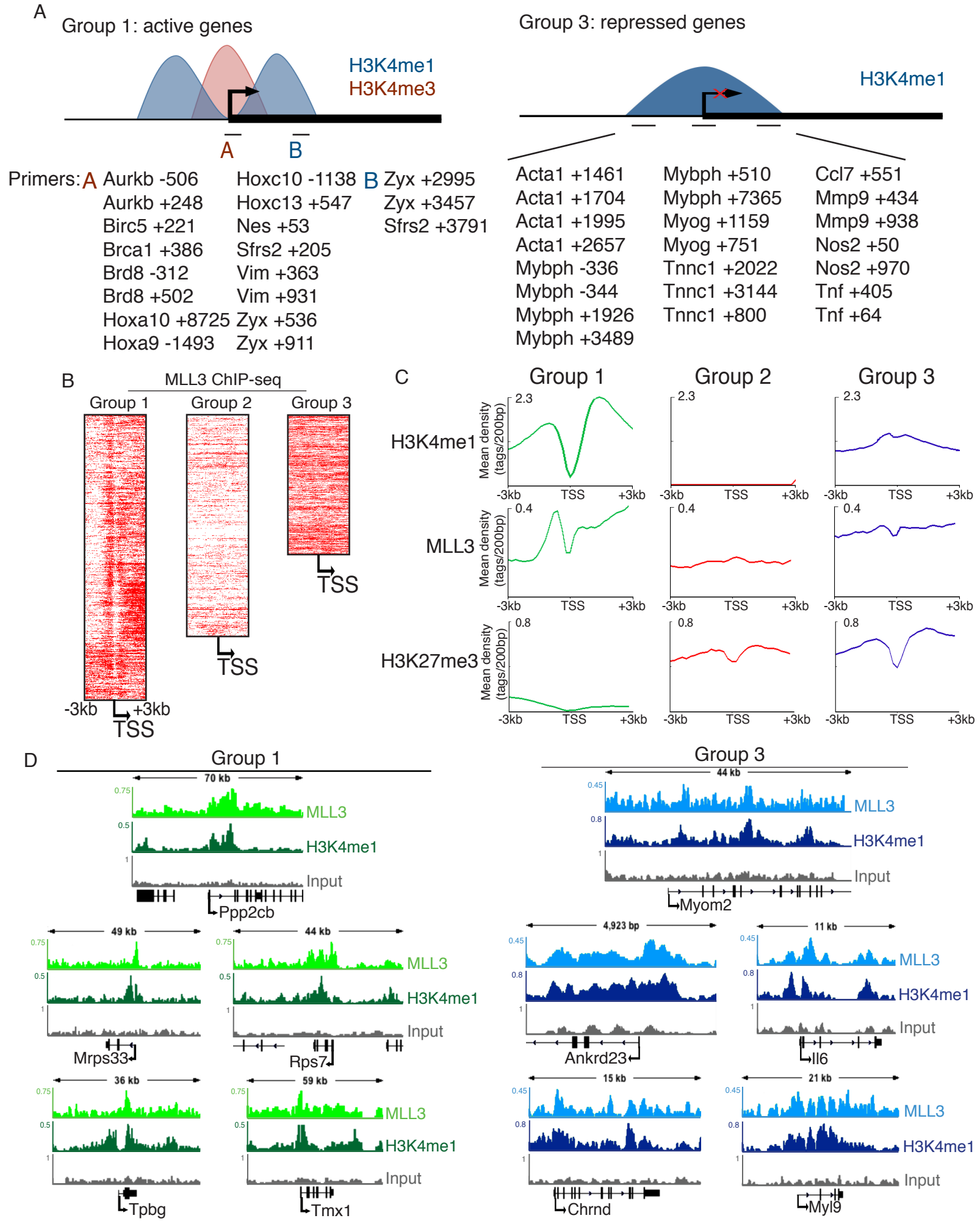
