Molecular Cell, *Volume 44*

Article

SAGA and ATAC Histone Acetyl Transferase Complexes Regulate Distinct Sets of Genes and ATAC Defines a Class of p300-Independent Enhancers

Arnaud R. Krebs, Krishanpal Karmodiya, Marianne Lindahl-Allen, Kevin Struhl, and Làszlò Tora

Supplemental Experimental Procedures

Antibodies

Anti-SPT20 (3006) polyclonal antibody was obtained by immunization of rabbit with C-terminal region (330-531 amino acids) of mSPT20 (NP_064379.1). The fragment was PCR amplified using forward primers 5'-CCCCATATGGCTCATGATGTAAAAGACGAT-3' and reverse primer 5'-CCGCTCGAGGCTTTCATGTCTCTGGGATGA-3'. The amplified PCR product was cloned in pET28b (Novagen) using Nde I and Xho I. Protein was expressed in *E. Coli* (BL21) and purified using Ni-NTA column. Rabbits were immunized with purified protein. The antibody was characterized in Western blot, immunoprecipitation and ChIP-Western blot (Figure S1).

Luciferase reporter assay

Predicted ATAC sites enriched with H3K4me1 and reduced H3K4me3 and negative regions devoid of any regulatory marks were amplified from human genomic DNA and cloned into pGL3 promoter vector (Promega) upstream of luciferase gene and SV40 minimal promoter. Co-ordinates of the cloned sites and primer information are provided into Table S4. 25 μg of cloned construct in pGL3 vector and 25 μg of PCH111 vector were electroporated into 5 million GM12878 cells at 350 V, 400 Ohm, and 960 μF. Cells were diluted to 0.5 million cells/ml into GM medium. 48 h post electroporation lucifearse activity was assessed and normalized with β-gal activity.

Transient transfections

siRNAs are from Dharmacon: scramble (D-001810-10), anti-hZZZ3 (L-013939-01-000), anti-hATAC2 (L-008481-00-000), anti-hPCAF (L-005055-00-000) and anti-hSPT20 (L-013820-00-000). The siRNAs were transfected into HeLa cells using Lipofectamine 2000 and OptiMEM serum free medium following the recommendations of the supplier. Cells were harvested after 72 hours post-transfection for protein and mRNA analysis by western blot or Q-RT-PCR, respectively. Inducible shGCN5 HeLa cells were used as described earlier (Orpinell et al., 2010).

Chromatin immunoprecipation (ChIP)

All ChIP experiments were carried out on $2X10^7$ cells per antibody. Cells were cross-linked with 1% formaldehyde, lysed and sonnicated in sonnication buffer (10mM Tris-HCl pH 7.5, 200mM NaCl, 1% SDS, 4% NP-40, 1mM PMSF) to obtain an average chromatin size of 1000bp. Chromatin was pre-cleared using 50ul of a 50% protein A sepharose (GE healthcare) slurry for 1h at 4°C with gentle inverting. Immunoprecipitations were carried out in 10ml of

IP buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 2mM EDTA, 1% Triton-X 100). 15ul of ZZZ3 and SPT20 serum antibodies were used. Input chromatin was obtained after pre-clearing, by de-crosslinking and purifying input DNA using a Qiaquick column (Qiagen) according to manufacturer's instructions. Immunoprecipitations were carried out with inverting at 4°C for 14–16 h. The samples were then incubated with 50 uL of a 50% Protein A sepharose slurry for 3 h at 4°C with gentle inverting. IP samples were reverse-crosslinked and the DNA was purified using a Qiaquick column (Qiagen). Q-PCR using SYBR green was used to validate known target sites before and after sequencing.

Library preparation and sequencing

We followed the manufacturer's (Solexa) protocol for creating genomic DNA libraries and as previously described in (Auerbach, R. K., et al. 2009). ChIP DNA and input DNA were first band-isolated on a 2% agarose to obtain fragments between 150 and 350 base pairs and DNA was extracted using the QIAquick gel extraction kit (Qiagen) and eluted in 34 µL. For input DNA, because DNA amounts are higher than DNA recovered by ChIP, input DNA after gel extraction was diluted 1:5. After end-repair and addition of a single adenosine ("A") nucleotide, adapters were ligated to samples for 15 min at room temperature in the following fashion: the samples eluted from the MinElute column in 10 µL were ligated to 1 µL of adapters using 1.3 µL of LigaFast T4 DNA Ligase (3 Units/µL; Promega) and 12.3 µL Rapid Ligation Buffer (Promega). For DNA libraries, the Illumina genomic DNA adapters were diluted 1:10. After 15 min, samples were purified with the MinElute PCR purification kit (Qiagen).
Adapters in excess were eliminated by using gel purification on a 2% agarose E-Gel (Invitrogen) for 20 min, together with Track-It 50 bp DNA ladder (Invitrogen). DNA fragments ranging from 150 base pairs to 500 base pairs were extracted and recovered in 28 µL EB with a QIAquick gel extraction kit (Qiagen). To amplify the library, PCR was performed using Illumina genomic DNA primer "1.1" and Illumina genomic DNA primer "2.1" with 15 cycles (Input DNA) or 17 cycles (ChIP DNA) of amplification. A final size selection was performed using a 2% agarose E-Gel to obtain a library with a median length of ~230 bp which is within the recommended size range for cluster generation on Illumina's flowcell. The library was recovered in 20 µL EB using MinElute Gel Extraction kit (Qiagen). Finally, DNA concentrations and purities (A260/280 nm ratios) were measured on a Nanodrop spectrophotometer. Sequencing was carried out on the Illumina (Solexa) platform at a sequencing depth of 1 lane averaging 10 million reads, read length of 27+bp, single-end reads and mapped to human genome build (hg18).

Bioinformatics procedure

Bioinformatics was performed using existing tools (cited in document) as well as *ad hoc* Python, Java and R scripts available upon request.

Data pre-processing

Establishment of list of reference loci
ChIP-seq data were mapped using the ELAND software (Illumina) allowing one mismatch. Mapped read data were used as an input and to establish list of loci using MACS (Using default parameters except: mfold 12; tag size according to platform; band width 100) (Zhang et al., 2008). Input DNA file was used as a control in all peak detection analyses. List of detected peaks were further filtered for previously described artefacts (Pepke et al., 2009) by selecting on the peak length (100< length <1000). Peaks were ranked based on the number of contributing reads/peak and cut-off was experimentally defined by ChIP-qPCR validation to establish list of various confidence sites (high>25tag/peak, medium>20tag/peak). The high

confidence binding site list was used for downstream analyses.

Creation of density files for genome browser data visualisation
Raw BED files are used as input for *ad hoc* (WIG) density file creation script as described in ((Ye et al., 2011). Reads are directionally extended of their theoretical length (200bp), and 25bp bins are created. In each bin, the maximal number of overlapping reads is computed. Tracks were uploaded and displayed using fixed scale representation in the UCSC genome browser (Kent et al., 2002).

Expression calculation using RNA-seq data
Aligned read file were obtained from the ENCODE servers. A Read/Kilobase/Million reads score (PRKM) (Mortazavi et al., 2008) was then computed for each transcript using Refseq as a reference database.

