Supplemental methods and results.

Genome wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding.

A Gordon Robertson, Mikhail Bilenky, Angela Tam, Yongjun Zhao, Thomas Zeng, Nina Thiessen, Timothee Cezard, Anthony P Fejes, Elizabeth D. Wederell, Rebecca Cullum, Ghia Euskirchen, Martin Krzywinski, Inanc Birol, Michael Snyder, Pamela A Hoodless, Martin Hirst, Marco A Marra, Steven JM Jones.

Contents

1. Supplemental methods

Calculating enrichment (XSET coverage) profiles For each enriched STAT1 or FoxA2 binding site we calculated H3K4me1 and H3K4me3 coverage profiles for a ±5-kb region around the location of a region's maximum height. For each 5-kb region we calculated the coverage at each nucleotide as the number of 200 bp directionally extended single-end tags (XSETs) that overlapped this nucleotide (Wederell et al. 2008; Robertson et al. 2007). For each TF and histone modification we summed all individual profiles to generate a metaprofile, then normalized each metaprofile by the total coverage (i.e. the area under the curve) to allow comparing profiles for libraries that had been sequenced to different depths. Coverage profiles were not thresholded using an FDR. We assigned no strand to distal regions, and displayed the results as 'folded' half-profiles that showed coverage as a function of absolute distance from a TF binding site; this presentation suppressed strand-related asymmetry for proximal sites.

Filtering against simple tandem repeats and segmental duplications Certain types of repetitive genomic regions are potential sources of artifacts in ChIP-chip experiments (Johnson, et al. 2008). To assess the potential for such artifacts, we determined the fractional overlap of all enriched regions with all RepeatMasker annotation types, and with repeats identified by Tandem Repeat Finder (TRF), using UCSC hg18 and mm8 data (Karolchik et al. 2008). Relatively few regions had a high fraction of their length overlapped by a TRF-repeat. For HeLa data, an overlap threshold of 80% identified <0.7% of H3K4me1 regions in unstimulated and stimulated cells, and 5.2% of STAT1 regions in unstimulated cells and 1.8% in IFNG-stimulated cells. For mouse adult liver cells, this threshold identified 0.4% of H3K4me1 regions and 1.4% of Foxa2 sites. Profiling read mappability across segmental duplications showed that reads generally can be mapped into such duplications, and flags cases in which mappability is low (data not shown). Given these two factors, we retained all enriched regions in the analyses.

Region overlap and association We calculated overlap and association rates for whole-genome region sets and for ENCODE regions (Supplemental document, sections AD and EN). For whole-genome results a STAT1 or Foxa2 region was considered to be associated with an H3K4me1 region if the distance from the transcription factor region maximum to the nearest H3K4me1 region maximum was less than 1kb. We used the same approach for HeLa cells to calculate overlap between untreated and treated STAT1 or H3K4me1 regions. We generated random expectations for overlap/association rates in a pairwise region set comparison by randomly reassigning locations for one region set on each chromosome. For ENCODE regions, we estimated random expectations by this approach, and also by a block bootstrap approach, reporting 99.9% confidence intervals (The ENCODE Project Consortium).

For comparisons with distal predicted enhancers and p300 sites reported by Heintzman et al. (2007), we considered only ChIP-seq H3K4me1- and me3-enriched regions whose maxima were within the ENCODE regions. We identified distal locations as those further than ±2.5 kb from a UCSC hg18 known gene TSS.

Distinguishing proximal and distal regions Assigning a distance to a TSS for a ChIP-seq enriched region is complicated by genes having multiple transcripts, genes being closely adjacent or overlapping, and different annotation systems identifying some

unique transcripts. Because the first two factors made it difficult to assign both an unambiguous 'genic' status and a distance to a nearest transcriptional start site (TSS) to a STAT1 or Foxa2 binding site, we chose not to distinguish intergenic from genic sites; instead, we identified an site as 'distal' if the absolute distance from it to the nearest TSS was greater than a threshold value. The third factor required that we choose one or more transcript annotation systems to work with. In the results reported, we assigned to each STAT1 or Foxa2 region a distance to the closest TSS for a UCSC hg18 or mm8 known gene, and distinguished proximal and distal regions using a distance threshold of ±2.5 kb (Heintzmann et al. 2007; The ENCODE Project Consortium).

Filtering distal TF binding sites against additional predicted TSSs and CAGE tags
We reclassified, as proximal, STAT1 binding sites in stimulated HeLa cells that were distal to hg18 known genes, by considering predicted AceView (Thierry-Mieg and Thierry-Mieg 2006), Genscan (Burge and Karlin 1997) and SwitchGear TSSs (SwitchGear Transcription Start Site Predictions, hosted at www.switchdb.com) from the UCSC hg18 genome browser. We also tested filtering by a set of ~123k CAGE tag clusters from (Abeel et al. 2008), using the centre of each tag cluster as a reference location. For mouse Foxa2 sites, we considered UCSC mm8 SIB gene (Benson et al. 2004) and Genscan predictions.

Effect of more stringent FDR thresholds on TF-associated H3K4me1-me3 patterns
We assessed whether results for proportions of histone modification patterns for distal STAT1 sites was robust to using a more stringent profile-specific FDR thresholds. The FindPeaks v2.0 profiling application (Fejes et al. 2008) generated an FDR-region height relationship for each profile that was close to linear over most of its range, with each integer increase in height corresponding to approximately half an order of magnitude in FDR. We compared cumulative distributions of peak height for STAT1, H3K4me1 and me3 for proximal and distal STAT1 binding sites. We then profiled the proportion of the four different combinations of histone modifications (or no modifications), associated with distal STAT1 sites for the default FDR threshold and height increments up to 4 higher, corresponding approximately to an FDR range that decreased from ~0.01 to 100 times smaller than this (Supplemental document, section FE).

TFBS enrichment in genomic regions flanked by H3K4me1 We identified a high confidence global set of H3K4me1-associated regions as follows. We generated density distributions of the distance between consecutive pairs of H3K4me1 regions (Supplemental Fig. TF2). From these distributions, we selected minimum and maximum distance criteria of 200 and 1000 bp to identify flanked regions. We considered such a region to be distal if its center was further than ±2.5 kb from the TSS of a UCSC hg18 or mm8 known gene (Heintzman et al. 2007).
　　　　For each region we selected the genomic sequence between locations of maxima of the flanking H3K4me1 regions, and we also generated zeroth order Markov randomized sequences that had the same overall (A+T) content of the genomic sequence. We generated five random sequences per region in order to determine overall enrichment trends.
　　　　We used a custom 'gScan' algorithm (Wederell et al. 2008) to scan genomic and randomized sequences with transcription factor binding sequences (TFBS) for the 319 mammalian TRANSFAC v9.3 models for which site sequences were available (Matys et

al. 2006). For a stringent first assessment, we required that a region sequence contain an exact match to at least one of a TFBS sequence from a TRANSFAC model. For each transcription factor model, we also estimated two enrichment parameters: a ratio of the number of such sites in the genomic sequences to the mean number per randomized sequence set, and a Z-score for the number of matches in actual and randomized sequences. We considered a motif enriched if it had a ratio of at least 2.0 and a Z-score of at least 3.0, and had exact sequence matches in at least 50 flanked regions (Supplemental Fig. TF4). We determined sets of enriched models separately for distal and proximal flanked regions.

TF binding sites associated with published CTCF data   We used three published sources of CTCF data. From ChIP-chip data in primary human fibroblasts, Kim et al. (2007) reported 12.8k conserved CTCF motifs for human, as well as 6.6k conserved CTCF motifs for mouse. Barski et al. (2007) made available ChIP-seq CTCF read data for human T cells, from which we generated an enrichment profile that contained 38.4k enriched regions at an FDR of ~0.01. Chen et al. (2008) reported 64k ChIP-seq CTCF locations from mouse ES cells.

For both the human conserved motifs and enriched regions, we determined a 150 bp association distance from the cumulative distribution of distances between a STAT1 region maximum and a CTCF motif centre or region maximum. For proximal and distal STAT1 regions in stimulated HeLa cells, we used this distance threshold to determine proportions of combinations of H3K4me1 and H3K4me3 modifications that were associated with all STAT1 sites and with STAT1 sites that were not associated with either a conserved CTCF site or a ChIP-seq CTCF enriched region.

For mouse, we used the UCSC mm8 genome browser Table Browser intersection tool with BED-format files for STAT1 and CTCF conserved motifs or ChIP-seq regions to determine that only 86 (0.8%) of 10,970 Foxa2 enriched regions overlapped a conserved CTCF site. Given that this number was low, and that no strong spatial relationship was evident between our Foxa2 locations and Chen et al. (2008) locations, we did not pursue further calculations for Foxa2 and CTCF.

Concordance between unstimulated and IFNG-stimulated HeLa cells   We used two approaches to estimate concordance rates for H3K4me1, H3K4me3 and STAT1. Because of differences between region sets between stimulated and unstimulated cells, results were asymmetric; e.g. the percentage concordance for stimulated H3K4me1 regions against unstimulated regions was much lower than for unstimulated H3K4me1 regions against stimulated regions.

For the first approach we generated cumulative distributions of distances between locations of region maxima in IFNG-stimulated and unstimulated HeLa cells, for a range of profile threshold heights that corresponded approximately to the upper 70%, 50% and 30% of region heights. We then thresholded the distributions at the median width of the stimulated regions from Table 1 (Supplemental Fig. MC1A, 2A, 3A). For the second approach we calculated the distribution overlap for stimulated and unstimulated regions directly, and generated a random expectation by shuffling the genomic coordinates of one of the sets of regions (Supplemental Fig. MC1B, 2B, 3B).

References

Abeel T, Saeys Y, Rouzé P, Van de Peer Y. 2008. ProSOM: core promoter prediction based on unsupervised clustering of DNA physical profiles. *Bioinformatics*. **24**: i24-31.

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.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2004. GenBank: update. *Nucleic Acids Res.* **32**: D23-6.

Burge C. Modeling Dependencies in Pre-mRNA Splicing Signals. In Salzberg S, Searls D, Kasif S, eds. Computational Methods in Molecular Biology, Elsevier Science, Amsterdam. 1998; 127-163.

Chen X, Xu H, Yuan P, Fang F, Huss M, et al. (2008) Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* **133**: 1106-1117.

The ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799-816.

Fejes AP, Robertson AG, Bilenky MB, Varhol R, Bainbridge MN, et al. 2008. FindPeaks 3.1: A Java Application for Identifying Areas of Enrichment from Massively Parallel Short-Read Sequencing Technology. *Bioinformatics*. **24**: 1729-30.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39:** 311-318.

Johnson, D. S., Li, W., Gordon, D. B., Bhattacharjee, A., Curry, B., Ghosh, J., Brizuela, L., Carroll, J. S., Brown, M., Flicek, P., et al. 2008. Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. *Genome Res.* **18:** 393-403.

Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, et al. 2008. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res*. **36**(Database issue): D773-9.

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**: 1231-45.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**(Database issue): D108-10.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4:** 651-657.

Thierry-Mieg, D. and Thierry-Mieg, J. 2006. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* **7 Suppl 1:** S12.1-14.

Wederell, E. D., Bilenky, M., Cullum, R., Thiessen, N., Dagpinar, M., Delaney, A., Varhol, R., Zhao, Y., Zeng, T., Bernier, B., et al. 2008. Global Analysis of In Vivo FoxA2 Binding Sites in Mouse Adult Liver Using Massively Parallel Sequencing. *Nucleic Acids Res.*, Advance Access published on July 8, 2008.

2. Supplemental results

SA Saturation of histone modifications and transcription factors
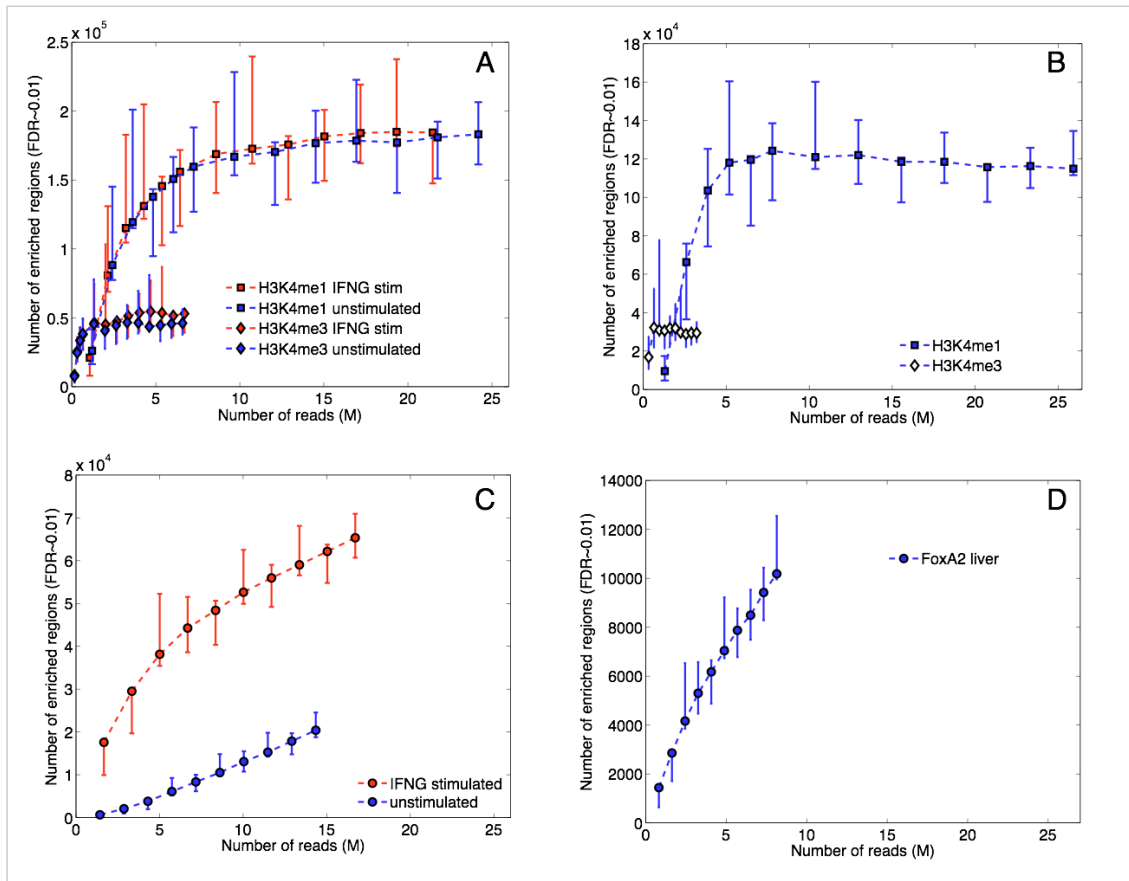
Figure SA1. Sequencing depth and profile saturation. (A) In both unstimulated and IFNG-stimulated HeLa S3 cells, simulated sequencing depths of ~8 to 10 M reads for H3K4me1 and ~1 M aligned reads H3K4me3 determined the same number of enriched regions as the full experimental sequencing depths. (B) Similarly, we estimated that ~6 to 8M and ~1M aligned reads, respectively, saturated these histone modification profiles in mouse adult liver cells. In contrast, the number of TF-enriched regions increased continuously with simulated sequencing depth for (C) STAT1-enriched regions in unstimulated and stimulated HeLa cells, and (D) Foxa2 regions in adult mouse liver. All saturation curves shown are for a constant estimated FDR of ~ 0.01 (Methods).

GA Global association of STAT1 and Foxa2 with H3K4me1 and H3K4me3.

From Fig. GA1 we set an association distance threshold of 1 kb.