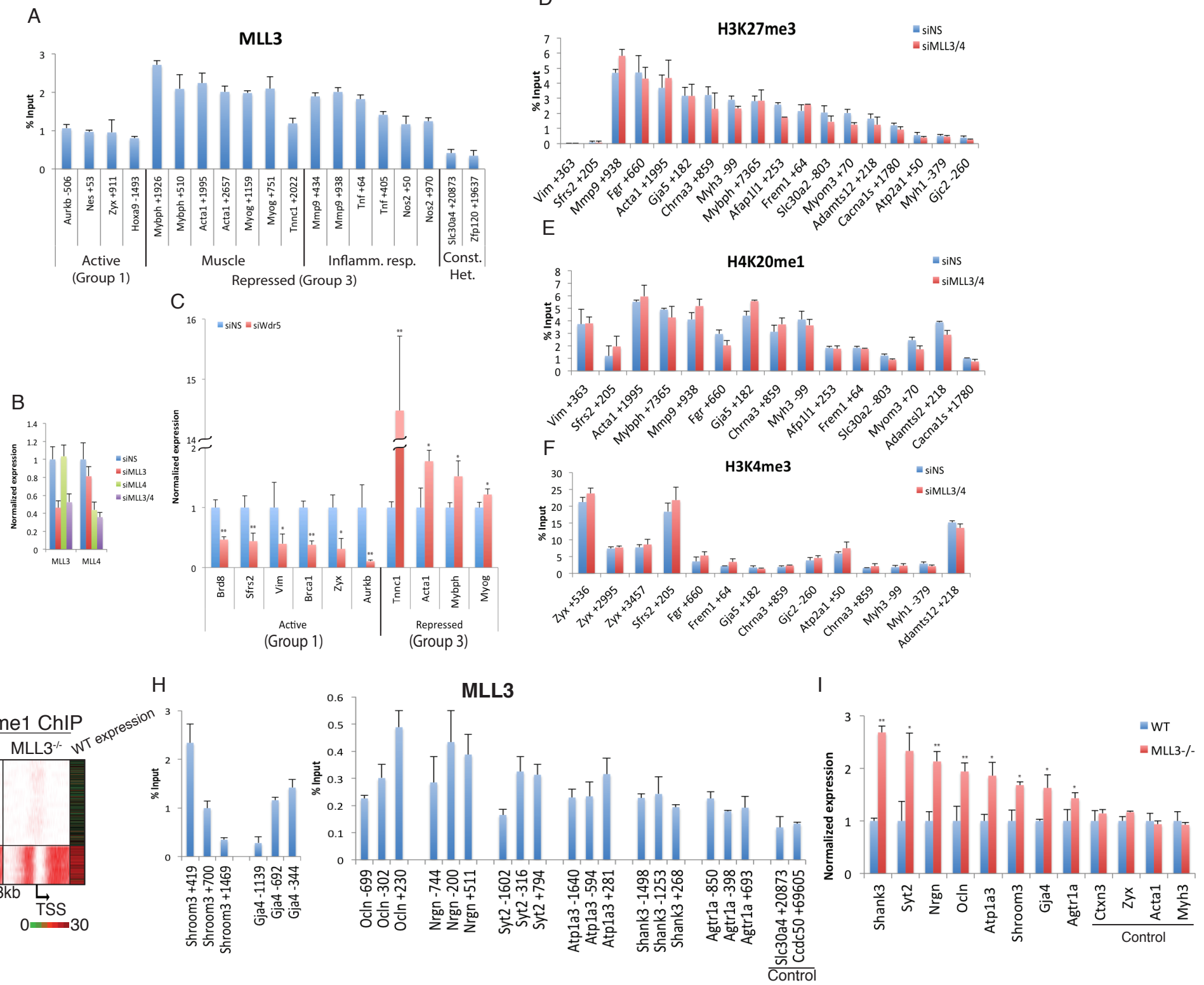