Data analysis

Enrichment calculation
Enrichment is calculated as previously described . Briefly, enrichments are defined as the number of tags present in a defined window (1kb) around the reference site. ChIP-seq enrichments (e) are defined as e = log2 ((foreground tags+q)/(background tags+q)). q is defined as an empirical constant (q=10 in the present analyses); foreground tags are the density value computed in the data track ; background tags, the density value in the control track. q is used to lower the contribution of noise variations that is assumed to be higher at low count levels. The use of the constant q reduces the influence of the signal variation in the noise measurement on the ratio calculation (Ye et al., 2011).

Overlap calculation
In order to avoid biases linked to peak detection algorithm in the calculation of overlap between complexes (false negative contribution), we use a method based on enrichment calculation. We use previously established high confidence binding sites lists and pooled them, correcting for sites being present in both lists (within 500bp) to establish a pooled list of binding sites. Then we collect enrichments over the control (input) track (see above for details in enrichment calculation) for the datasets to compare (i.e. SPT20 and ZZZ3). Bound (enrichment>=2) and unbound events (<2) are then separated and Venn diagram is plotted.

Average profile calculations
For average gene profiles, ATAC or SAGA associated genes (+/-2500 bp from binding site) were divided in 100 bins of length relative to the gene length. Moreover 10 equally sized (50bp) bins were created on the 5' and 3' of the gene and ChIP-seq densities were collected for each dataset in each bin. The average of enrichments over input in each bin was plotted for each dataset. For average TSS profile, the same genes as above were aligned on their TSS and densities were collected in 100 equally sized bins around the TSS (-/+2500bp)

Distance plot calculation
For binding sites of each category the distance to the 5' of the next gene is computed using Refseq as a reference gene set (method adapted from (Krebs et al., 2008)). Then fixed bin (1000bp) histogram was plotted, showing the frequency of events in each bin.

Complex binding/expression Correlation
Signal enrichment around 5' of all Refseq transcripts was calculated (-/+ 1Kb). Genes were ranked according to their RPKM score and grouped in bins of genes having similar expression levels. For each bin, a boxplot of the enrichment values at the corresponding genes is plotted using R boxplot function with default parameters (outliers > 1.5 IRQ (Inter Quartile Range)).

Gene expression spectra analysis

Pre-processed expression microarray data were downloaded from BioGPS repository (formerly known as GNF atlas) (Wu et al., 2009). A single relative expression value was computed for each gene in each of the 81 human tissues available. Genes not showing any signal in any of the tissues tested were excluded from the analysis. Genes identified as bound by the complexes were matched to the established reference expression list. In each set, more than 80% of the total set was represented on the expression array and thus matched with expression data. Genes categories were classified based on the number of tissues where significant signal was observed (cutoff 200). The definition of Tissue Specific (observed in less than 11/81 tissues), House Keeping (present in more than 65/81 tissues), and broadly expressed (11<present <65) was adapted from (Mohn and Schubeler, 2009). Control heatmaps (Figure S5) and piled barplot of the proportion of each category in each dataset was plotted.

Transcription Factor (TF) binding motif screen

DNA sequences spanned by ATAC and SAGA peaks were retrieved as a foreground set. As a background set, we used Dnase Hypersensitivity Sites (DHS) to isolate active regulatory regions in GM, that we classified as promoter (~8 000) or enhancer (~20 000) based on their H3K4me1/3 ratio. DNA sequences for each set were then retrieved.

All higher eukaryotes Position Frequency Matrixes were retrieved from JASPAR 2010 (Portales-Casamar et al., 2010) through the MotIV package, Postion Weigth Matrixes (PWM), were then calculated according to (Wasserman and Sandelin, 2004). The matchPWM function from Biostrings was then used to screen each foreground and background datasets for all PWM using a minimal matrix matching score of 95%. The number of occurrence of a given matrix within each set was then calculated. In order to account for differences in number of input sequences between sets, in the pairwise comparison, counts were normalized to the minimum sum. Each foreground set was then plotted to its corresponding background set (separating enhancer/promoters) and elements enriched over 2 fold were extracted. For each candidate TF, the RPKM value was extracted from the RNA-seq and wideness of expression was estimated using BioGPS (see Figure 5).

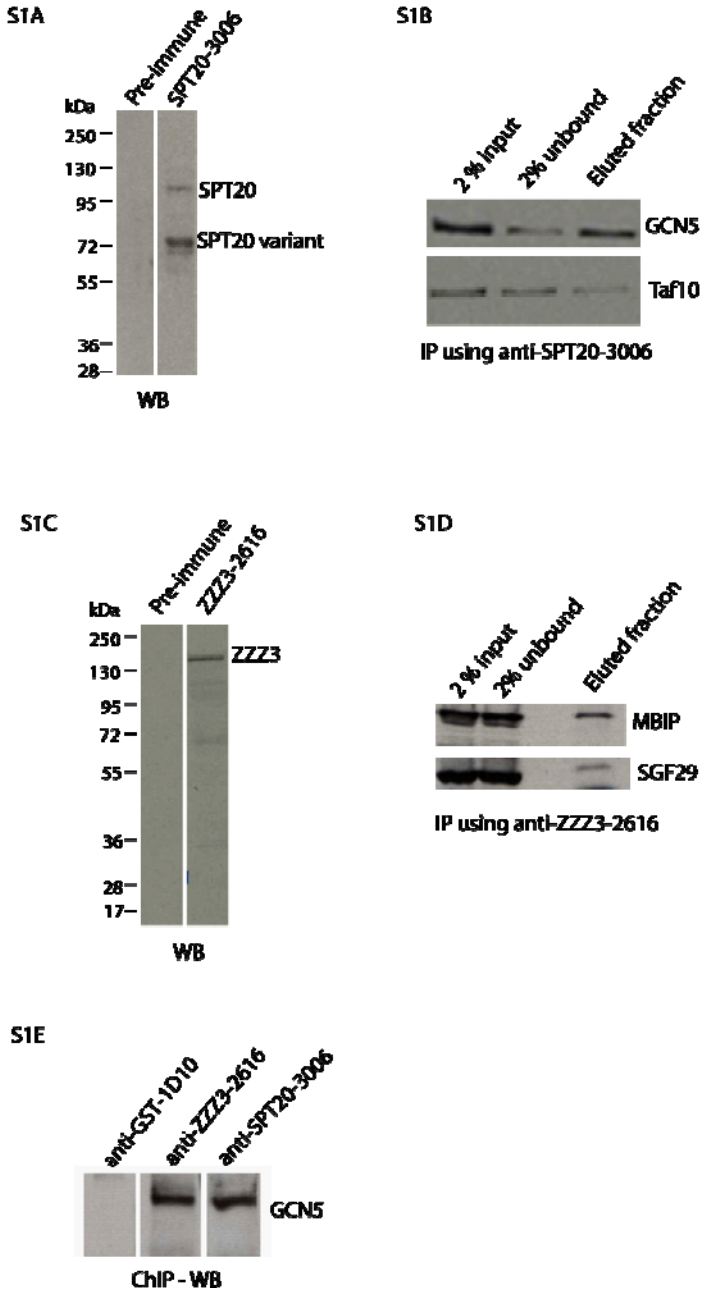


Figure S1. Characterization of Polyclonal Anti-mSPT20 Antibody Raised in This Study and Anti-hZZZ3 Antibody for ChIP, related to Figure 1

(**A**) Comparison of the detection of SPT20 by preimmune and immune sera (3006; crude) using Western blot analysis detection on HeLa cell nuclear extract (NE). (**B**) Immunoprecipitation was carried out from HeLa cell NE using the developed anti-SPT20 antibody. GCN5 and TAF10 (subunits of SAGA complex) were detected by western blot using specific antibodies on the indicated fractions. (**C**) WB analysis of ZZZ3 using preimmune and ZZZ3-2616 immune sera (purified). (**D**) Immunoprecipitation was carried out from HeLa cell NE using the anti-ZZZ3 antibody. MBIP and SGF29 (subunits of ATAC complex) were detected by western blot using specific antibodies on the indicated fractions. (**E**) ChIP-WB analysis using GST-1D10, SPT20-3006 and ZZZ3-2616 antibodies. Detection of GCN5 in SPT20-3006 and ZZZ3-2616 ChIPs are shown.