A   **B**

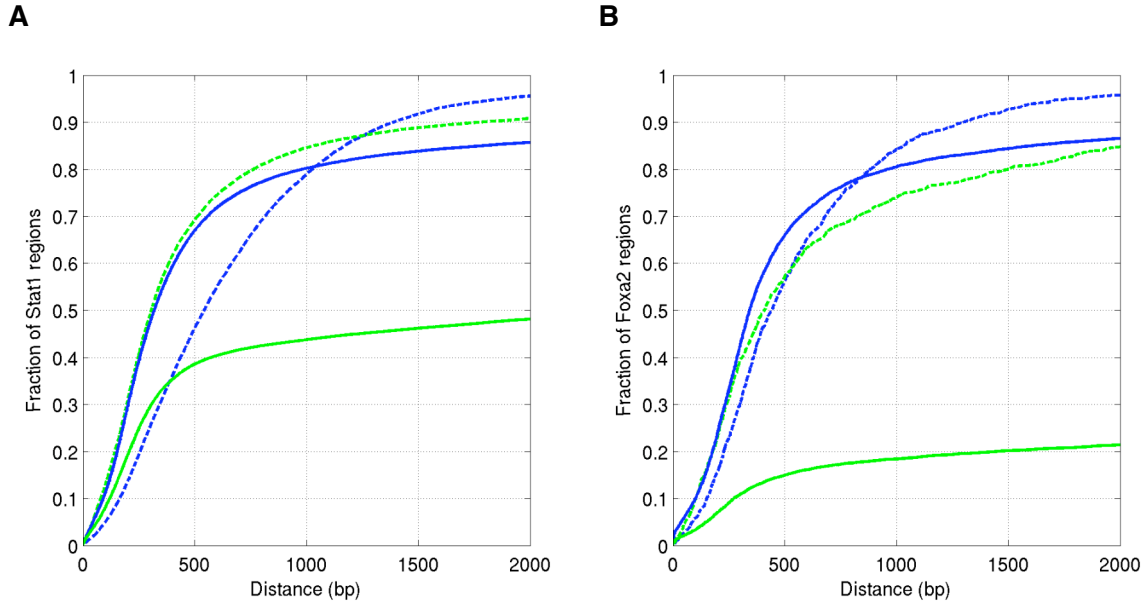**Figure GA1.** Association distance distributions for A) STAT1 and B) Foxa2 binding sites. The graphs show cumulative distributions for the distance between the location of maximum coverage of a distal (solid) or proximal (dotted) TF-enriched region and the closest maximum of an H3K4me1 (blue) or H3K4me3 (green) region.

Tables GA1 and 2 show the number and percentage of STAT1 (IFNG) sites associated with different combinations of H3K4me1 and H3K4me3.

Table GA1. Association rates for all, proximal and distal STAT1 (IFNG) sites with H3K4me1 or me3 regions.

	All sites		Proximal sites		Distal sites	
	Nbr	% total	Nbr	% prox.	Nbr	% distal
Total	70,292	Na	17,296	24.6%	52,996	75.4%
Overall assocn						
No marks	9,446	13.4%	656	3.8%	8,790	16.6%
H3K4me1	56,152	79.9%	13,659	79.0%	42,493	80.2%
H3K4me3	37,811	53.8%	14,626	84.6%	23,185	43.7%
Either mark	60,846	86.6%	16,640	96.2%	44,206	83.4%
Pattern assocn						
H3K4me1 only	23,035	32.8%	2,014	11.6%	21,021	39.7%
H3K4me3 only	4,694	6.7%	2,981	17.2%	1,713	3.2%
Both marks	33,117	47.1%	11,645	67.3%	21,472	40.5%

Of distal binding sites, 83.4% were associated with at least one of the two histone modifications, and H3K4me1 was associated with over 42,493/44,206 = 96.1% of these marked sites.

Table GA2. Association rates for all, proximal and distal Foxa2 sites with H3K4me1 or me3 regions.

	All sites		Proximal sites		Distal sites	
	Nbr	% total	Nbr	% prox.	Nbr	% distal
Total	10,970	Na	1,404	12.7%	9,566	87.2%
Overall assocn						
No marks	1,243	11.3%	30	2.1%	1,677	17.5%
H3K4me1	8,887	81.0%	1,183	84.3%	7,704	80.5%
H3K4me3	2,796	25.5%	1,038	73.9%	1,758	18.4%
Either mark	9,237	84.2%	1,219	86.8%	7,889	82.5%
Pattern assocn						
H3K4me1 only	6,441	58.7%	181	12.9%	6,131	64.1%
H3K4me3 only	350	3.2%	36	2.5%	185	2.0%
Both marks	2,446	22.3%	1,002	71.4%	1,573	16.4%

Of distal binding sites, 82.5% were associated with at least one of the two histone modifications, and H3K4me1 was associated with over 7,704/7,889 = 98.9% of these marked sites.

EN   Associated histone modifications in ENCODE regions.

From Fig. EN1 we set a 1 kb association distance threshold. This threshold returned the results in Tables EN1 and 2, which are shown as bar graphs in Fig. EN2.
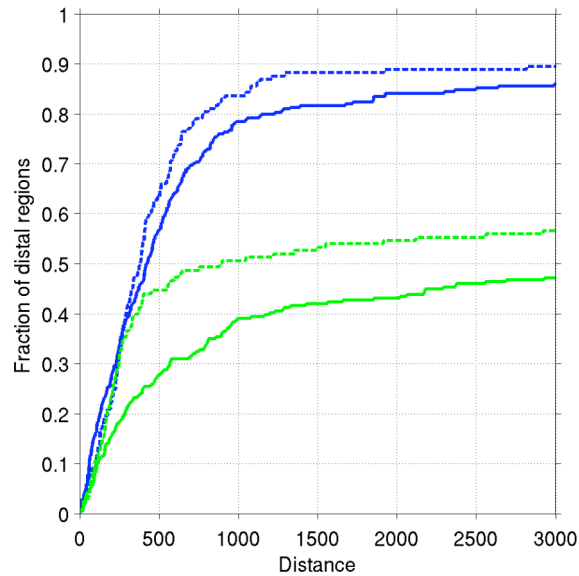

Figure EN1. Association distances for distal p300 (solid lines) and distal predicted enhancers (dashed lines) from Heintzman et al. 2007 with our ChIP-Seq H3K4me1 (blue) and H3K4me3 (green) regions from stimulated HeLa cells. The calculation that generated the cumulative distributions is explained in the caption of Supplemental Fig. AD1.

Table EN1. Numbers of histone modification combinations associated with proximal and distal p300 sites or distal predicted enhancers from Heintzman et al. (2007), using a 1kb association distance.

	me3 only	me1+me3	me1 only	no mark	Total
Proximal p300	4	15	6	4	29
Proximal pred Enh	0	18	12	10	40
Distal p300	5	71	57	20	153
Distal pred Enh	9	97	126	52	284

Table EN2. As table EN1, but percentages.

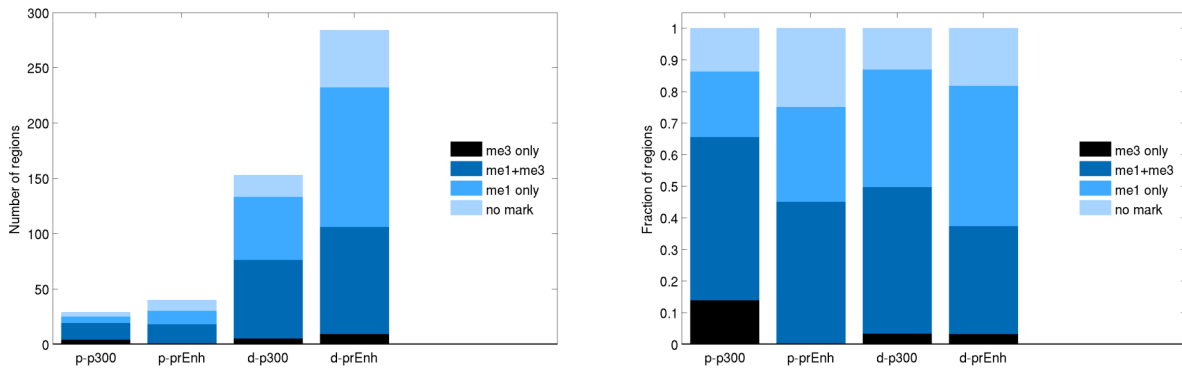|	me3 only	me1+me3	me1 only	no mark	Total
Proximal p300	13.8%	51.7%	20.7%	13.8%	100%
Proximal pred Enh	0.0%	45.0%	30.0%	25.0%	100%
Distal p300	3.3%	46.4%	37.2%	13.1%	100%
Distal pred Enh	3.2%	34.2%	44.4%	18.3%	100%

Figure EN2. Barchart display of Tables EN1 and 2. Proximal and distal are indicated by 'p' and 'd'.

MP   Read mappability of genomic regions flanking STAT1 sites in IFNG-stimulated HeLa cells.

Sets of enriched ChIP-seq regions can contain false negatives that are caused by local mappability deficits that reduce aligned read densities. Fig. 5 shows examples of genomic regions that have relatively small fractions of low mappability regions. To assess the potential role of mappability deficits on results for associated histone modifications, we profiled the average read mappability around subsets of distal STAT1 sites in stimulated HeLa cells and Foxa2 binding sites in mouse adult liver. We considered sites that were associated with H3K4me1 only, or with neither H3K4me1 nor me3. We divided H3K4me1-associated sites into those that had the expected flanking modifications, and those that had the modification only on one side.

Except for the Foxa2 binding sites that were associated with neither modification, TF sites were at genomic locations that had high read mappability. For both IFNG-stimulated STAT1 and Foxa2, average mappability profiles were consistent with some cases of asymmetric associated H3K4me1 being caused by flanking mappability deficits. However, for both STAT1 and Foxa2 sites, average mappabilities in flanking regions were between ~0.8 and ~0.9, i.e. were relatively high. This suggests that most cases in which a distal TF site lacked associated H3K4me1 were caused by the TF have been associated either with no modifications, or with different modifications than those we profiled, rather than being false negatives caused by read mappability.
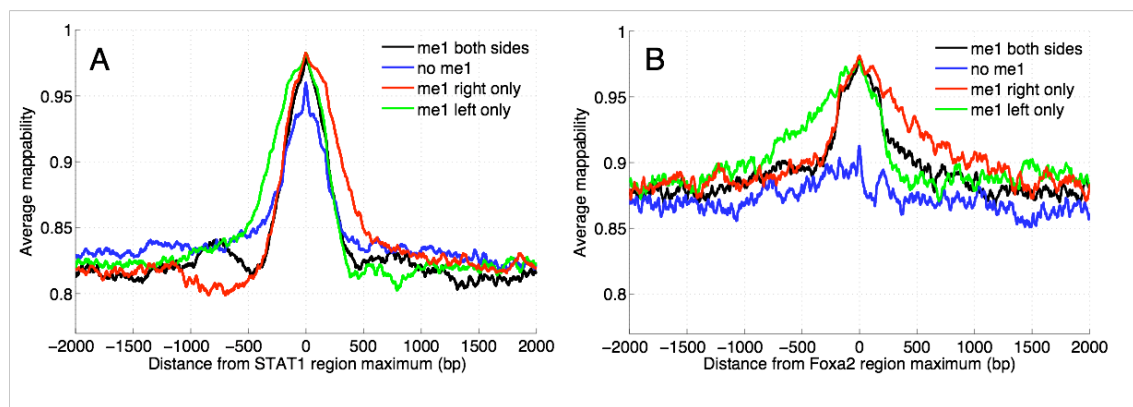

Fig. MP1   Average mappability profiles around distal (A) STAT1 sites in stimulated HeLa cells and (B) Foxa2 sites in mouse adult liver peaks. Data are shown only for TF sites that were associated with H3K4me1 only, or with neither H3K4me1 nor me3. TF sites with associated H3K4me1 were divided into those with the expected flanking modifications ('both sides'), and with the modification present on only one side ('left only', 'right only'). 'No me1' means that neither H3K4me1 nor me3 were associated with the TF site. Mappability profiles were calculated from data resources that characterized the fraction of unique 27mers at each base pair in hg18 and mm8 reference genome sequences.

PT Effect of predicted TSSs and CAGE tags on proportions of H3K4me1/me3 patterns for distal and proximal TF sites

In previous work involving ChIP-seq profiles around distal DHS, UCSC known genes were considered to be an incomplete set of promoters, and input sets of DHS were filtered against Pol II sites or expression evidence that was distal to TSSs of known genes (Boyle et al. 2008; Wang et al. 2008; Barski et al. 2007). To facilitate comparing our results to this work, we tested whether the proportions of histone modifications associated with our distal STAT1 (IFNG) and Foxa2 sites were sensitive to similar filtering.

PT1 STAT1 in IFNG-stimulated HeLa cells: Filtering against predicted TSSs and CAGE tags

Having defined proximal and distal STAT1 (IFNG) sites using known gene TSSs, we compared the 53.0k distal sites against the combined set of AceView, Genscan and Switchgear TSS predictions from the UCSC hg18 genome browser.

Table PT1 Number of STAT1 sites with different histone modification patterns as a function of predicted TSSs and CAGE tags. Prox (KG) means a STAT1 region that is proximal to UCSC hg18 known gene transcripts. Prox(DEE) means a STAT1 region that is distal to known genes but proximal to a least one of AceView predictions, Genscan predictions or SwitchGear TSSs. Dist(DEE) means distal TSSs of all known genes, as well as AceView, Genscan and SwitchGear TSSs. Dist(CAGE) means distal to all these TSSs and CAGE tag clusters from Abeel et al. (2008). DEE filtering reclassified 33.9% of 52996 STAT1 distal sites as proximal.

	me3 only	me1+me3	me1 only	no mark	Total
Prox (KG)	2981	11645	2013	656	17295
Prox (DEE)	638	7795	6775	2780	17988
Dist (DEE)	1075	13677	14246	6010	35008
Dist (CAGE)	1000	12523	13255	5600	32378

Table PT2 As Table PT4, but for the fraction of Stat1 sites in the different pattern when including more expression evidence.

	me3 only	me1+me3	me1 only	no mark	Total
Prox (KG)	17.2%	67.3%	11.6%	3.8%	100%
Prox (DEE)	3.6%	43.3%	37.7%	15.5%	100%
Dist (DEE)	3.1%	39.1%	40.7%	17.2%	100%
Dist (CAGE)	3.1%	38.7%	40.9%	17.3%	100%

H3K4me1 was associated with 25778/26778 = 96.3% of the marked 'Dist(CAGE)' binding sites, and H3K4me3 was associated with 13523/26778 = 50.5% of these sites.

The ratio of H3K4me1 to H3K4me3 associated with sites distal to known genes but proximal to AceView, Genscan and SwitchGear TSS's (Prox DEE) was 1.7 (81%/46.9%), for STAT1 sites distal to DEE was 1.9 (79.8%/42.2%).
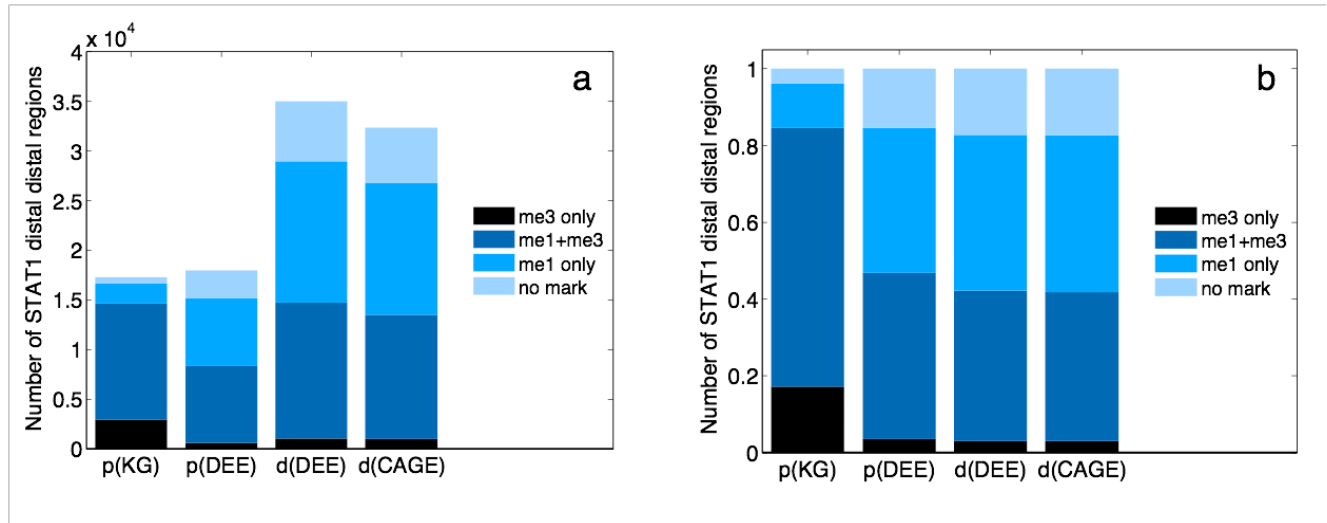

Figure PT1. Number (a) and fraction (b) of STAT1 sites with different histone modification patterns as a function of expression evidence. See caption for Table FE4.

