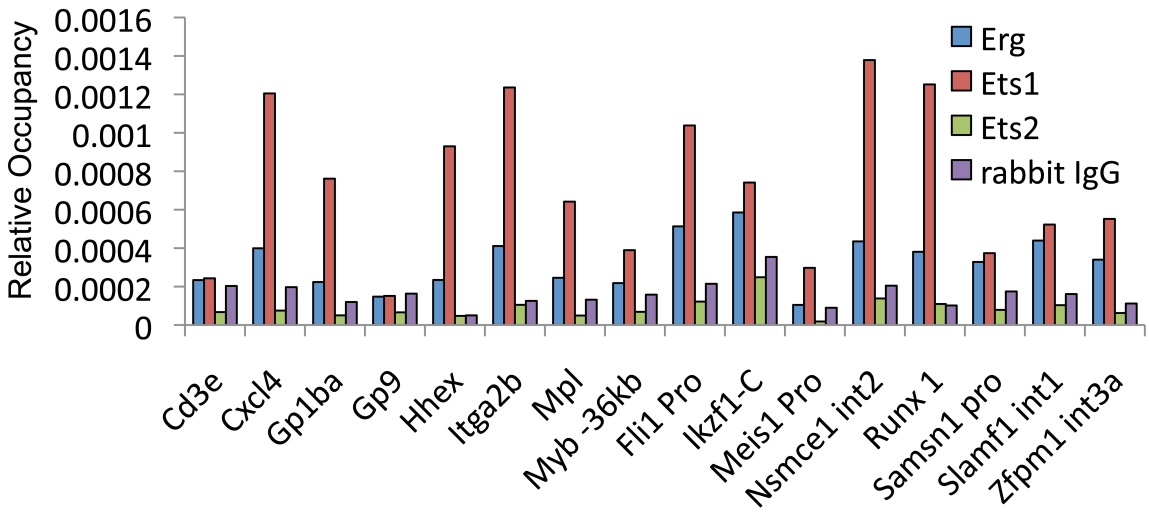
SUPPLEMENTARY FIGURE 1



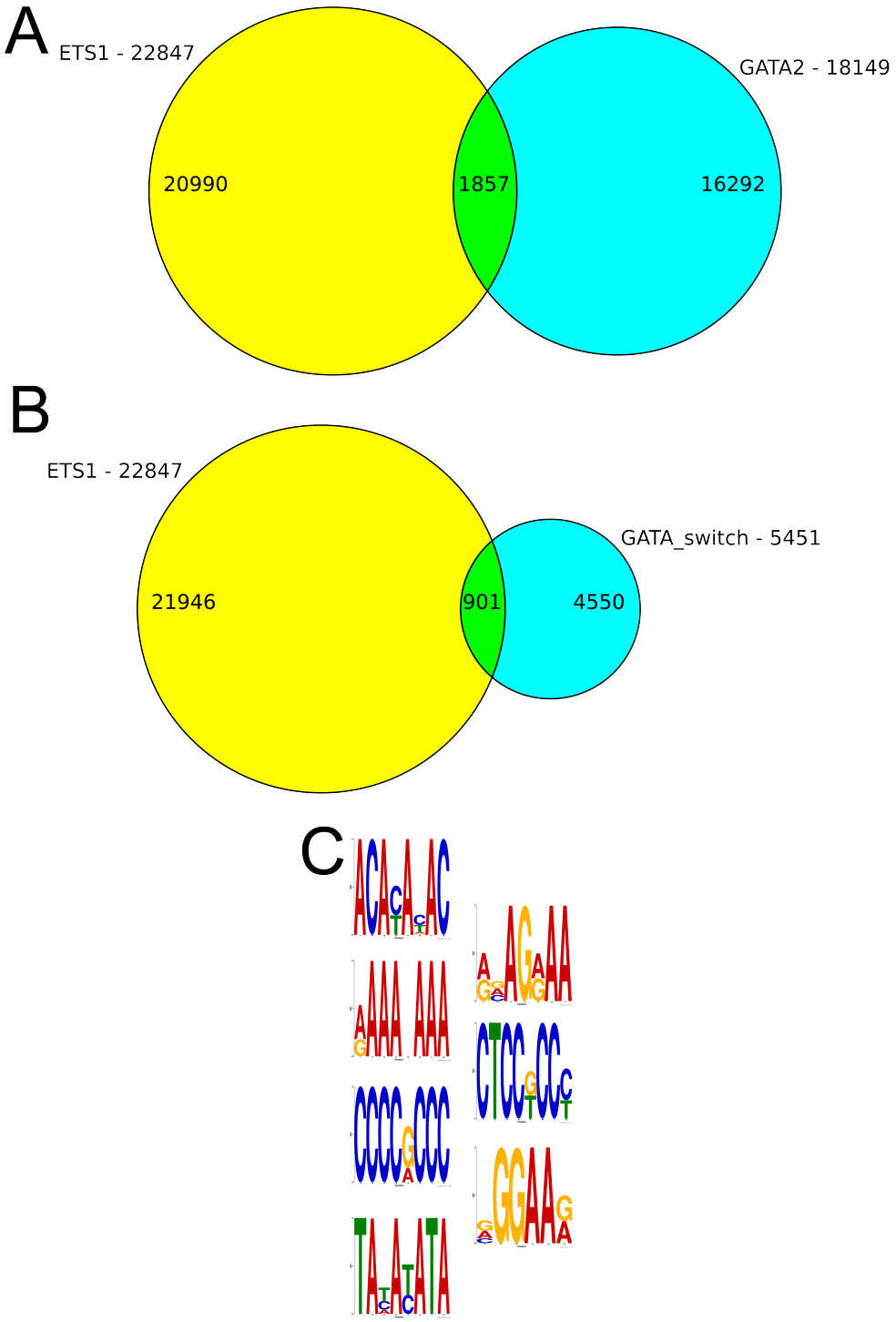