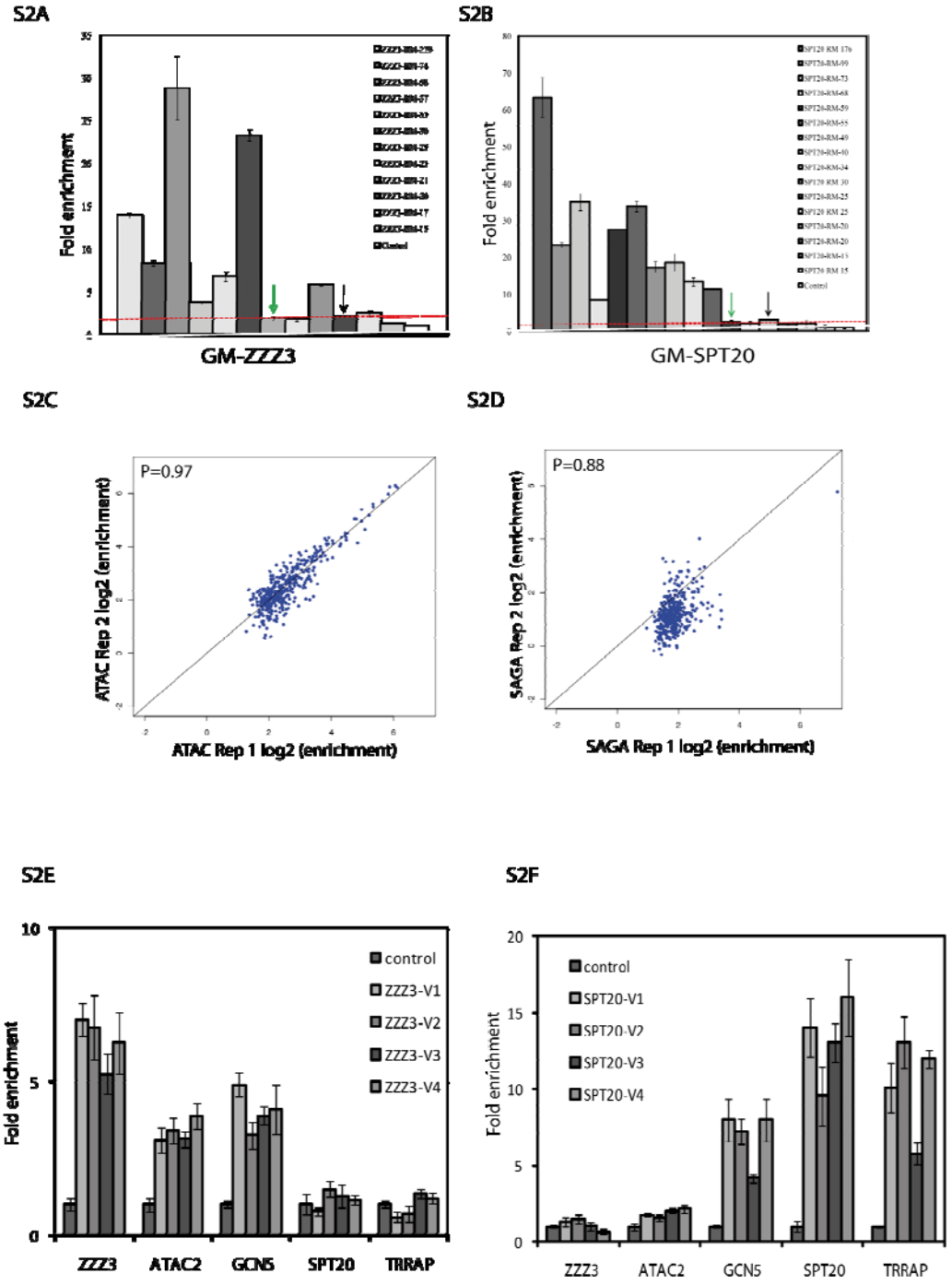


Figure S2. Establishment of Lists of Reference Loci and Replicate Analysis, Related to Figures 2 and 3

Validation of (**A**) ZZZ3 and (**B**) SPT20 ChIP-seq in GM12878 by ChIP-qPCR. Tag density per peak (shown on the right of the graph) is plotted as fold enrichment over an arbitrarily chosen control genomic region. Tag density above 25 and 20 were used for selecting the highly enriched sites (green arrow) and medium enriched sites (black arrow), respectively. The horizontal red dotted line indicates the two folds enrichments over the control genomic region. (**C, D**) Plot of the enrichment values for ATAC (**C**) and SAGA (**D**) in each replicate over the high confidence binding sites. Pearson correlation values are given. **(E, F)** Validation of the four ATAC **(E)** and SAGA **(F)** bound sites by ChIP-qPCR using the antibodies against the GCN5 (common subunit of the two complexes), ATAC2 (ATAC specific subunit) and TRRAP (SAGA specific subunit). Error bars represent the standard deviation for three biological replicates.