PT2 Foxa2 in mouse adult liver: Filtering against predicted TSSs

Table PT3 Number of Foxa2 sites with different associated histone modification patterns as a function of expression evidence. Prox (KG) means a Foxa2 region that is proximal to a UCSC mm8 known gene transcript TSS. Prox(DEE) means a Foxa2 region that is distal to known genes but proximal to a TSS from either an SIBgene or a Genscan prediction. Dist(DEE) means distal to all known genes, as well as SIBgene and Genscan predictions. DEE filtering reclassified 2384 (24.9%) of 9566 Foxa2 distal sites as proximal.

	me3 only	me1+me3	me1 only	no mark	Total
Prox (KG)	165	873	310	56	1404
Prox (DEE)	48	297	1662	377	2384
Dist (DEE)	22	295	5450	1415	7182

Table PT4   As Table PT3, but for the fraction of Foxa2 sites.

	me3 only	me1+me3	me1 only	no mark	Total
Prox (KG)	11.8%	62.2%	22.1%	4.0%	100%
Prox (DEE)	2.0%	12.5%	69.7%	15.8%	100%
Dist (DEE)	0.3%	4.1%	75.9%	19.7%	100%

H3K4me1 was associated with 5745/5767 = 99.6% of the marked 'Dist(DEE)' binding sites, and H3K4me3 was associated with 317/5767 = 0.5% of these sites.

The ratio of H3K4me1 to H3K4me3 associated with Foxa2 sites distal to known genes but proximal to SIB gene and Genscan TSSs (Prox DEE) was 5.7 (82.2%/14.5%), while for Foxa2 sites distal to these TSSs it was 18.2 (80.0%/4.4%).
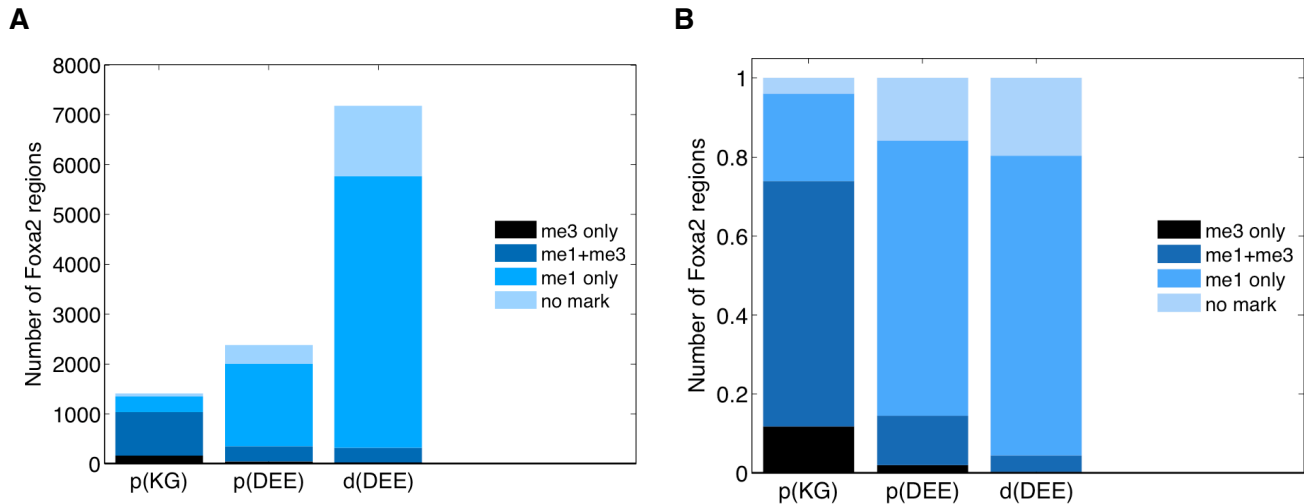
A

B

Figure PT2.   Number (A) and fraction (B) of Foxa2 sites associated with different histone modification patterns as a function of expression evidence. See caption for Table FE9.

FD   Relationship of FDR threshold to proportions of H3K4me1/me3 patterns for distal and proximal TF sites

FD1   HeLa (IFNG)

FD1.1   Region height distributions

Fig. FD1 shows cumulative height distributions of STAT, and STAT1-associated H3K4me1 and H3K4me3 in IFNG-stimulated HeLa cells.
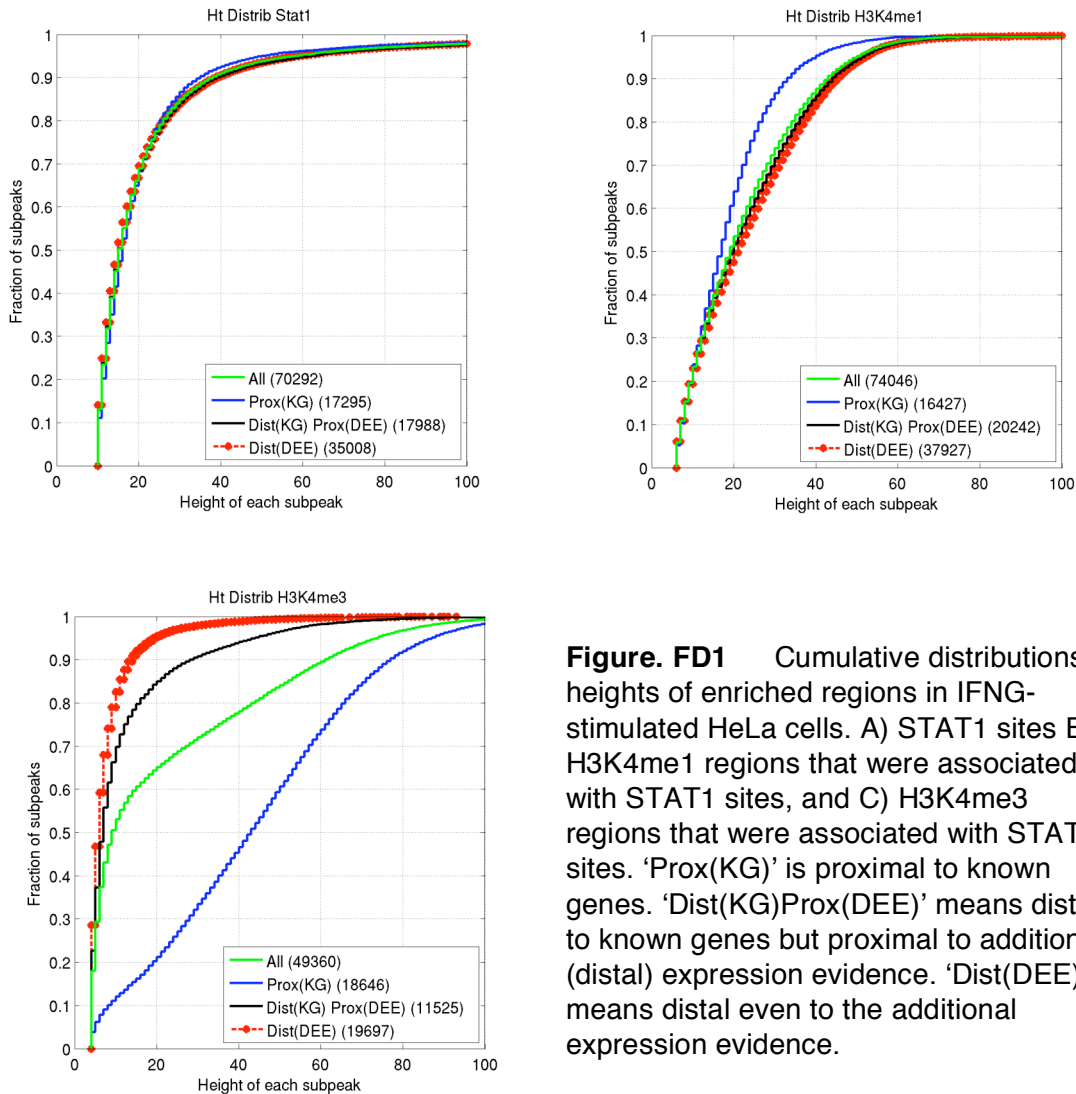

Figure. FD1   Cumulative distributions of heights of enriched regions in IFNG-stimulated HeLa cells. A) STAT1 sites B) H3K4me1 regions that were associated with STAT1 sites, and C) H3K4me3 regions that were associated with STAT1 sites. 'Prox(KG)' is proximal to known genes. 'Dist(KG)Prox(DEE)' means distal to known genes but proximal to additional (distal) expression evidence. 'Dist(DEE)' means distal even to the additional expression evidence.

FD1.2   Effect of FDR threshold stringency

Table FD1   Relationship between estimated FDR and XSET height threshold of H3K4me1 and H3K4me3 profiles in IFNG-stimulated HeLa cells. The default height threshold corresponds to FDR ≈ 0.01, and the more stringent thresholds of FDR ≈ $3 \cdot 10^{-3}$ and ≈ $3 \cdot 10^{-4}$ are marked as '+1' and '+2' respectively.

Height threshold	FDR, H3K4me1	FDR, H3K4me3
1	0.859	0.951
2	0.261	0.131
3	0.170	0.062
4	0.085	**0.014 (default)**
5	0.034	**0.0021 (+1)**
6	**0.011 (default)**	**2.9E-4 (+2)**
7	**0.0032 (+1)**	3.40E-05
8	8.40E-04	3.8E-6
9	**2.1E-4 (+2)**	3.80E-07
10	5.0E-5	4.20E-08
11	1.10E-05	4.40E-09
12	2.50E-06	3.60E-10

Table FD2   Number of STAT1 sites with different H3K4me1/me3 patterns as a function of the profile height threshold for H3K4me1 and H3K4me3 profiles.

Prox/distal	Appr FDR	me3 only	me1+me3	me1 only	no mark	Total
Proximal	**1e-2**	2981	11645	2013	656	17295
	3e-3	3210	11079	2214	792	17295
	3e-4	3799	10271	2198	1027	17295
Distal	**1e-2**	1713	21472	21021	8790	52996
	3e-3	1571	17131	23753	10541	52996
	3e-4	1597	13760	24424	13215	52996

Table FD3   As table FD2, but for the percent of STAT1 sites proximal or distal.

Prox/distal	Appr FDR	me3 only	me1+me3	me1 only	no mark	Total
Proximal	**1e-2**	17.2%	67.3%	11.6%	3.8%	100%
	3e-3	18.6%	64.1%	12.8%	4.6%	100%
	3e-4	22.0%	59.4%	12.7%	5.9%	100%
Distal	**1e-2**	3.2%	40.5%	39.7%	16.6%	100%
	3e-3	3.0%	32.3%	44.8%	19.9%	100%
	3e-4	3.0%	26.0%	46.1%	24.9%	100%

At FDR ~ 0.01, 43.7% of distal STAT1 sites were associated with H3K4me1 and 80.1% with H3K4me1 (equivalent to an me1:me3 ratio of 1.8), while at FDR ~ 0.0001, 29.0% of distal STAT1 sites were associated with H3K4me1 and 73% with H3K4me1 (equivalent to an me1:me3 ratio of 2.5).
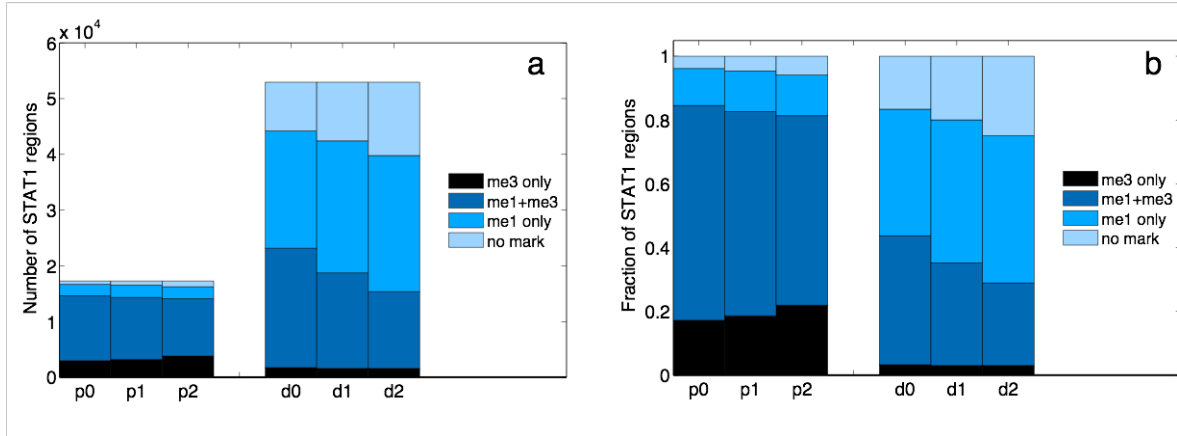

Figure FD2. Number (a) and fraction (b) of proximal and distal STAT that were associated with different combinations of H3K4me1 and me3. p0 and d0: modification profiles were thresholded with an FDR ≈ 0.01. p1, d1, and p2, d2: modification profiles were segmented using height thresholds corresponding to FDR ≈ $3 \cdot 10^{-3}$ and ≈ $3 \cdot 10^{-4}$ (Table 1 above).

FD2   Mouse adult liver

FD2.1   Region height distributions

Fig. FD3 shows cumulative height distributions of Foxa2, and Foxa2-associated H3K4me1 and H3K4me3 in mouse adult liver.
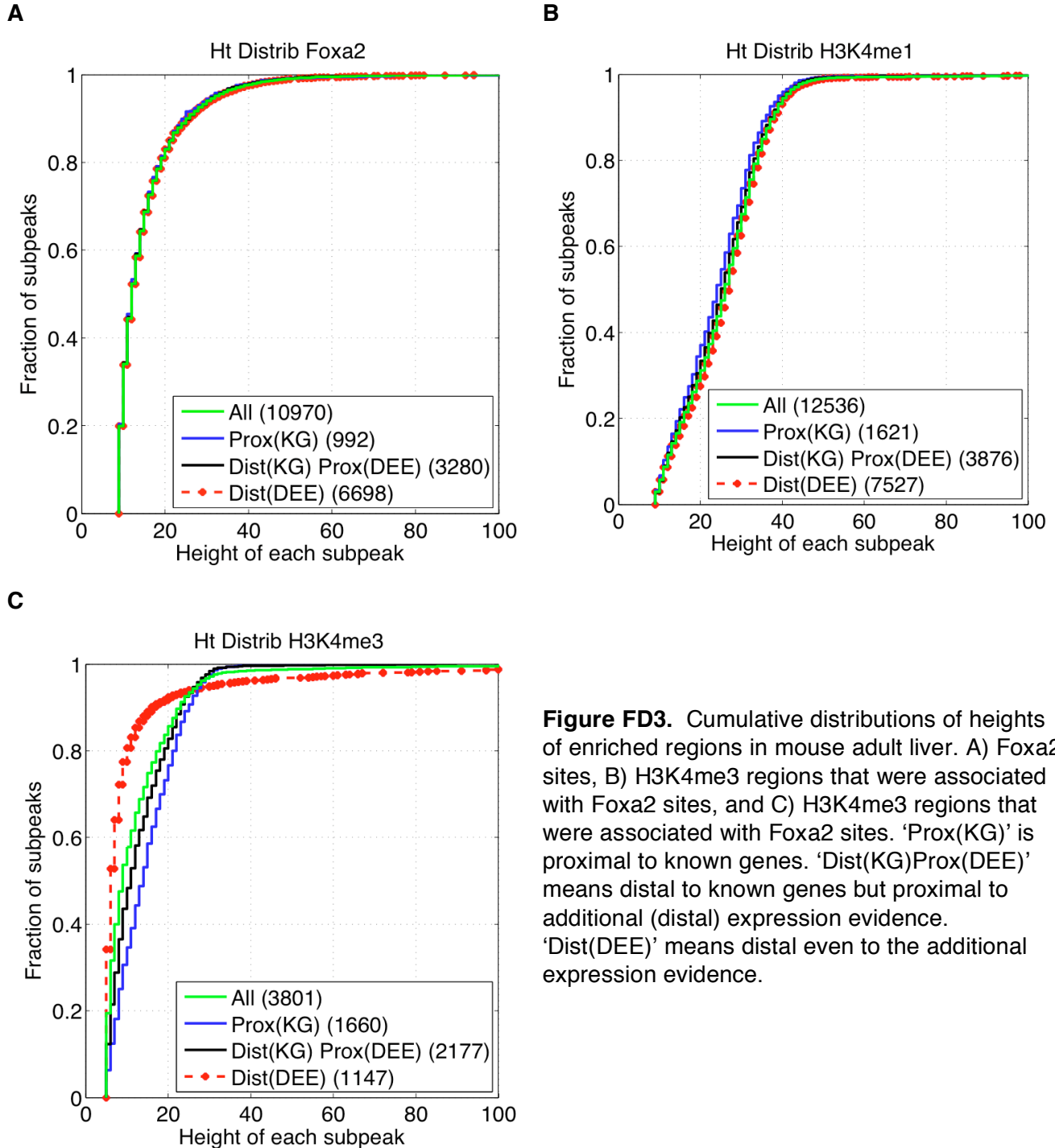
A

B

C

Figure FD3. Cumulative distributions of heights of enriched regions in mouse adult liver. A) Foxa2 sites, B) H3K4me3 regions that were associated with Foxa2 sites, and C) H3K4me3 regions that were associated with Foxa2 sites. 'Prox(KG)' is proximal to known genes. 'Dist(KG)Prox(DEE)' means distal to known genes but proximal to additional (distal) expression evidence. 'Dist(DEE)' means distal even to the additional expression evidence.