SUPPLEMENTARY FIGURE 2



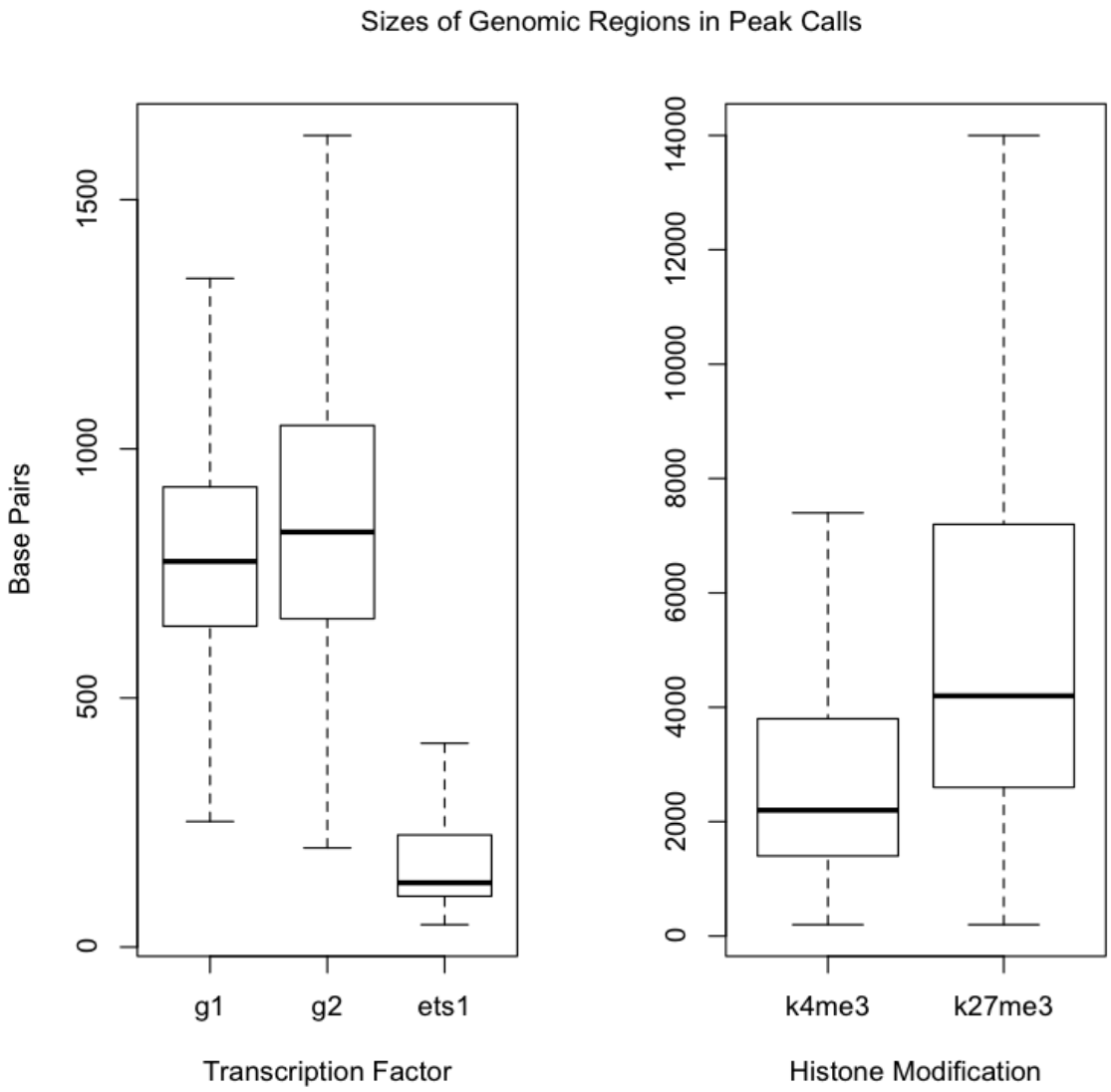
SUPPLEMENTARY FIGURE 3



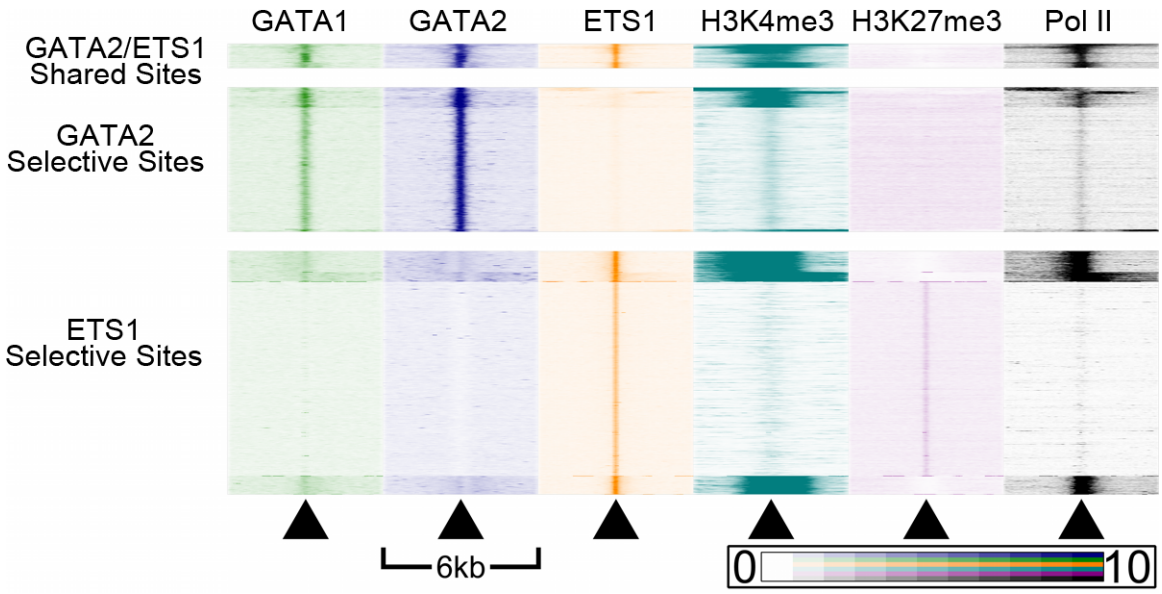
SUPPLEMENTARY FIGURE 4



SUPPLEMENTARY FIGURE 5



SUPPLEMENTARY FIGURE 6



SUPPLEMENTARY FIGURE 7