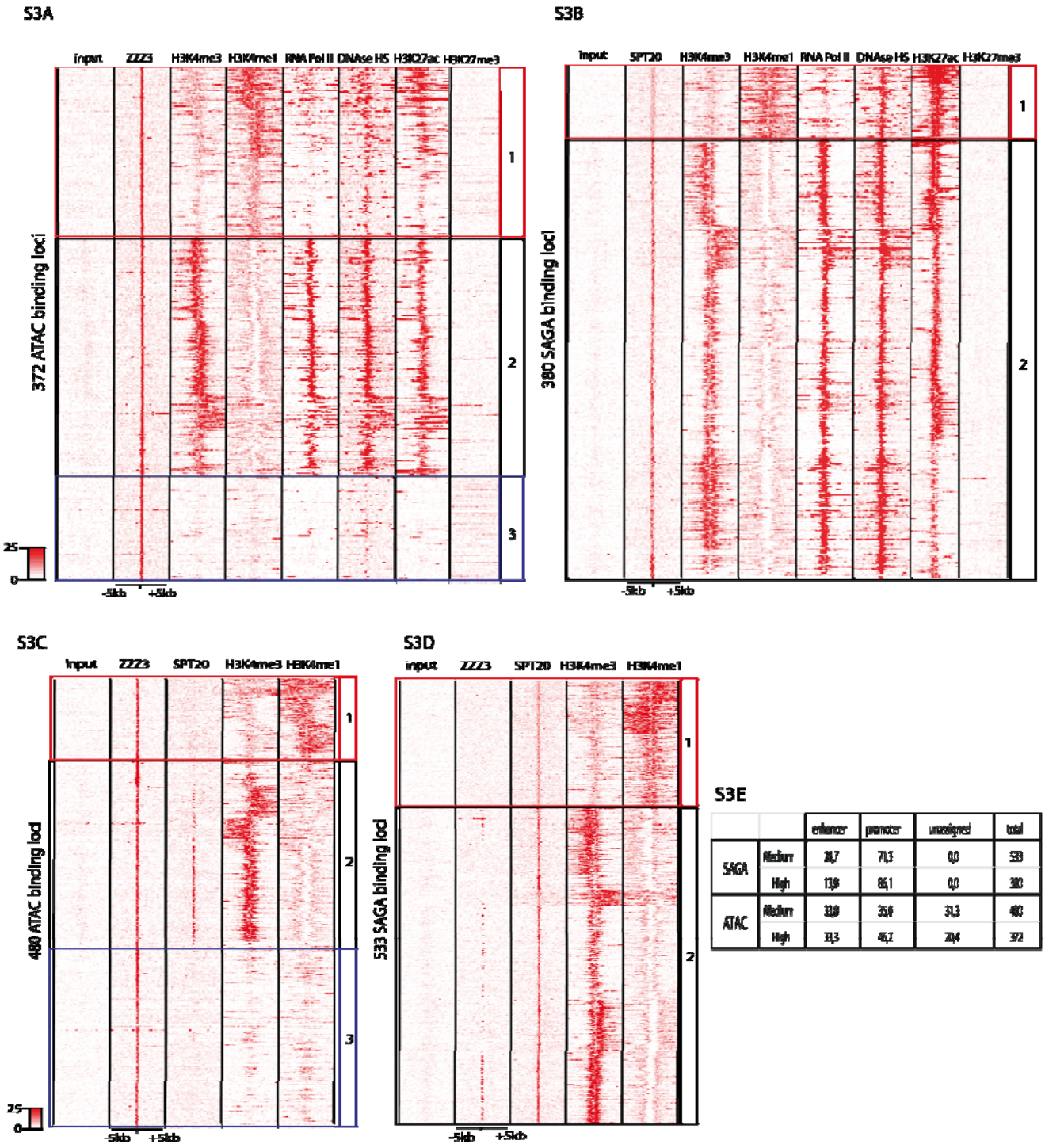


Figure S3. ATAC and SAGA Associate with Active Enhancers, Related to Figures 2 and 3

Distribution of H3K27ac and H3K27me3 over (**A**) the three categories of 372 ATAC bound sites (see Figure 2) and (**B**) the two categories of 380 SAGA bound sites (see Figure 3). **(C, D)** Heatmap of the signal density observed on regions surrounding the total ZZZ3 (480 ATAC) or SPT20 (533 SAGA) binding sites (±5 kb) for different genomic features (as indicated). The density map was subjected to clustering in order to create groups of loci sharing the same genomic profile. Loci were classified the same way as in Figure 2 and 3, showing no significant difference in the behavior of the medium/high confidence binding sites lists. **(E)** Table representing the proportion of sites found in each functional category using the

medium or the high confidence binding sites lists for ATAC and SAGA. It shows that proportions of each functional category in medium/high confidence is not significantly changed with the exception of SAGA enhancers that are more frequent in the lower confidence sites (in accordance with the fact that these sites are from lower intensity).

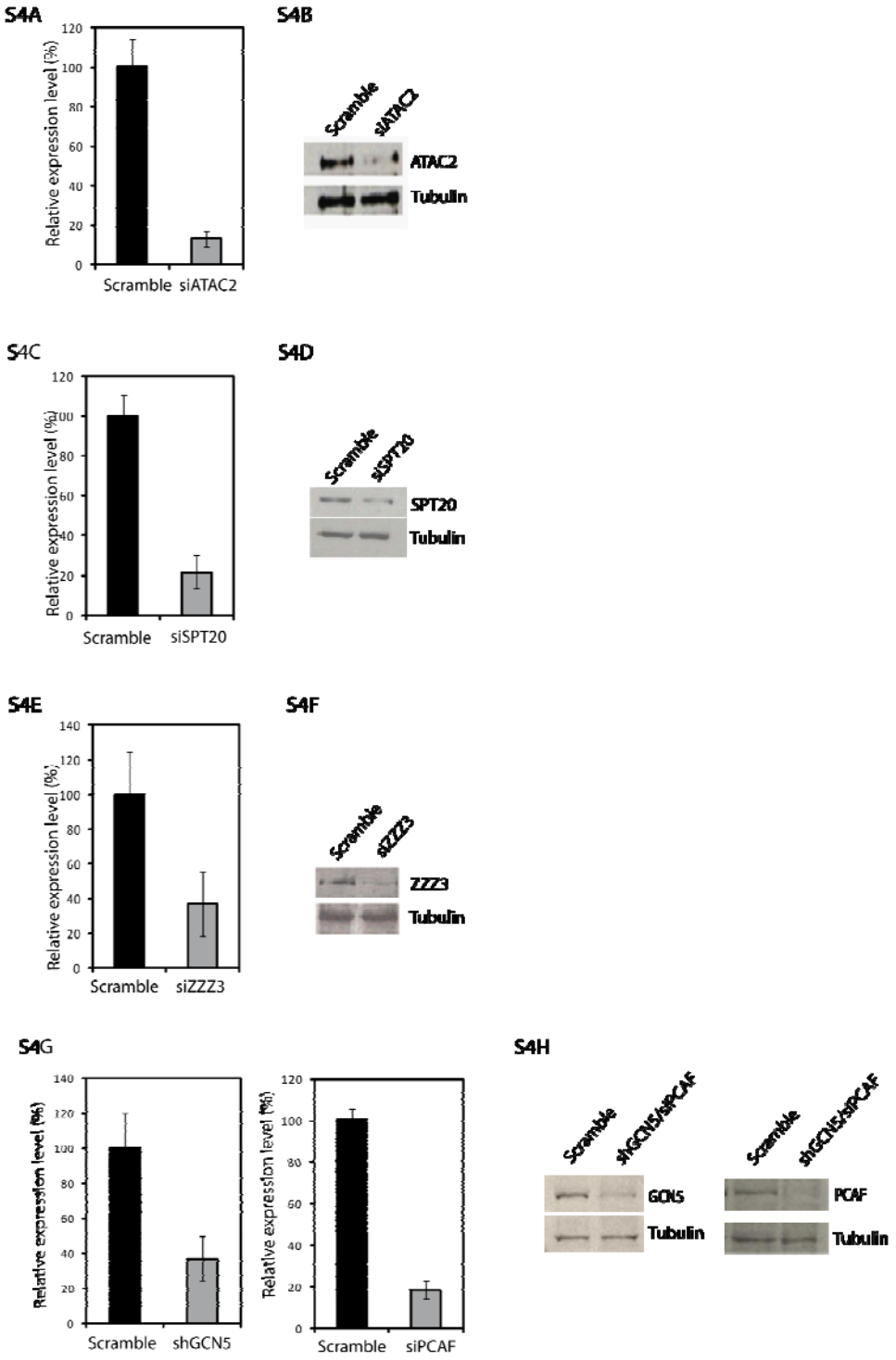
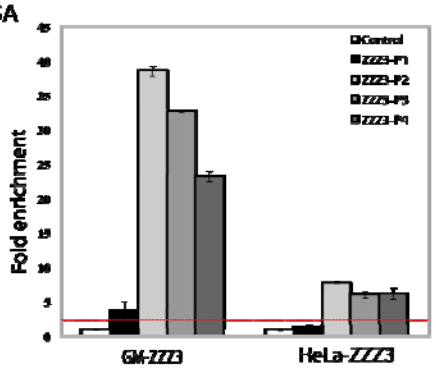