FD2.2   Effect of FDR threshold stringency

Table FD4   Relationship between estimated FDR and XSET height threshold of H3K4me1 and H3K4me3 profiles in IFNG-stimulated HeLa cells. The default height threshold corresponds to FDR $\approx$ 0.01, and the more stringent thresholds of FDR $\approx 1\cdot10^{-3}$ and $\approx 1\cdot10^{-4}$ are marked as '+1' and '+2' respectively.

| Height threshold	FDR, H3K4me1	FDR, H3K4me1
1	0.844	0.970
2	0.411	0.152
3	0.341	0.093
4	0.272	0.026
5	0.185	**0.0043 (default)**
6	0.107	**5.78E-4 (+1)**
7	0.051	**6.94E-5 (+2)**
8	0.022	7.78E-06
9	0.0083	8.62E-07
10	**0.0033 (default)**	8.45E-08
11	**0.0011 (+1)**	9.40E-09
12	3.89E-04	9.65E-10
13	**1.26E-4 (+2)**	1.34E-10
14	3.99E-05	1.20E-11
15	1.34E-05	9.50E-13
16	4.06E-06	2.86E-14

Table FD5   Number of Foxa2 sites with different H3K4me1/me3 patterns as a function of the height threshold for H3K4me1 and H3K4me3 profiles. 'Prox 10e-2' means proximal sites with the default FDR height threshold. 'Prox 1', 2, 3 and 4 mean thresholds that are 1, 2, 3 and 4 higher than the default.

Prox/distal	Appr FDR	me3 only	me1+me3	me1 only	no mark	Total
Proximal	**1e-2**	165	873	310	56	1404
	1e-3	181	822	335	66	1404
	1e-4	202	761	358	83	1404
Distal	**1e-2**	185	1573	6131	1677	9566
	1e-3	161	1116	6377	1912	9566
	1e-4	155	829	6464	2118	9566

Table FD6 As table FE5, but for the fraction of Foxa2 sites.

Prox/distal	Appr FDR	me3 only	me1+me3	me1 only	no mark	Total
Proximal	**1e-2**	11.8%	62.2%	22.1%	4.0%	100%
	1e-3	12.9%	58.6%	23.9%	4.7%	100%
	1e-4	14.4%	54.2%	25.5%	5.9%	100%
Distal	**1e-2**	1.9%	16.4%	64.1%	17.5%	100%
	1e-3	1.7%	11.7%	66.7%	20.0%	100%
	1e-4	1.6%	8.7%	67.6%	22.1%	100%

At FDR ~ 0.01, 18.3% of distal Foxa2 sites were associated with H3K4me1 and 80.5% with H3K4me1 (equivalent to an me1:me3 ratio of 4.4), while at FDR ~ 0.0001, 10.3% of distal Foxa2 sites were associated with H3K4me1 and 76.3% with H3K4me1 (equivalent to an me1:me3 ratio of 7.4).
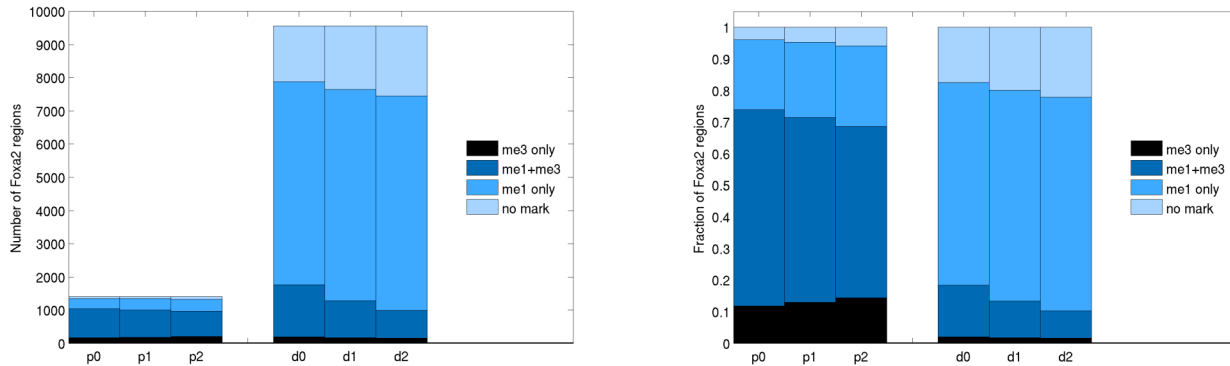
A   **B**

Figure FD4. Number (A) and fraction (B) of proximal and distal Foxa2 associated with different combinations of H3K4me1 and me3 as a function of profile threshold. p0 and d0: modification profiles were thresholded with an FDR ≈ 0.01. p1, d1, and p2, d2: modification profiles were segmented using height thresholds corresponding to FDR ≈ $1 \cdot 10^{-3}$ and ≈ $1 \cdot 10^{-4}$ (Table FE6 above).

TF   Distal H3K4me1-flanked genomic regions were enriched in known functional TF binding site sequences

TF1   Selecting regions flanked by H3K4me1

We first analyzed widths of regions that contained an enriched location for STAT1 or FoxA2 and had an enriched H3K4me1 region on both sides (**Fig. TF1**). For both IFNG-stimulated HeLa cells and mouse adult liver, distal regions were more compact than proximal, consistent with **Figs. 4** and **5**. The dominant spacing between the nearest H3K4me1 region maximum on either side of a STAT1 or Foxa2 region maximum was ~400 bp and 500 bp respectively.
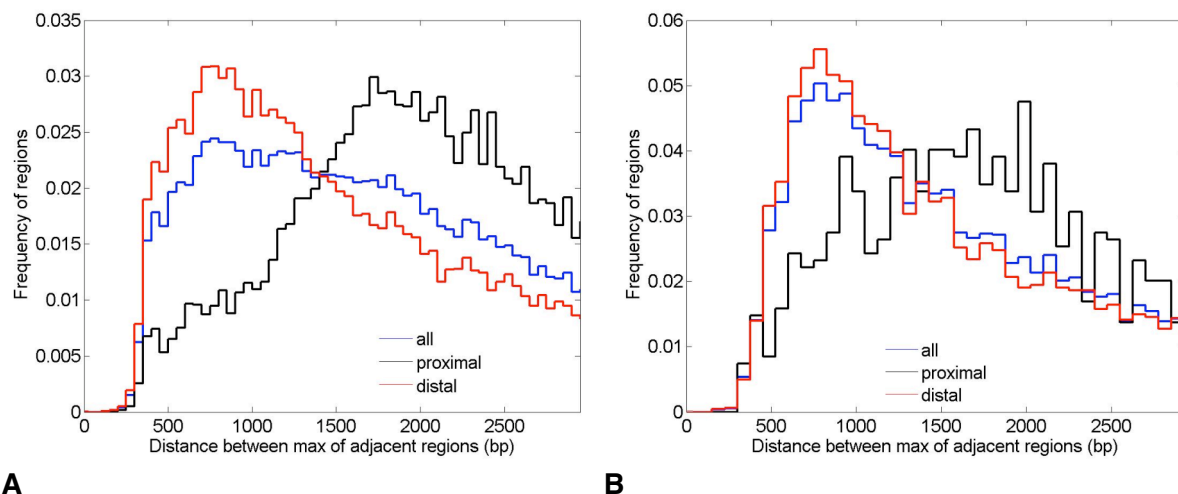

A   B

Figure TF1. Frequency distributions of widths of regions defined maxima of two adjacent H3K4me1 enriched regions that flanked enriched STAT1 (A) and FoxA2 (B) sites. The blue line corresponds to all TF locations, and red and black lines for proximal and distal TFs enriched locations respectively, using a distance of ±2.5 kb from the TSS of a UCSC hg18 or mm8 known gene.

Next we analyzed all genomic locations that were flanked by H3K4me1-enriched regions. The dominant distance between adjacent H3K4me1 region maxima was between ~400 and 600 bp in both stimulated HeLa cells and mouse adult liver (Fig. TF2). These length scales were consistent with Fig. TF1. From Figs. TF1 and TF2, we set criteria for selecting a global subset of H3K4me1-flanked regions: the distance between adjacent H3K4me1 region maxima should be between 200bp and 1000bp. Fig. TF3 shows examples of regions selected with this criterion.
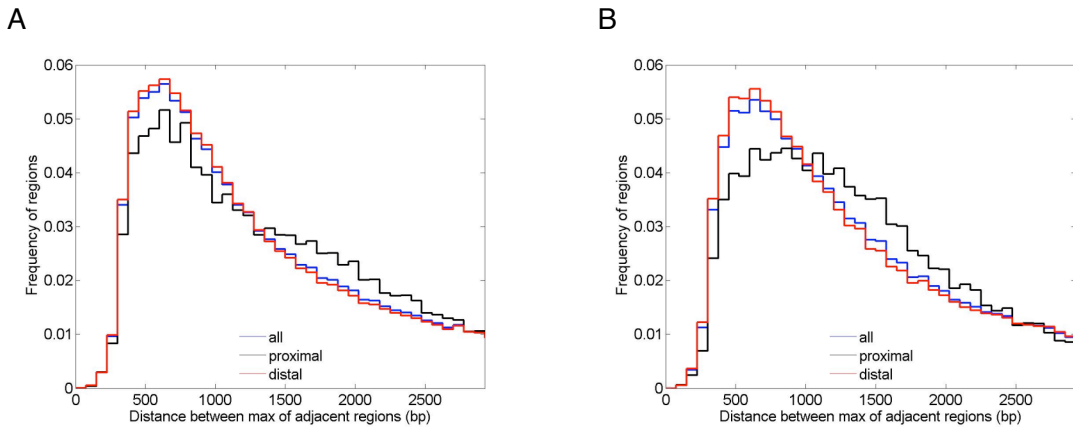
Figure TF2. Distances between locations of maxima of adjacent HeK4me1 regions in A) in stimulated HeLa cells and B) in mouse adult liver.
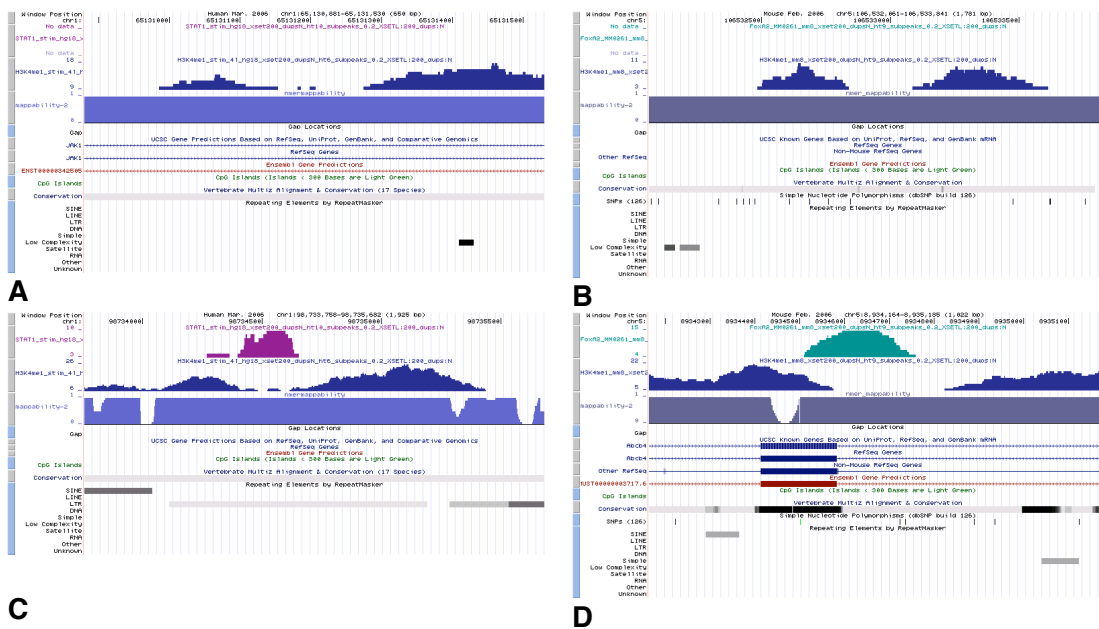

Figure TF3. Examples of regions with flanking H3K4me1 that were identified by the separation criteria for adjacent H3K4me1 of between 200bp and 1000 bp. (A,C for HeLa cells and B,D for mouse adult liver). All cases are for regions flanked by H3K4me1 distal from UCSC known genes. Mappability profiles shown were calculated for 27mers. FDR thresholds for these profiles are given in Table 1.