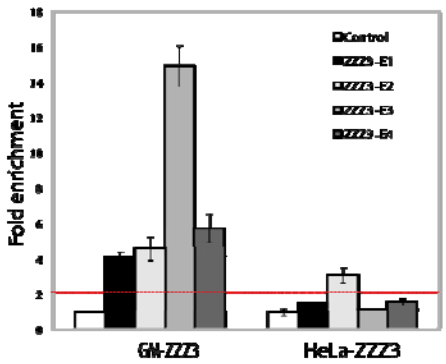
Figure S4. Measurement of Knockdown Efficiency of siRNA Treatment against ATAC2, SPT20, ZZZ3 and GCN5/PCAF Subunits, Related to Figure 4

(A, C, E, G) Total RNA extracts were prepared after 72 hours of treatment by the corresponding siRNA. The extracts were then reverse transcribed and relative transcript abundance was measured by qPCR using primers specific for the targeted mRNA. Results are presented relative to mRNA level in cells transfected with scramble siRNA. Error bars represent the standard deviation for three biological replicates. **(B, D, F, H)** Whole cell extracts were prepared after 72 hours of treatment by the corresponding siRNA (horizontal legend). The extracts were then analysed by Western blot with the indicated antibodies (vertical legend).

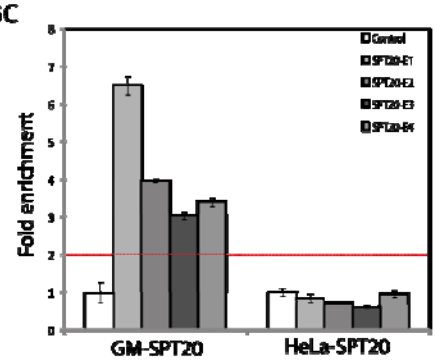
S5A



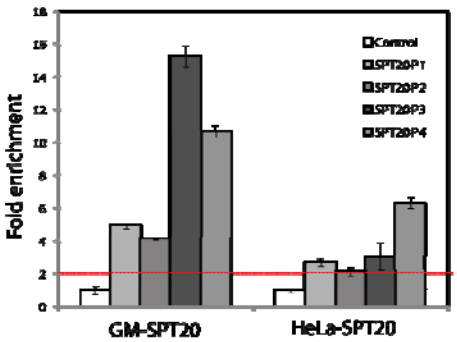
S5B



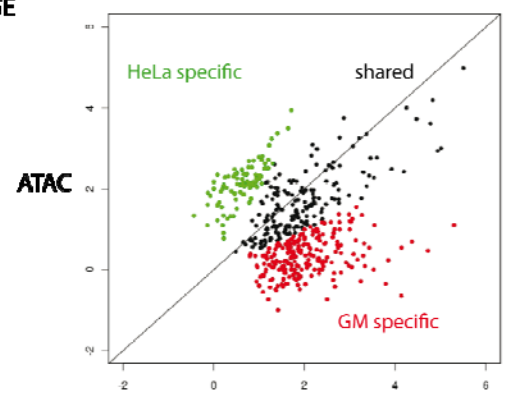
S5C



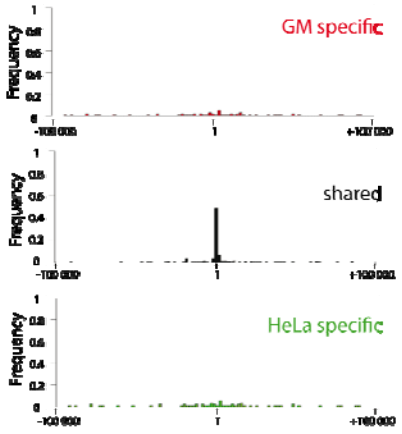
S5D



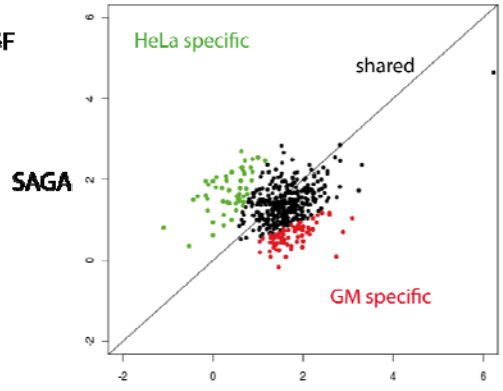
S5E



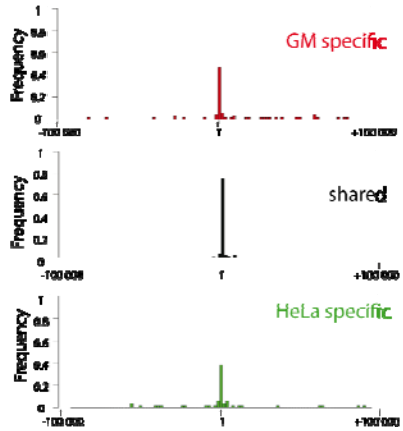
S5G



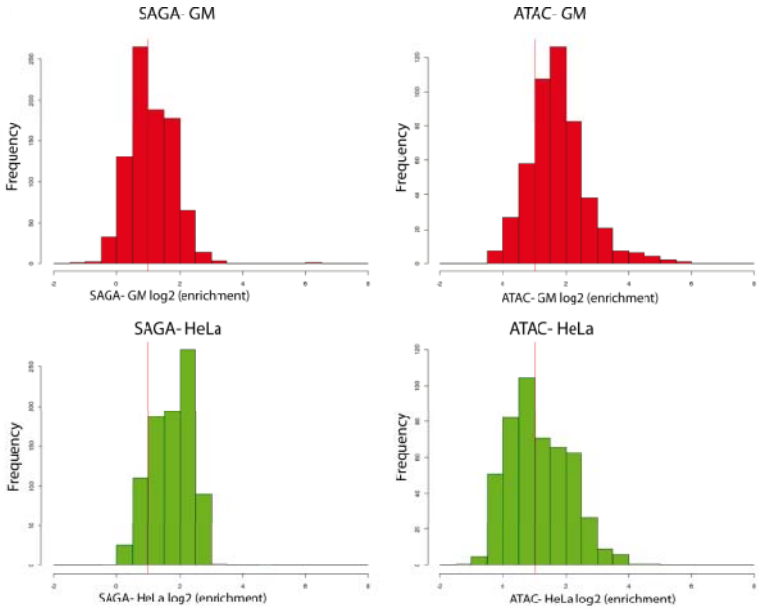
S5F



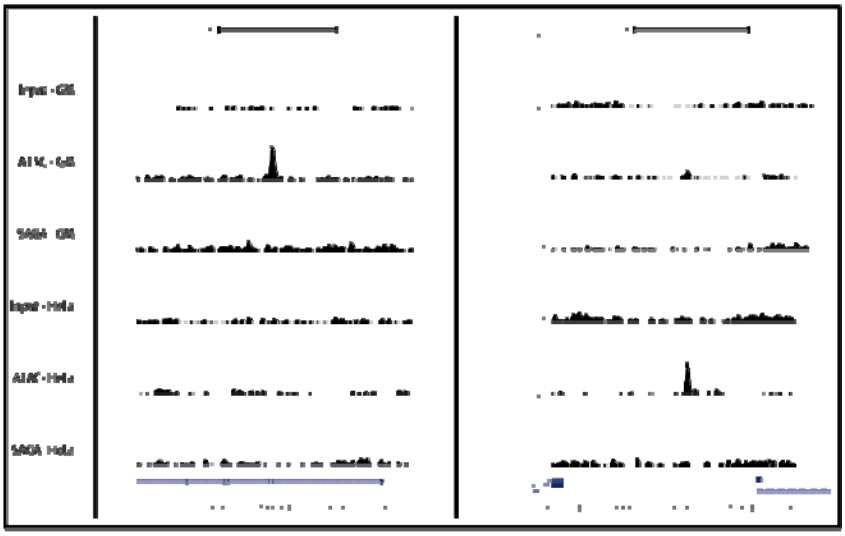
S5H



I.