TF2　　TFBS sequence scans

TF2.1　　Identifying enriched TFBS models

Applying thresholds of 2.0 on distributions of enrichment scores and 3.0 on Z-scores, then requiring that an enriched model have exact sequence matches on at least 50 regions, returned TFBS models from TRANSFAC v9.3 that were enriched in regions flanked by H3K4me1 in HeLa (IFNG) and mouse liver respectively (Supplemental Figs. TF4 to TF6; Supplemental Table TF1).
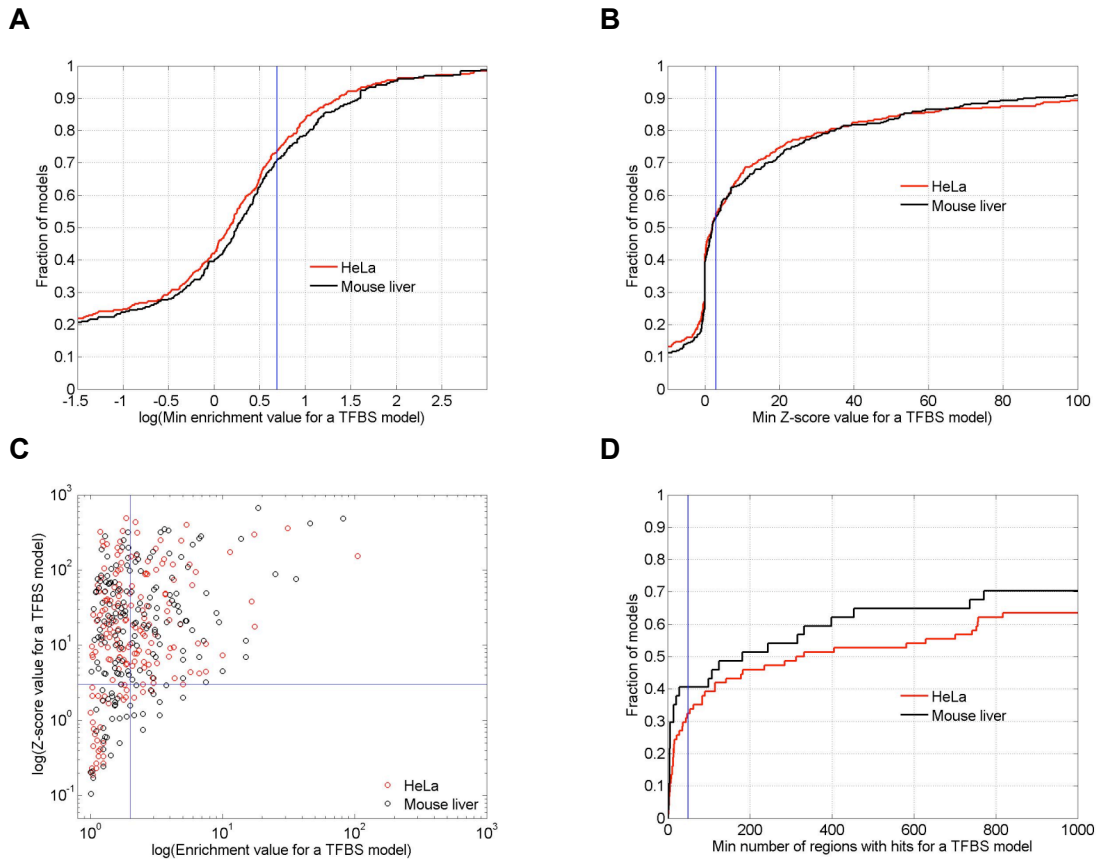

Figure TF4.　For distal H3K4me1-flanked regions, results for scans with site sequences from 319 TRANSFAC v9.3 models were similar for HeLa and mouse liver. Results show exact matches to at least one known transcription factor binding site sequence in a model. A) Cumulative distribution for the natural logarithm of the enrichment score (i.e. ratio of total matches in regions to average of total matches for randomized regions) with blue line showing the threshold of 2. B) Cumulative distribution for the Z-score (i.e. the difference between the number of exact matches for genomic regions and average for randomized sequences, measured in standard deviations for the randomized results) with the blue line indicating the threshold of 3. C) Scatterplot of enrichment vs. Z-score, with thresholds. D) Number of regions with at least one exact match for a TFBS model, with the blue line showing models with matches on 50 regions (blue line).
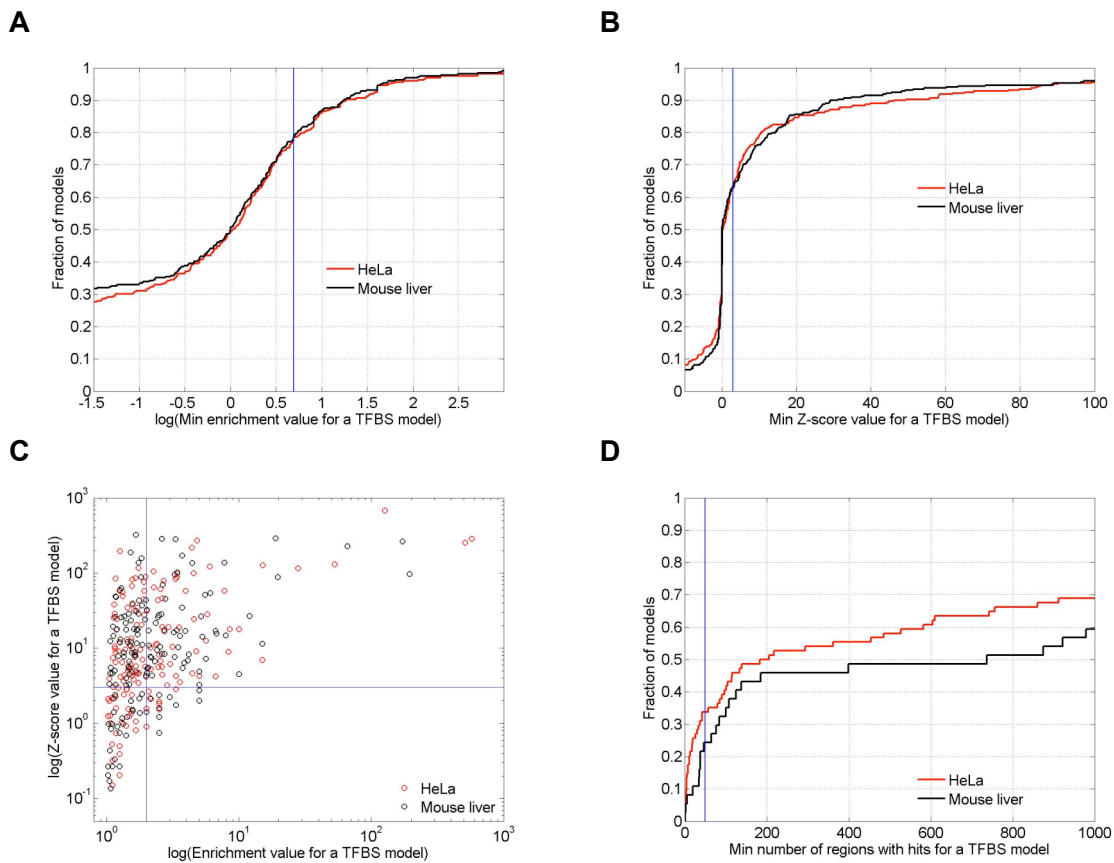
Figure TF5. As Fig. TF4, but for proximal H3K4me1-flanked regions.
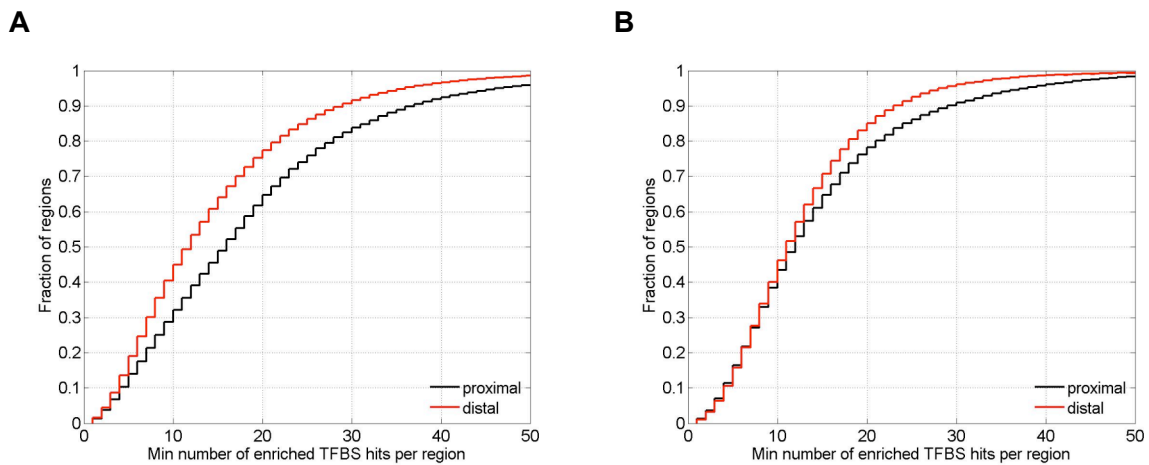

Figure TF6. Cumulative distributions of the number of globally enriched TFBS models with at least one exact sequence match in proximal and distal flanked regions. (A) HeLa, IFNG-stimulated. (B) Adult mouse liver.

Table TF1   Number of enriched models

Regions from	HeLa (IFNG)	Mouse adult liver
Proximal	35	34
Distal	51	50
All	52	51

Table TF2. Summary of TFBS sequence scan results for H3K4me1-flanked regions. Regions were classified as distal and proximal using a distance of ±2.5 kb between the region centre and TSSs of UCSC hg18 or mm8 known genes.

Regions scanned/with model matches	HeLa (IFNG)		Mouse liver	
Number of flanked regions scanned	**90,121**	**Na**	**63,708**	**Na**
All regions with a match for ≥ 1 model	80,736	90%	58,582	92%
≥ 3 models	58,370	65%	42,567	67%
Distal flanked regions	**77,924**	**Na**	**54,058**	**Na**
Distal regions with a match for ≥ 1 model	77,582	99.6%	53,824	99.6%
≥ 3 models	74,101	95%	52,121	96.4%
Proximal flanked regions	12,197	Na	9,727	Na
Proximal regions with a match for ≥ 1 model	12,136	99.5%	9,681	99.5%
≥ 3 models	11,648	95.5%	9,334	96%

CT   Proportions of histone modification patterns associated with STAT1 in IFNG-stimulated HeLa cells were insensitive to STAT1-CTCF association

Fig. CT1 shows cumulative distributions of the distance from a STAT1 (IFNG) region maximum to the maximum of the closest CTCF enriched region or conserved CTCF motif. From the distributions we chose a distance of 150 bp between Stat1 and CTCF maximum as a criterion for considering that STAT1 was associated with CTCF.

Fig. CT2 shows proportions of different histone modification patterns that were associated with a STAT1 (IFNG) site for all STAT1 sites, and for the subset of STAT1 sites that was not associated with CTCF.
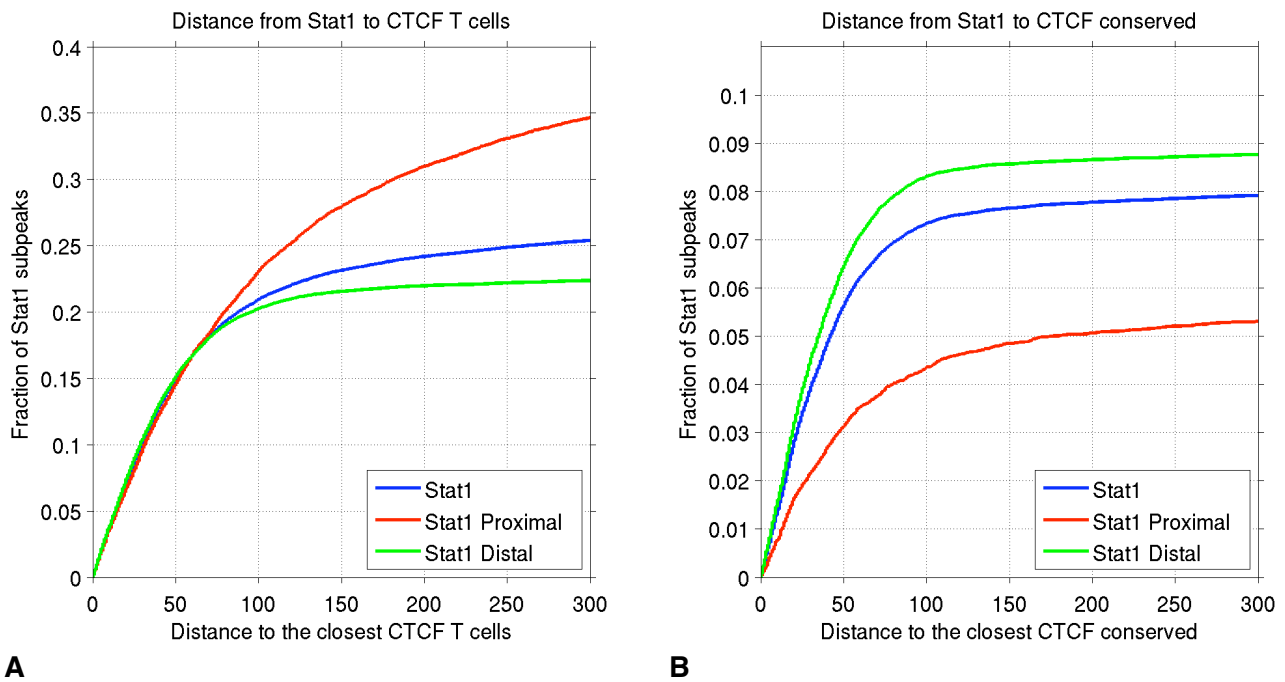

Figure CT1.   Distance between a STAT1 (IFNG) region maximum and the closest CTCF region maximum or conserved CTCF motif centre. A) Region maxima for CTCF-enriched regions from a FindPeaks 2.0 profile from Barski et al. (2007) T cell read data, B) CTCF conserved human motif locations from Kim et al. (2007).
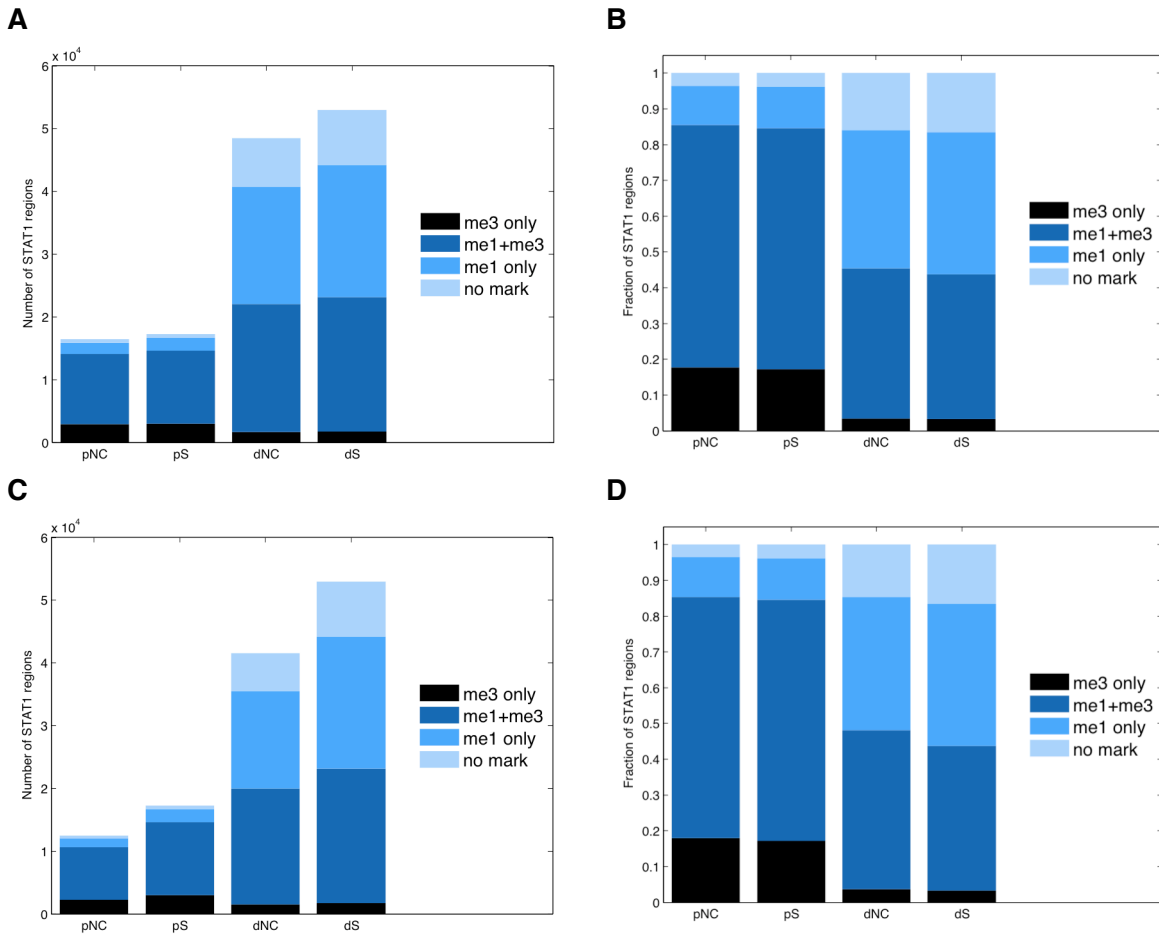
Figure CT2. Combinations of histone modifications CTCF-associated STAT1 sites in stimulated HeLa cells. (A,B) STAT1 associated (or not) with CTCF conserved motifs (Kim et al. 2007). (C,D) STAT1 associated (or not) with CTCF enriched regions from T cell data (Barski et al. 2007). pNC: proximal STAT1 sites that were not associated with CTCF. pS: all proximal STAT1 sites. dNC: distal STAT1 sites that were not associated with CTCF. dS: all distal STAT1 sites.

Table CT1 Number of STAT1 (IFNG) sites associated with different H3K4me1/me3 patterns, for all STAT1 sites and those not associated with a conserved CTCF site from (Kim 2007). 'p No CTCF' means a proximal STAT1 region that was not associated with CTCF, 'p Stat1' means all proximal STAT1 sites, 'd No CTCF' means distal STAT1 sites that were not associated with CTCF, and 'd STAT1' means all distal STAT1 sites.

	me3 only	me1+me3	me1 only	no mark	Total
p No CTCF	2922	11150	1799	584	16455
p STAT1	2981	11645	2013	656	17295
d No CTCF	1646	20373	18743	7693	48455
d STAT1	1713	21472	21021	8790	52996

Table CT2 As Table CT1 but with percentages. See caption of Table CT1.

	me3 only	me1+me3	me1 only	no mark	Total
p No CTCF	17.7%	67.8%	10.9%	3.6%	100%
p STAT1	17.2%	67.3%	11.6%	3.8%	100%
d No CTCF	3.4%	42.1%	38.7%	15.9%	100%
d STAT1	3.2%	40.5%	39.7%	16.6%	100%

Table CT3 As Table CT1 but using enriched CTCF sites calculated from (Barski et al. 2007) T cell read data. See caption of Table CT1.

	Me3 only	me1+me3	me1 only	no mark	Total
p No CTCF	2233	8391	1397	427	12448
p STAT1	2981	11645	2013	656	17295
d No CTCF	1524	18454	15527	6063	41568
d STAT1	1713	21472	21021	8790	52996

Table CT4 As Table CT3 but with percentages. See caption of Table CT1.

	me3 only	me1+me3	me1 only	no mark	Total
p No CTCF	17.9%	67.4%	11.2%	3.4%	100%
p STAT1	17.2%	67.3%	11.6%	3.8%	100%
d No CTCF	3.7%	44.4%	37.4%	14.6%	100%
d STAT1	3.2%	40.5%	39.7%	16.6%	100%

For mouse liver, only 0.8% of the 11.0k enriched Foxa2 sites were associated with the 6.6k conserved CTCF motifs reported by Kim et al. (2007). As well, while an association distance of 250 bp identified 5.9% of 10970 Foxa2 peaks as associated with Chen et al (2008)'s ~64k CTCF enriched regions from mouse ES cells, the association distance distribution showed much weaker evidence that these CTCF locations were spatially coordinated with our Foxa2 sites than we had seen above for STAT1 (Supplemental Fig. CT3). Given the low spatial association indicated, we did not compare histone modifications for Foxa2 sites that were or were not associated with CTCF data.
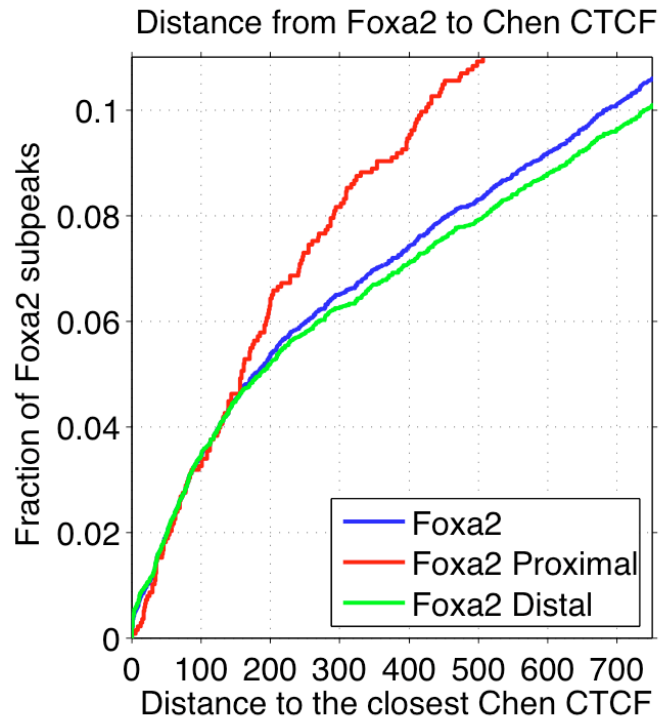

Figure CT3. Distance between a Foxa2 region maximum and the closest CTCF location from Chen et al. (2008)'s mouse ES data.

Figs. CT4 and 5 show coverage profiles for H3K4me1 and H3K4me3 around CTCF locations from four published datasets. Profiles were generally similar to those seen for STAT1 sites in stimulated HeLa cells and for Foxa2 sites in mouse adult liver (Fig. 3).

Figure CT4. Coverage profiles for H3K4me1 and me3 in stimulated HeLa cells around 12,799 conserved CTCF motifs from (Kim et al. 2007), and 38,423 CTCF-enriched regions identified by FindPeaks 2.0 with FDR ~0.01 with read data from Barski et al. (2007) data in human T-Cells.
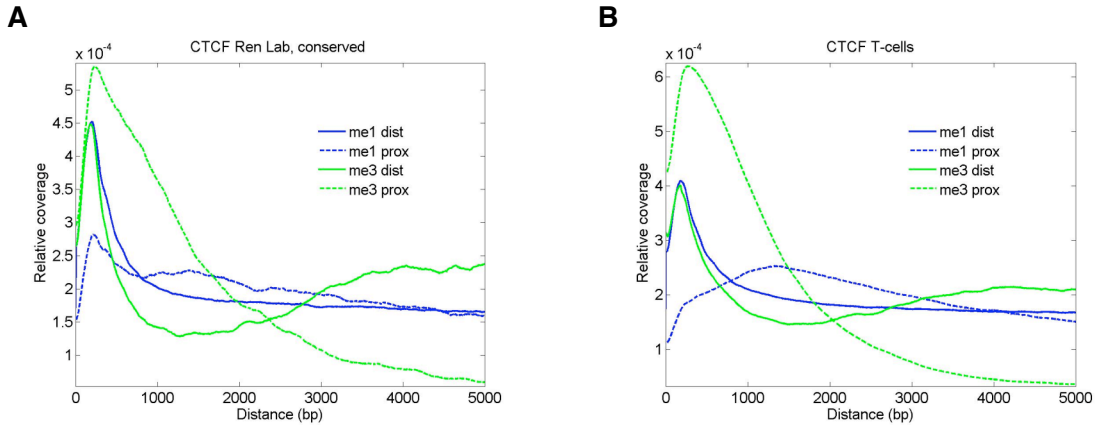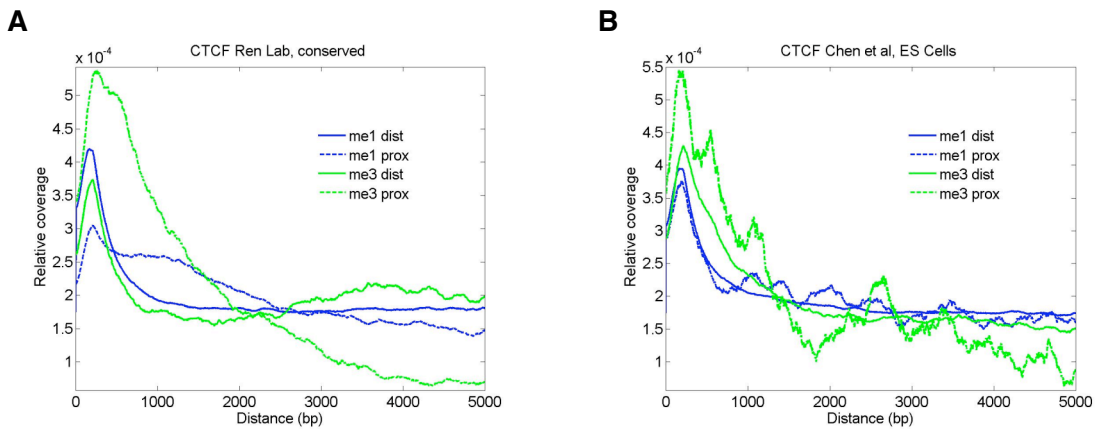

Figure CT5. Coverage profiles for H3K4me1 and me3 in mouse adult liver around 6,573 conserved CTCF motifs from Kim et al. (2007) and 63,542 ChIP-seq locations reported for mouse ES cells (Chen et al. 2008).

MC   Individual histone modifications were concordant between stimulated and unstimulated HeLa cells.

We used two approaches to estimate concordance rates. For the first approach, we generated distributions of distances between locations of enriched region maxima in IFNG-stimulated and unstimulated HeLa cells, for a range of FDR thresholds (Fig. MC1A). As the distribution curves did not indicate a threshold distance, we used the median width of stimulated regions as a threshold. For H3K4me1, this was 575 bp (Table 1). Applying this threshold to the association distance distributions returned H3K4me1 concordance rates of between ~70% and ~85% for the IFNG-stimulated profile for the default FDR threshold and the upper ~50% of region heights.

For the second approach we calculated the distribution overlap for stimulated and unstimulated regions directly, and generated a random expectation by shuffling the genomic coordinates of one of the sets of regions. We plotted the distances as cumulative distributions, from which, given a fractional overlap threshold, the concordance rate was the vertical distance from the cumulative distribution to 1.0. For H3K4me1 (Fig. MC2), a 50% overlap threshold indicated that 67.4% of 301k H3K4me1 regions in stimulated cells (re = 6.2%) overlapped H3K4me1 regions in unstimulated cells. Considered relative to unstimulated cells, 72.0% of H3K4me1 regions in unstimulated cells (re = 7.5%) overlapped H3K4me1 regions in stimulated cells.

Applying the first approach for H3K4me3, and using the ~400 bp median region width for stimulated regions as a threshold, the association distance distributions returned concordance rates of between ~56% and ~82% for the IFNG-stimulated profile for the default FDR threshold and the upper ~50% of region heights. Applying the second approach (Fig. MC2), a 50% overlap threshold indicated that 54.7% of 76.0 H3K4me3 regions in stimulated cells (re = 0.9%) overlapped H3K4me3 regions in unstimulated cells. Considered for unstimulated cells, 77.4% of 54.5k H3K4me3 regions in unstimulated cells (re = 1.5%) overlapped H3K4me3 regions in stimulated cells.

Applying the first approach for STAT1, and using the ~400 bp median region width for stimulated sites as a threshold, the association distance distributions returned concordance rates of between ~27% and ~39% for the IFNG-stimulated profile for the default FDR threshold and the upper ~50% of region heights. For STAT1 (Fig. MC3), a 50% overlap threshold indicated that 25.7% of 70.3k H3K4me3 regions in stimulated cells (re = 0.1%) overlapped STAT1 sites in unstimulated cells. Considered for unstimulated cells, 86.4% of 20.6k STAT1 sites in unstimulated cells (re = 1.3%) overlapped H3K4me3 regions in stimulated cells.
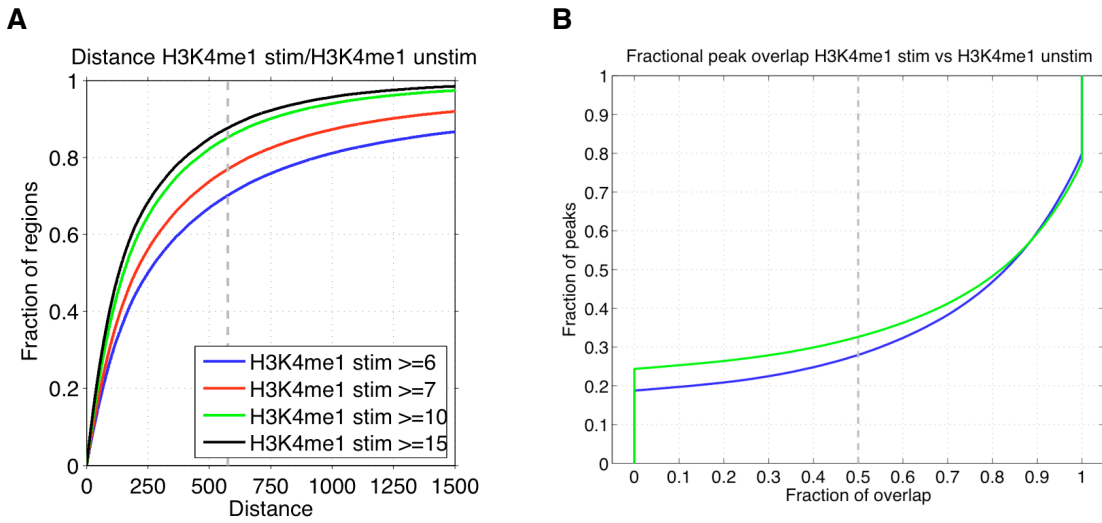
A

B

Figure MC1. Concordance parameters for H3K4me1. (A) Cumulative distribution of minimum distances between locations of a maximum region height in stimulated and unstimulated HeLa cells. Increasing profile height thresholds include all (FDR ≈ 0.01), then the approximate upper top 70%, 50% and 30% of peak heights. (B) Mark concordance for H3K4me1 by region overlap. Fractional overlap of H3K4me1 regions in HeLa cells with (green – 301,492 regions) or without (blue – 270,569 regions) IFNG stimulation. The concordance at 50% fractional overlap is the vertical distance above a blue or green curve.

A

B

Figure MC2. Concordance parameters for H3K4me3. (A) Cumulative distribution of minimum distances between locations of a maximum region height in stimulated and unstimulated HeLa cells. Increasing profile height thresholds include all (FDR ≈ 0.01), then the approximate upper top 70%, 50% and 30% of peak heights. (B) Fractional overlap of H3K4me3 in HeLa cells for 76,061 regions in IFNG-stimulated cells (green) and for 54,487 regions in unstimulated cells (blue). The concordance at 50% fractional overlap is the vertical distance above a blue or green curve.
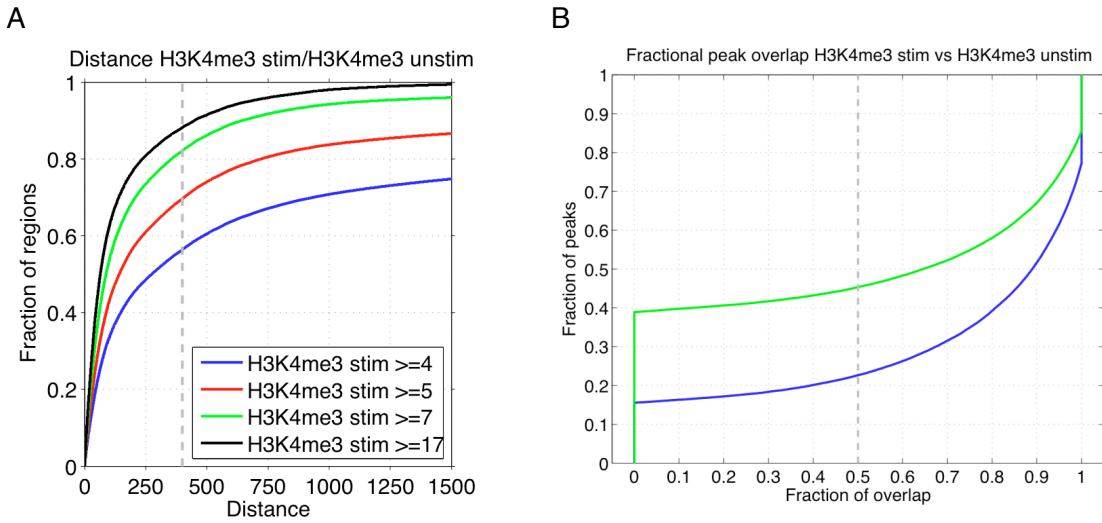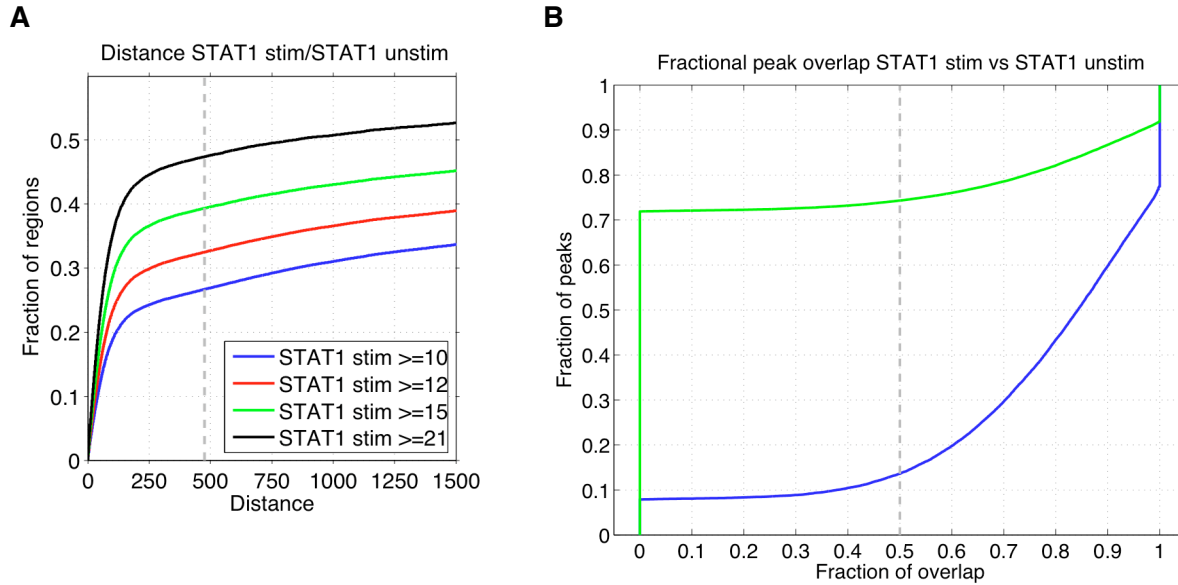
Figure MC3. Concordance parameters for STAT1. (A) Cumulative distribution of minimum distances between locations of a maximum region height in stimulated and unstimulated HeLa cells. Increasing profile height thresholds include all (FDR ≈ 0.01), then the approximate upper top 70%, 50% and 30% of peak heights. (B) Fractional overlap for 70,292 sites in IFNG-stimulated cells (green) and for 20,550 sites in unstimulated cells (blue). The concordance at 50% fractional overlap is the vertical distance above a blue or green curve.