SAGA- GM
ATAC- GM
Frequency
SAGA- GM log2 (enrichment)
ATAC- GM log2 (enrichment)
SAGA- HeLa
ATAC- HeLa
SAGA- HeLa log2 (enrichment)
ATAC- HeLa log2 (enrichment)

J.



K.

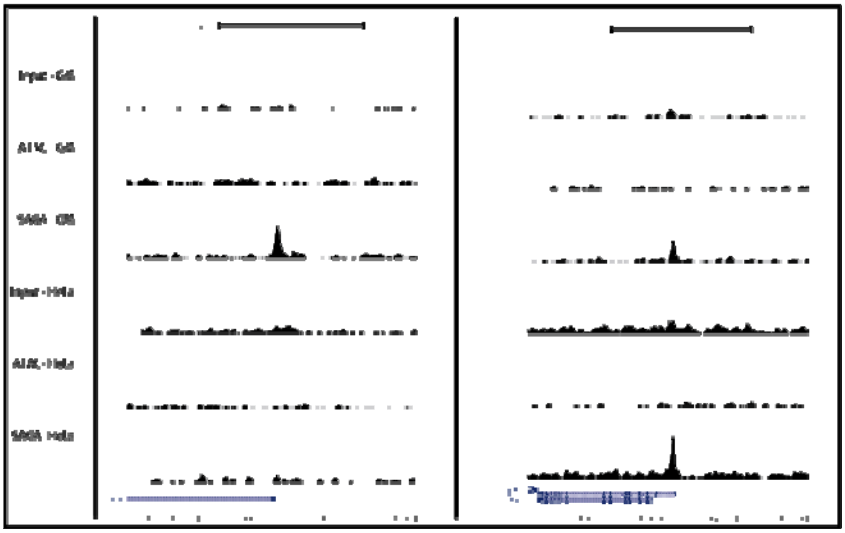


Figure S5. ATAC and SAGA Binding Exerts Higher Tissue Specificity on Enhancers Compared to Promoters, Related to Figure 5

Validation of binding of ATAC (ZZZ3) and SAGA (SPT20) complex to the promoters and enhancers in GM12878 and HeLa cells by ChIP-qPCR. Fold enrichment is calculated as described in Fig S2 for the predicted promoters (P) and enhancers (E). The horizontal red dotted line indicates the two folds enrichments over the control genomic region. (**A**) ATAC promoters predicted in GM12878 cells are conserved in GM12878 and HeLa cells. (**B**) ATAC enhancers predicted for GM12878 cells are not conserved in GM12878 and HeLa. Similarly, SAGA promoters (**C**) are conserved whereas SAGA enhancers (**D**) are not conserved in GM12878 and HeLa cells. Error bars represent the standard deviation for three biological replicates. (**E, F**) Dot plot representing ATAC or SAGA binding over high confidence binding sites from GM and HeLa cells. High confidence binding sites were extracted for ATAC and SAGA in GM and HeLa. These binding sites were pooled, correcting for binding sites found in both lists (within 500bp) to establish a pooled list of binding sites. Then enrichments over the control (input) track were collected and plotted. Loci found enriched specifically (over two fold) in GM (red dots) or in HeLa (green dots) were highlighted**. (G, H)** Frequency plot representing the distance of ATAC or SAGA binding sites to the transcription start sites (TSS, assumed to be the 5'of the transcript) of the closest gene in the genome in each category isolated. Binding sites conserved among the cell types tend to be preferentially found next to TSSs. Tissue specific binding events are enriched in events located distally from TSSs. (**I**) Comparative distribution of ATAC and SAGA enrichments in HeLa and GM cells. Enrichments were calculated at the detected peaks in each cell lines and the general intensity distribution was plotted. A similar dynamic range in the signal was observed for SAGA and ATAC in both cell lines, suggesting that the ChIP efficiency is similar in both cell lines. (**J, K**) UCSC genome browser tracks of representative examples of ATAC (J) and SAGA (K) peaks in HeLa and GM cells.