Table MC1. Concordance rates for enriched regions in stimulated HeLa cells as a function of region height threshold. See Figures MC1 to 3.

	Target height percentile	100%	70%	50%	30%
H3K4me1	Actual height percentile	100%	81.7%	53.3%	32.1%
	Height threshold	6	7	10	15
	Association at median distance	70%	77%	85%	88%
H3K4me3	Actual height percentile	100%	73.7%	52.4%	30.6%
	Height threshold	4	5	7	17
	Association at median distance	56%	69%	82%	88%
STAT1	Actual height percentile	100%	76.5%	55.0%	31.1%
	Height threshold	10	12	15	21
	Association at median distance	27%	33%	39%	47%

PC Combinations of H3K4me1 and me3 that were associated with STAT1 in stimulated HeLa cells were present at these locations in unstimulated cells.

Below, the upper pie chart on each page of figures shows the proportion of histone modification patterns for sets of STAT1 binding sites in IFNG-stimulated HeLa cells: PC1. All STAT1 sites; PC2. Distal STAT1 sites; PC3. Proximal STAT1 sites; PC4. STAT1 (IFNG) sites that were enriched in both stimulated cells unstimulated cells; PC5. As (PC4), but for distal STAT1 sites, PC6. STAT1 (IFNG) sites that were enriched in stimulated but not in unstimulated cells, and PC7. As (PC6) but for distal STAT1 sites. On each page, the subsequent four pie charts show the proportion of all four histone modification patterns at these locations in unstimulated cells for each of the four stimulated patterns. When a particular histone modification pattern is highly concordant between stimulated and unstimulated cells, the pattern-specific unstimulated pie chart for that stimulated pattern will be dominated by the stimulated pattern.

A   All Stat1=70292

13%

33%

7%

47%

B   Total me1 stim only=23035

10%

< 1%

8%

82%

C   Total me3 stim only=4694

7%

7%

33%

53%

D   Total me1 stim and me3 stim=33117

1%

5%

21%

73%

E   Total no stim mark=9446

17%

1%

1%

80%
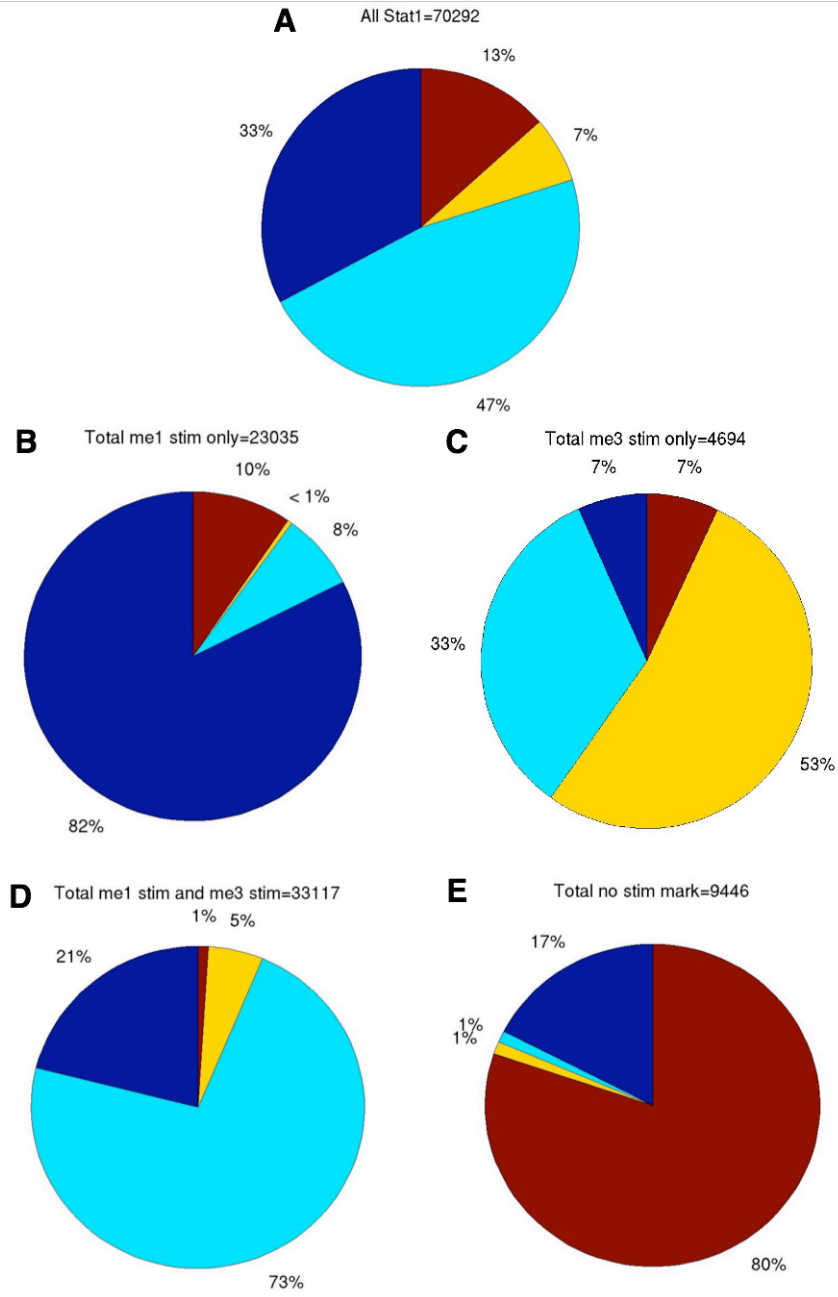
Figure PC1. Pattern concordance for all STAT1 (IFNG) sites.

A Distal Stat1=52996

17%

3%

40%

41%

B Total me1 stim only=21021

10%

< 1%

6%

84%

C Total me3 stim only=1713

16%   16%

30%

38%

D Total me1 stim and me3 stim=21472

1% 3%

29%

66%

E Total no stim mark=8790

18%

< 1%
< 1%

81%
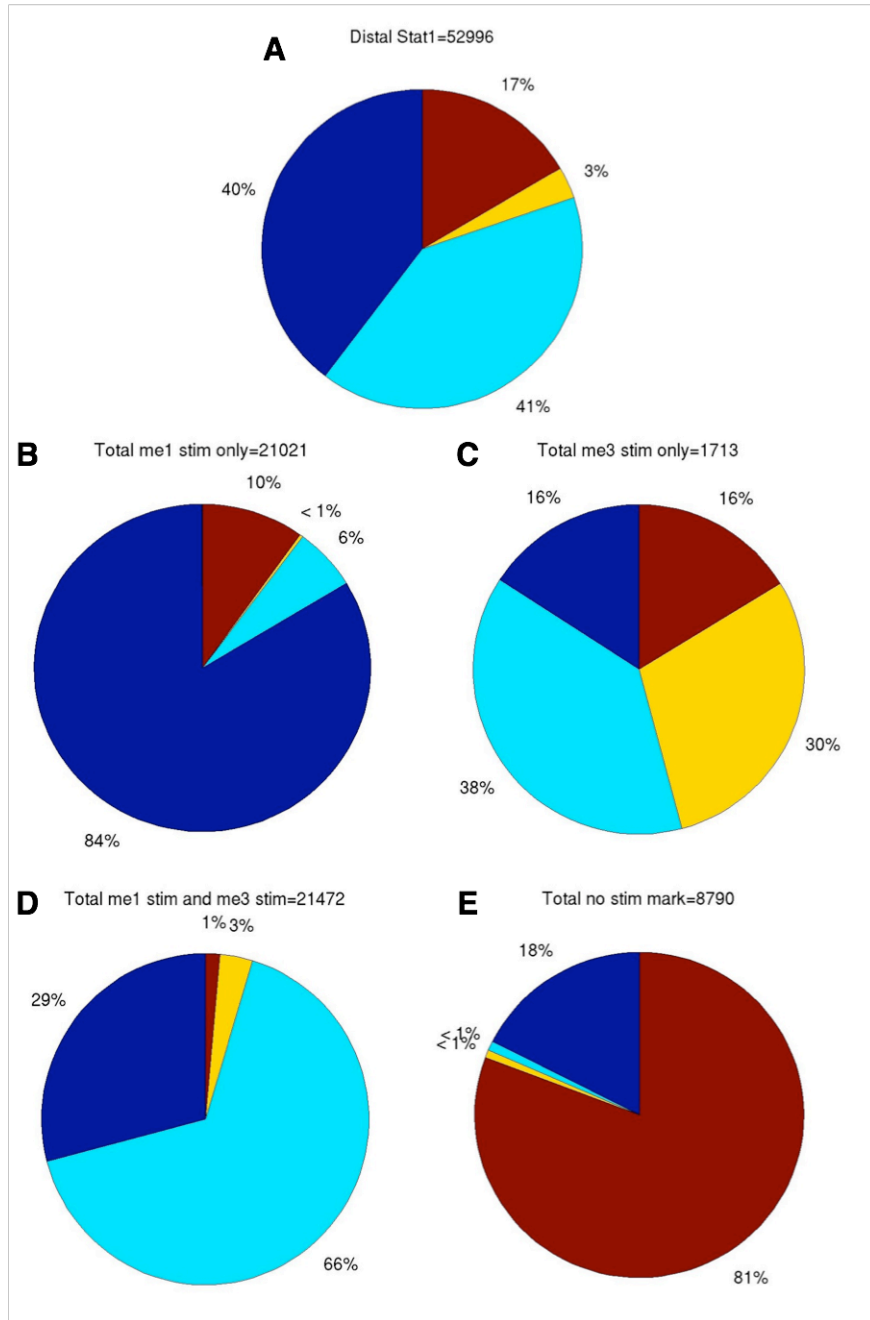
Figure PC2.   Pattern concordance for all distal STAT1 (IFNG) sites.

A   Proximal Stat1=17296

12%   4%

17%

67%

B   Total me1 stim only=2014

6%   1%

22%

70%

C   Total me3 stim only=2981

2% 2%

31%

66%

D   Total me1 stim and me3 stim=11645

6%   < 1%   9%

84%

E   Total no stim mark=656

16%

4%

7%

73%
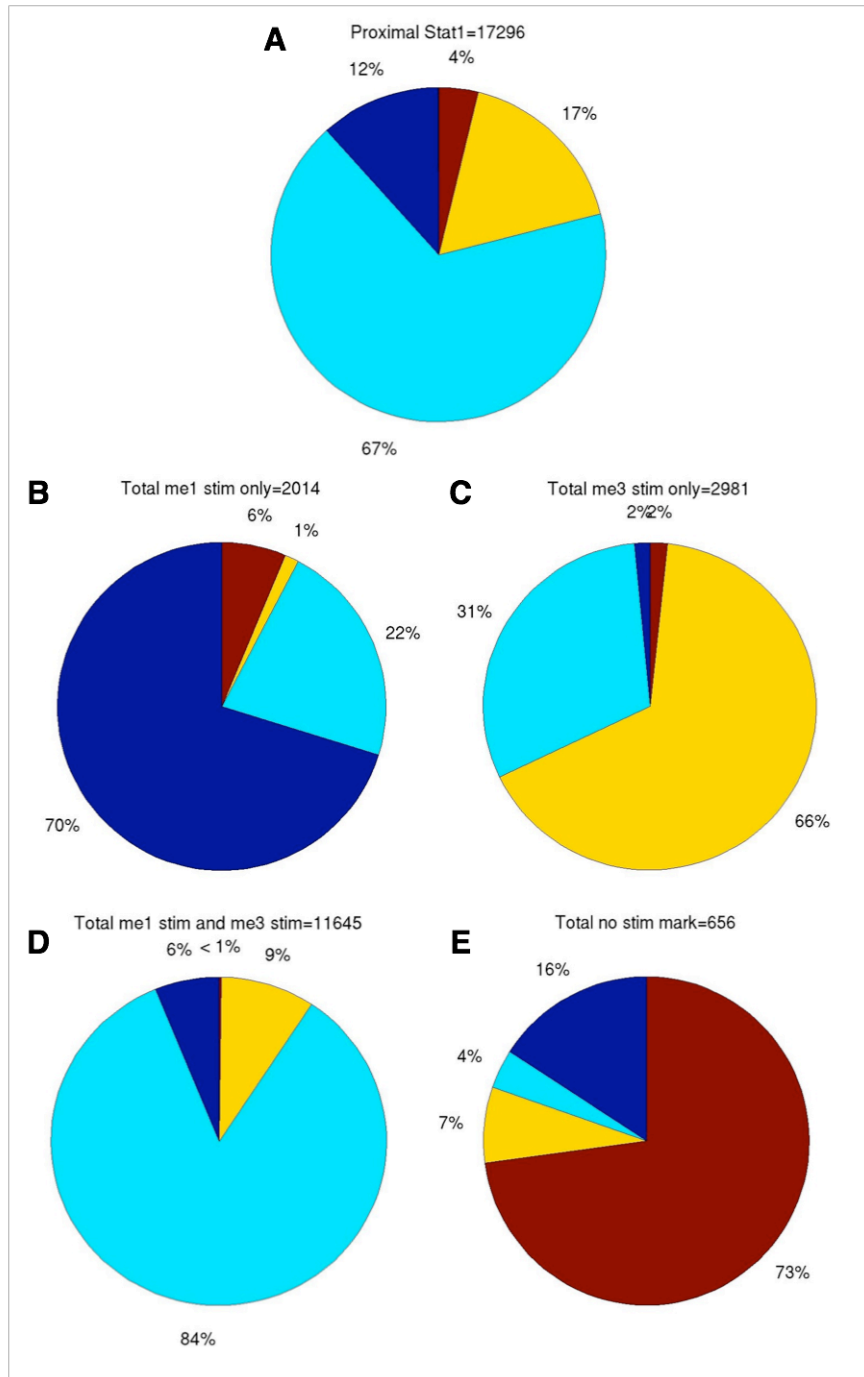
Figure PC3. Pattern concordance for all proximal STAT1 (IFNG) sites.