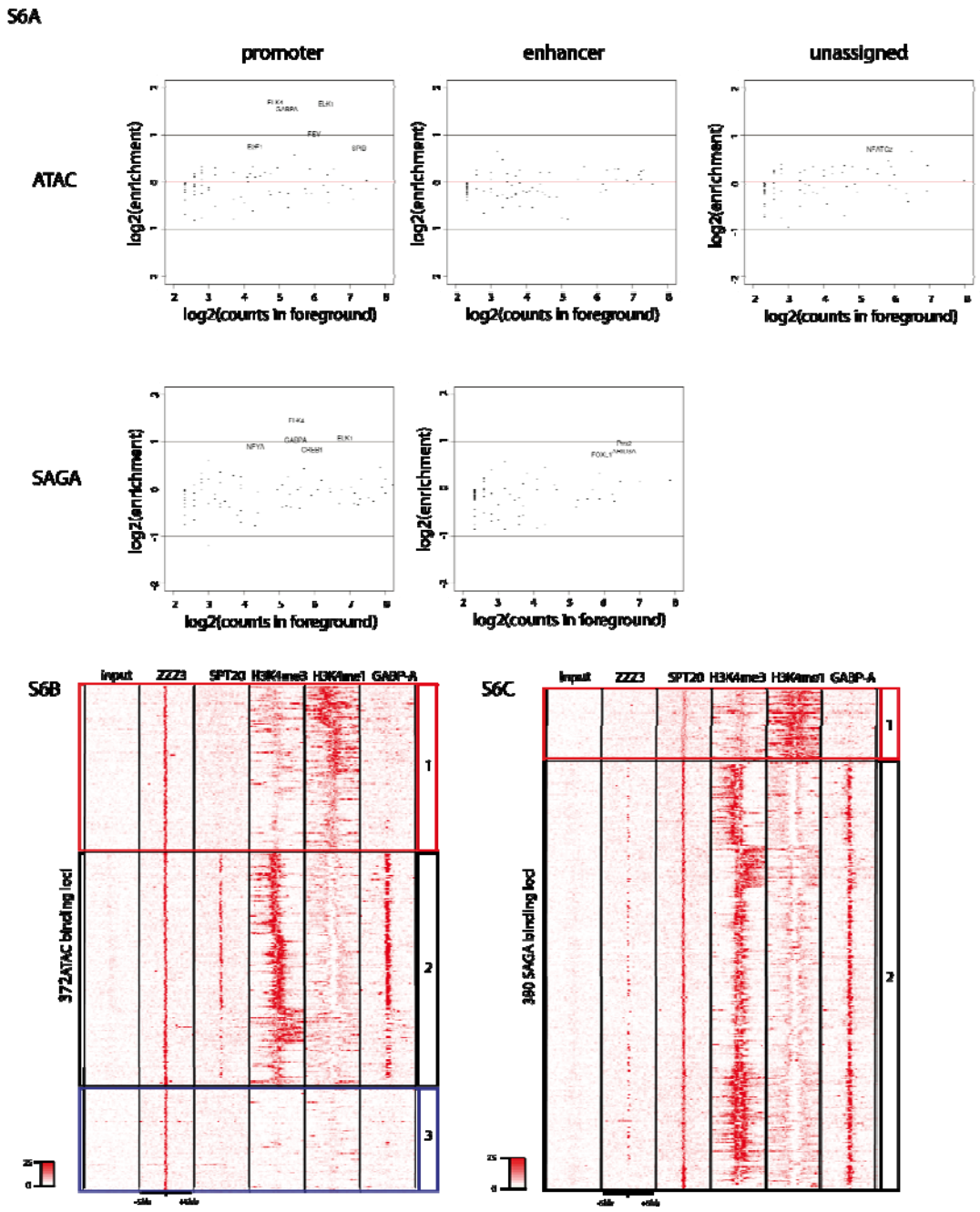


Figure S6. A Screen of Transcription Factor Motifs Identifies Candidate Transcription Factors for SAGA and ATAC Recruitment, Related to Figure 6

(A) Motif MA plots representing the distribution of all known JASPAR (2010 release) TF motifs in each analyzed subset compared to a control set of sequences. For TF motifs enriched over 1.5 fold, TF names were plotted. **(B, C)** Comparative heatmap of the ATAC (ZZZ3) or SAGA (SPT20) and GABPα ChIP-seq signal densities. Similar organization as in Figure 2A/3A was kept in order to illustrate the difference in GABPα binding on promoter and enhancers. Moreover SAGA and ATAC were plotted on both binding sites to contrast SAGA/GABPα signal correlation with coincidence of GABPα and ATAC signal over SAGA/ATAC shared loci.

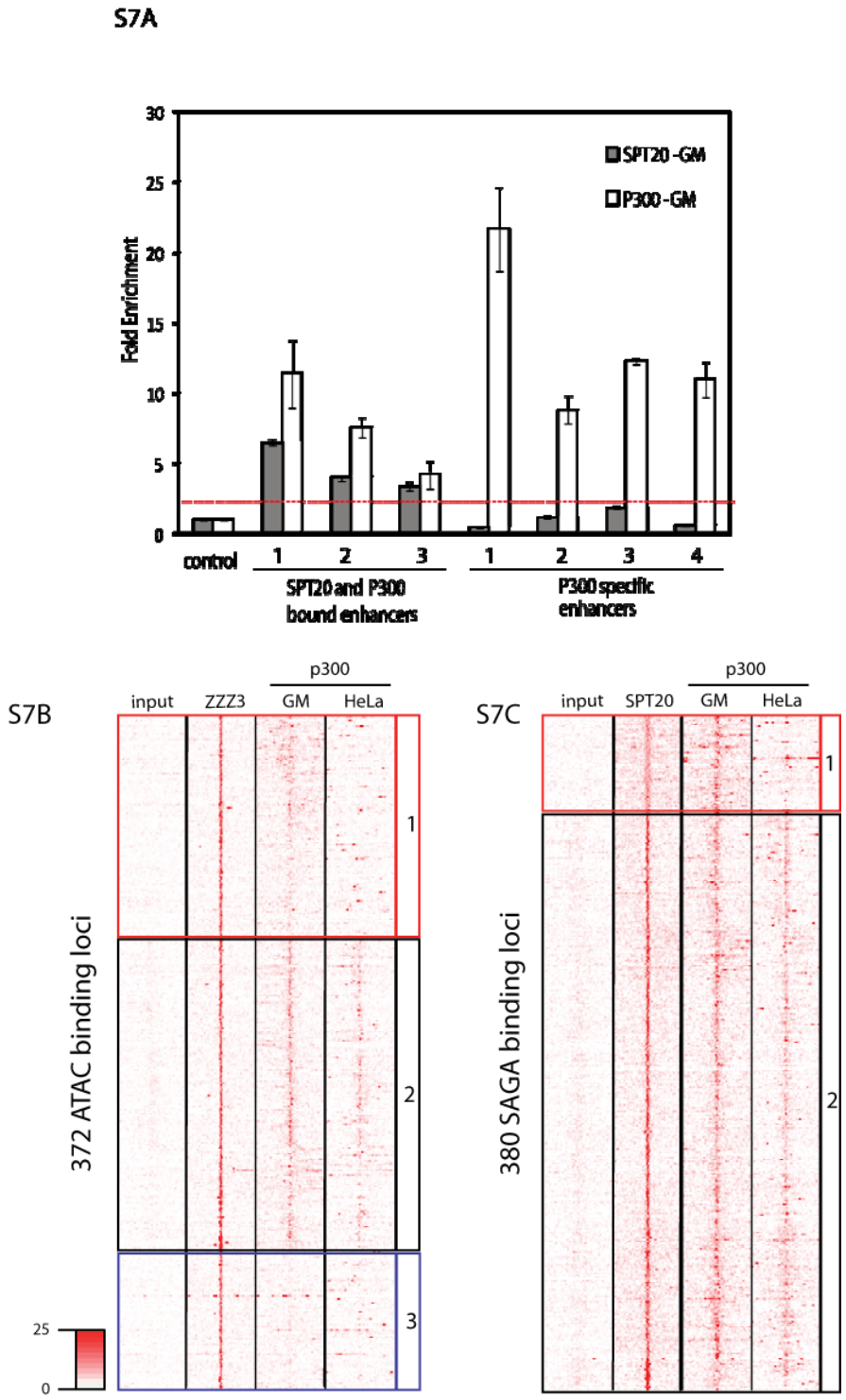


Figure S7. ATAC Is Defining a Class of Enhancers that Are Independent of p300, Related to Figure 7

(A) Validation of SPT20 and p300 bound enhancers by ChIP-qPCR quantification. Enrichment of SPT20 and p300 over, loci where no binding is expected (control) is calculated for SPT20 (grey bar) and p300 (white bar). Categories shown are enhancer loci predicted to be bound by SAGA and p300; and enhancer loci predicted to be bound only by p300. The horizontal red dotted line indicates the two folds enrichments over the control genomic region. Error bars represent the standard deviation for three biological replicates. **(B)** Comparative heatmap of the p300 ChIP-seq signal density in GM12878 and HeLa cells around the high confidence ZZZ3 binding sites. Similar loci organization as in Figure 2A was

conserved in order to illustrate differences in cell type binding specificity between enhancer (red), promoter (black) and undefined (blue) type of loci. **(C)** Similar analysis has been carried out for SAGA-bound loci. Note that newly detected binding sites in HeLa are not represented since H3K4me1/3 data are not available in this cell type.