38

A SandU Stat1=18762
3%
9%
20%
67%

B Total me1 stim only=3756
7%
< 1%
11%
82%

C Total me3 stim only=1742
3% 3%
34%
60%

D Total me1 stim and me3 stim=12611
< 1% 6%
13%
81%

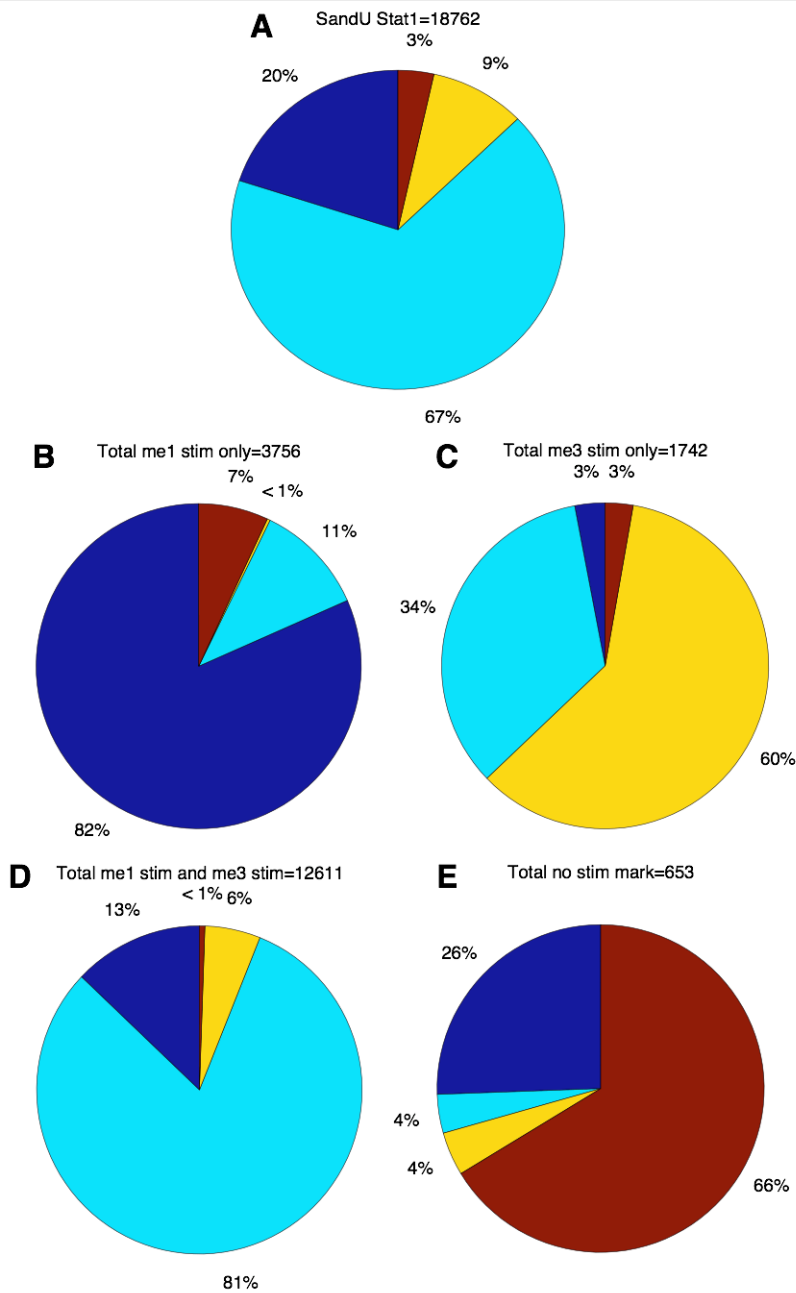**E** Total no stim mark=653
26%
4%
4%
66%

Figure PC4. Pattern concordance for all STAT1 sites that were enriched in stimulated cells at locations that were already enriched in unstimulated cells (see Figure PC5 for distal STAT1).
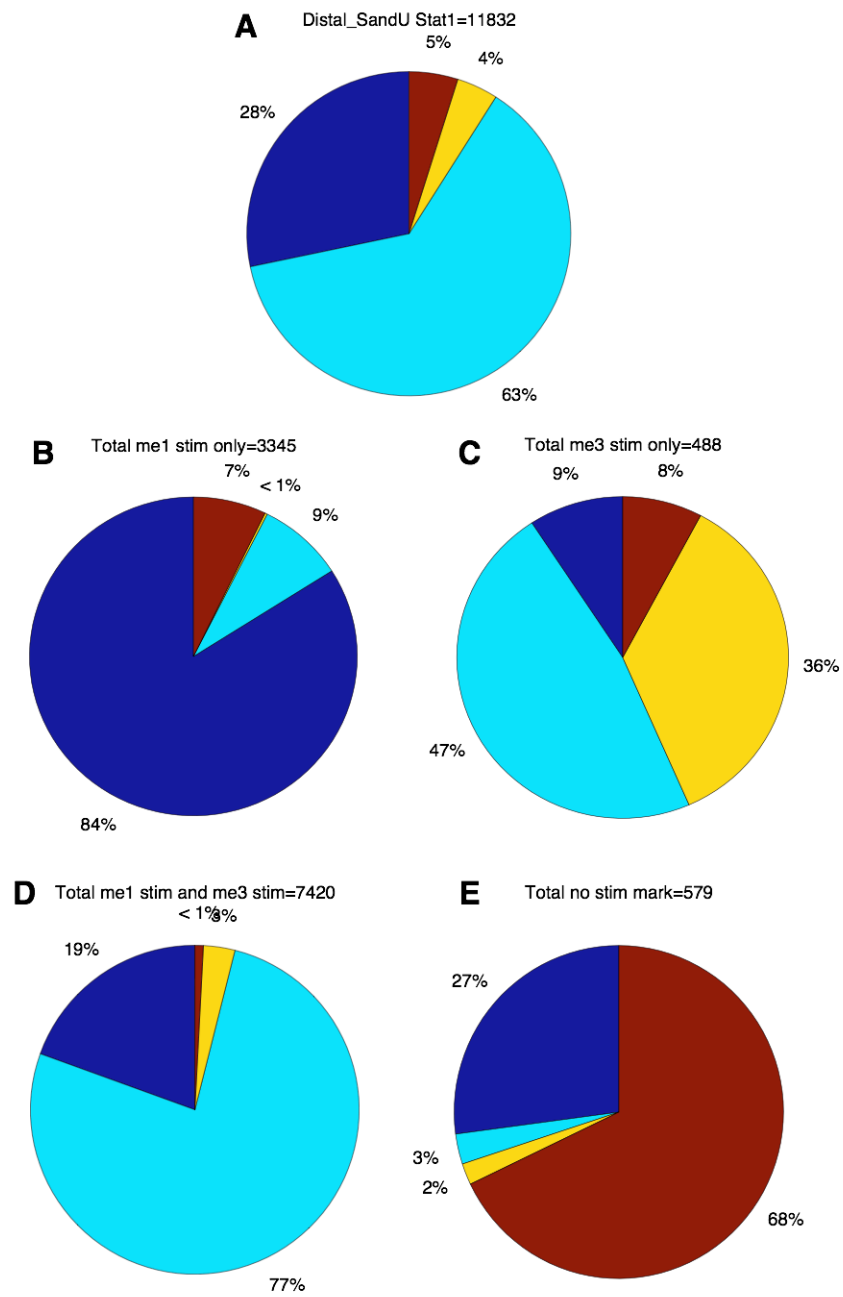
A   Distal_SandU Stat1=11832

5%

4%

28%

63%

B   Total me1 stim only=3345

7%   < 1%

9%

84%

C   Total me3 stim only=488

9%   8%

36%

47%

D   Total me1 stim and me3 stim=7420

< 1%   3%

19%

77%

E   Total no stim mark=579

27%

3%

2%

68%

Figure PC5. As Fig. PC4, but for distal STAT1 sites.

A Sonly Stat1=51530

17%
6%
37%
40%

B Total me1 stim only=19279

10%
< 1%
7%
83%

C Total me3 stim only=2952

9%
10%
33%
48%

D Total me1 stim and me3 stim=20506

1%
5%
26%
67%

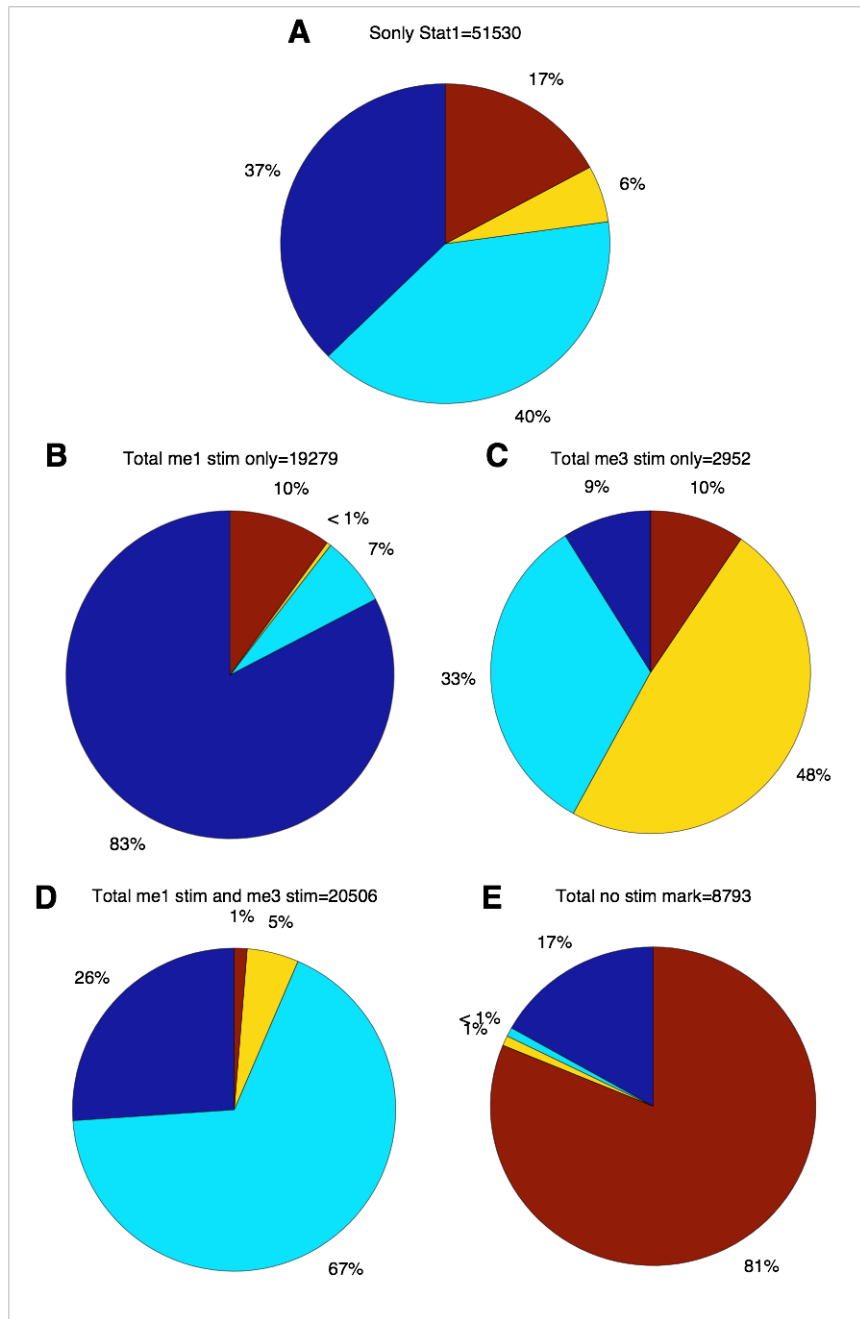**E** Total no stim mark=8793

17%
< 1%
1%
81%

Fig. PC6. Pattern concordance for STAT1 sites that were enriched in stimulated cells at locations that were not enriched in unstimulated cells (see Fig. PC7 for distal STAT1).
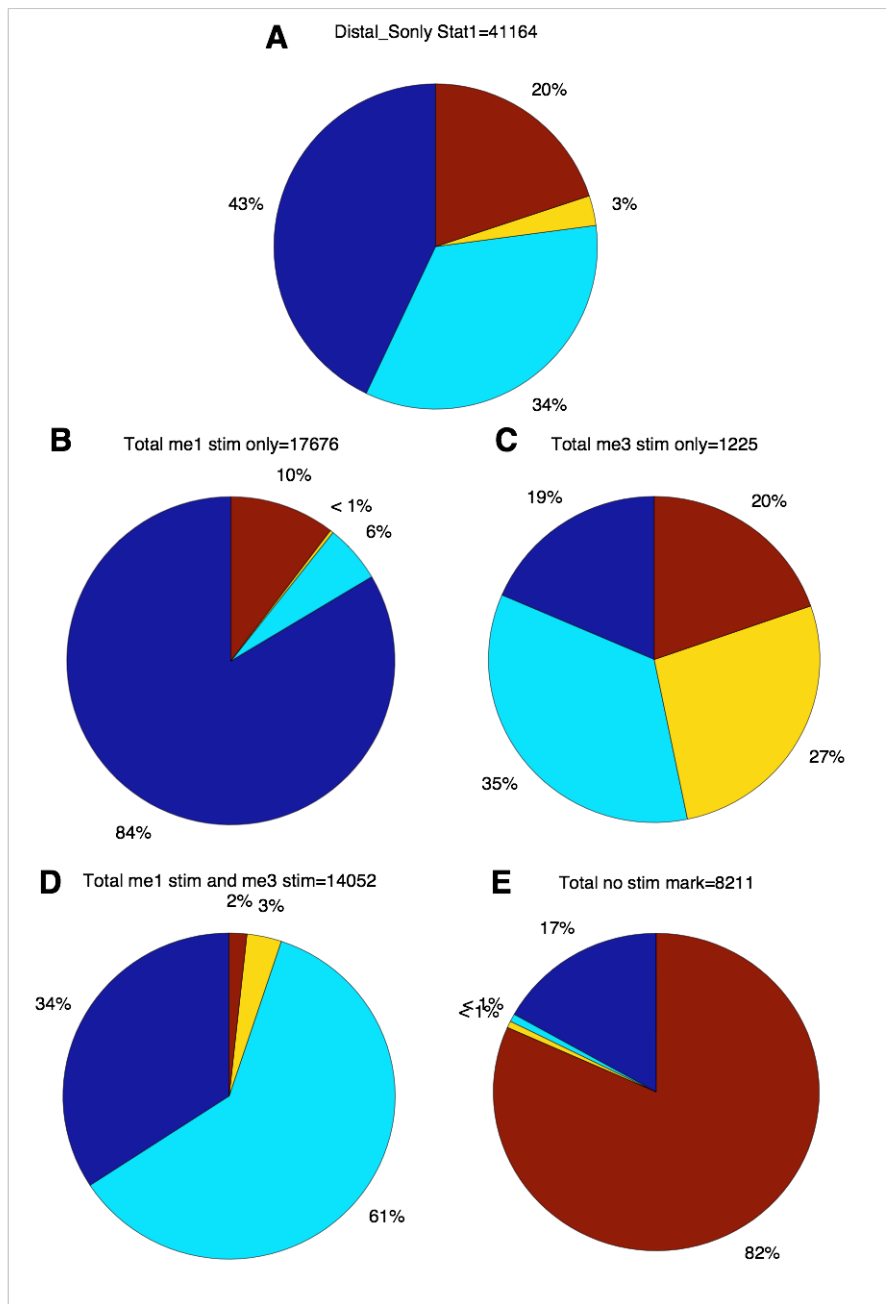
A   Distal_Sonly Stat1=41164

20%

3%

43%

34%

B   Total me1 stim only=17676

10%

< 1%

6%

84%

C   Total me3 stim only=1225

19%

20%

27%

35%

D   Total me1 stim and me3 stim=14052

2% 3%

34%

61%

E   Total no stim mark=8211

17%

≤ 1%

82%

Figure PC7. As PC6, but for distal STAT1 sites.