Table S1. Gene Ontology (GO) Analysis of the Genes Bound by the SAGA and ATAC Complexes, Related to Figure 5

ATAC bound promoters

Biological functions

Term	P Value	Benjamini
Protein biosynthesis	1.32E-12	9.73E-11
Protein metabolism and modification	1.82E-04	0.006724434
mRNA splicing	0.007388667	0.167173907
Pre-mRNA processing	0.026355995	0.389896399
Intracellular protein traffic	0.083208651	0.723558339
Peroxisome transport	0.08867741	0.681856533

Molecular functions

Ribosomal protein	7.36E-11	6.40E-09
Nucleic acid binding	6.85E-07	2.98E-05
mRNA splicing factor	0.021403906	0.46604966
SNARE protein	0.026880705	0.447143618
mRNA processing factor	0.052007651	0.605175508

SAGA bound promoters

Biological functions

Term	P Value	Benjamini
Protein biosynthesis	1.03E-07	1.35E-05
Nucleoside, nucleotide and nucleic acid metabolism	1.72E-05	0.00112455
mRNA splicing	3.44E-05	0.00150136
Translational regulation	4.21E-05	0.00137909
Protein metabolism and modification	3.66E-04	0.00953361
Pre-mRNA processing	5.85E-04	0.01269289
Nuclear transport	0.0034636	0.06286814
tRNA metabolism	0.01699954	0.2447909
Intracellular protein traffic	0.02112584	0.26713438
Endocytosis	0.05683226	0.53536161
Mitosis	0.08081111	0.63340713

Molecular functions

Nucleic acid binding	2.73E-07	3.16E-05
Ribosomal protein	1.62E-04	0.00937431
Translation factor	0.00196476	0.073226
Ribonucleoprotein	0.00465268	0.12649605
Other ligase	0.00866878	0.18289899
Other transcription factor	0.01699735	0.28211091
Other zinc finger transcription factor	0.04326621	0.51951336
Transcription cofactor	0.05787872	0.57874137
Transcription factor	0.06676555	0.58959347
mRNA processing factor	0.09099404	0.66934579

Gene Ontology (GO) classification of ATAC- or SAGA-bound genes did not reveal any specific pathways specifically enriched between complexes, suggesting that they are involved in a variety of processes rather than specific pathways.

Table S2. Sequencing Statistics and Origin of the Generated High Throughput Sequencing Data Sets, Related to Figures 1–7

Antibody	Cell type	Platform	Aligned reads	Accession ID
input	GM12878	Yale	1.15E+007	wgEncodeYaleChIPseqAlignmentsRep1Gm12878InputV2
input	HeLa	Yale	2.98E+007	wgEncodeYaleChIPseqAlignmentsHelas3Input
ZZZ3 R1	GM12878	Yale	1.83E+07	GSM769224
ZZZ3 R2	GM12878	Yale	1.67E+07	GSM769225
ZZZ3	HeLa	Yale	1.30E+07	GSM769223
SPT20 R1	GM12878	IGBMC	1.64E+07	GSM769221
SPT20 R2	GM12878	Yale	1.64E+07	GSM769222
SPT20	HeLa	IGBMC	2.19E+07	GSM769220
H3K4me3	GM12878	Hudson	1.17E+07	wgEncodeBroadChipSeqAlignmentsRep1Gm12878H3k4me3
H3K4me1	GM12878	Hudson	1.49E+07	wgEncodeBroadChipSeqAlignmentsRep1Gm12878H3k4me1
H3K36me3	GM12878	Broad	1.66E+07	wgEncodeBroadChipSeqAlignmentsRep1Gm12878H3k36me3V2
H3K27Ac	GM12878	Broad	1.23E+007	wgEncodeBroadChipSeqAlignmentsRep1Gm12878H3k27acV2
H3K27me3	GM12878	Broad	1.57E+007	wgEncodeBroadChipSeqAlignmentsRep1Gm12878H3k27me3V2
DNAseHS	GM12878	Uw	2.58E+07	wgEncodeUwDnaseSeqAlignmentsRep1Gm12878
RNA Pol II	GM12878	Hudson	1.29E+07	wgEncodeHudsonalphaChipSeqAlignmentsRep1Gm12878Pol2
p300	GM12878	Hudson	2.47E+07	wgEncodeHudsonalphaChipSeqAlignmentsRep2Gm12878P300
p300	HeLa	Hudson	1.38E+07	wgEncodeSydhTfbsHelas3P300sc584sc584IggrabAlnRep2
GABPA	GM12878	Hudson	1.27E+07	wgEncodeHudsonalphaChipSeqAlignmentsRep1Gm12878Gabp
RNA-seq	GM12878	Yale	2.92E+07	wgEncodeYaleRnaSeqPolyaAlignmentsGm12878

Sequencing data were produced at Institute of Genetics and Molecular and Cellular Biology (IGBMC). Or in the sequencing platforms from ENCODE participants: Yale University (Yale), Hudson Alpha Institute for Biotechnology (Hudson), California Institute of Technology (Caltech) and University of Washington (Uw).

Table S3. Primers Used for ChIP-qPCR, Related to Figures 1, 4, and 7
Primer sequences provided in a separate excel file.

Table S4. Co-ordinates of the Cloned Enhancers and Primer Used for Cloning, Related to Figure 7

Name	Genomic cordinates (hg18)	Amplicon Size	Forward primer	Reverse primer
E1	chr7:30526412-30527421	1010	CCGACGCGTATGATGCCCAGAGAGAGCAGGA	GGAAGATCTACCCCAGAGACCACCTTTGTGA
E2	chr11:72113355-72114502	1148	CCGACGCGTCCATGACTCCAACTGCCAAAGA	GGAAGATCTTTTATTGCTGGGCACTCTGCTG
E3	chr8:67596037-67597180	832	CCGACGCGTTTTGGGTGCACCTGTCTTGTAA	GGAAGATCTGCTTCTTTCTGAATGGAGTGGA
E4	chr11:10288152-10289335	1148	CCGACGCGTAATAAGCGGTGGAGCCTGGATT	GGAAGATCTGGCAACTGAGCAAAACCCTGTC
N1	chr1:49765597-49766669	1073	CCGACGCGTGAGCAGGCCCAACTTACTCATT	GGAAGATCTAGTGGCAGAAATCAACTCAAGC
N2	chr11:97241300-97242490	1191	CCGACGCGTACGGACATATCTGAGGCCCTTG	GGAAGATCTGCCTGTGTCAAAATATCCCACGTA
H4	chr21:40,393,095-40,394,169	1213	CCGACGCGTAAGGATGGGCAATGTGGAGAGA	GGAAGATCTGTTGCCAAGCATGTGAATGAGG
H9	chr10:128075866-128077317	1452	CCGACGCGTTCCCAATCCAGGCTTAAGGACA	GGAAGATCTACCATTTGGGGTACAAGGGTGA

Supplemental References

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Krebs, A., Frontini, M., and Tora, L. (2008). GPAT: retrieval of genomic annotation from large genomic position datasets. BMC Bioinformatics *9*, 533.

Mohn, F., and Schubeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet *25*, 129-136.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Orpinell, M., Fournier, M., Riss, A., Nagy, Z., Krebs, A. R., Frontini, M., and Tora, L. (2010) The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. EMBO Journal *29*, 2381-2394.

Pepke, S., Wold, B., and Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. Nat Methods *6*, S22-32.

Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res *38*, D105-110.

Wasserman, W.W., and Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. Nature reviews Genetics *5*, 276-287.

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C. L., Haase, J., Janes, J., Huss, J. W., 3rd, and Su, A. I. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol *10*, R130.

Ye, T., Krebs, A. R., Choukrallah, M. A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res, 39(6):e35.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.