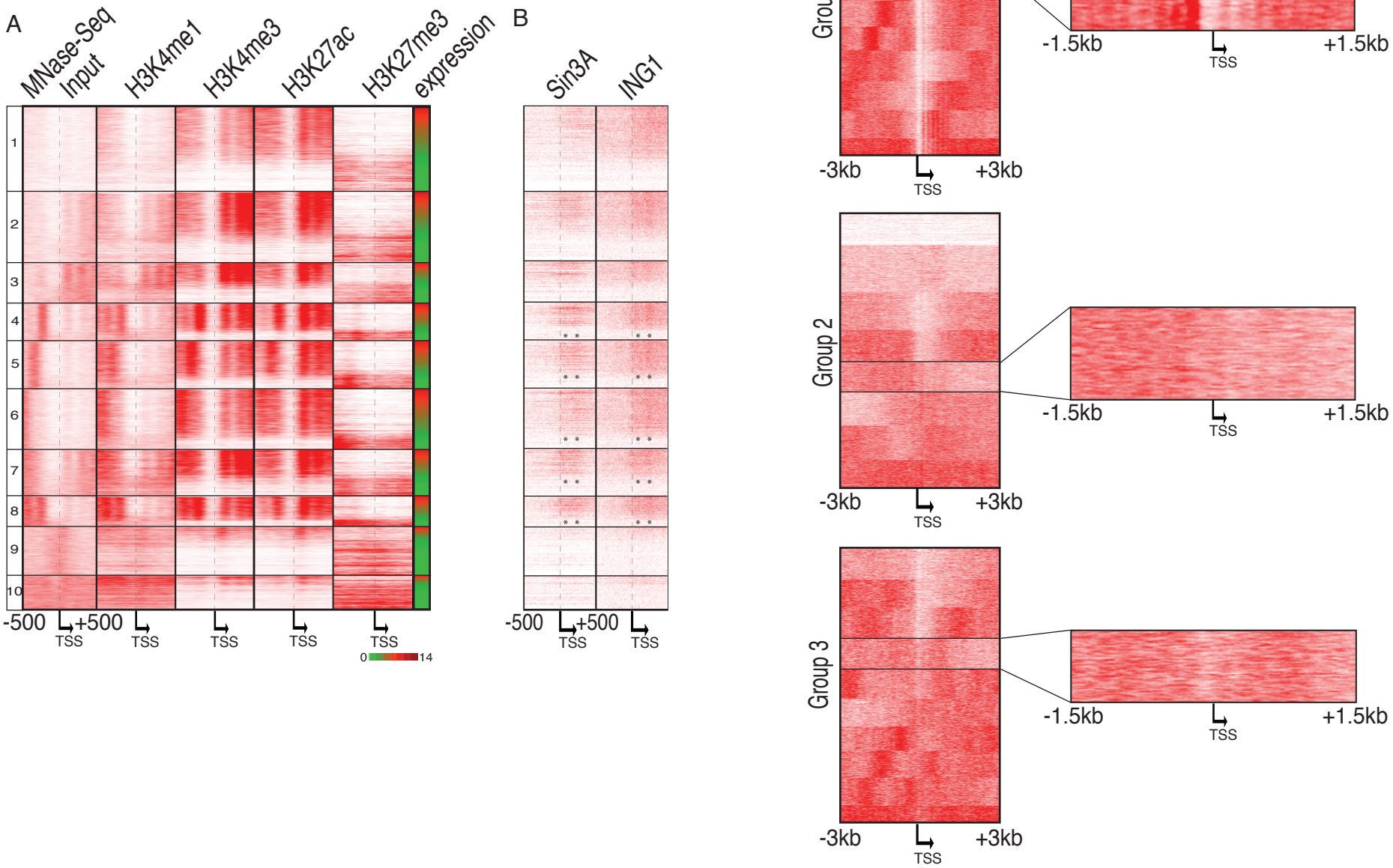
Cheng et al. Figure S1.



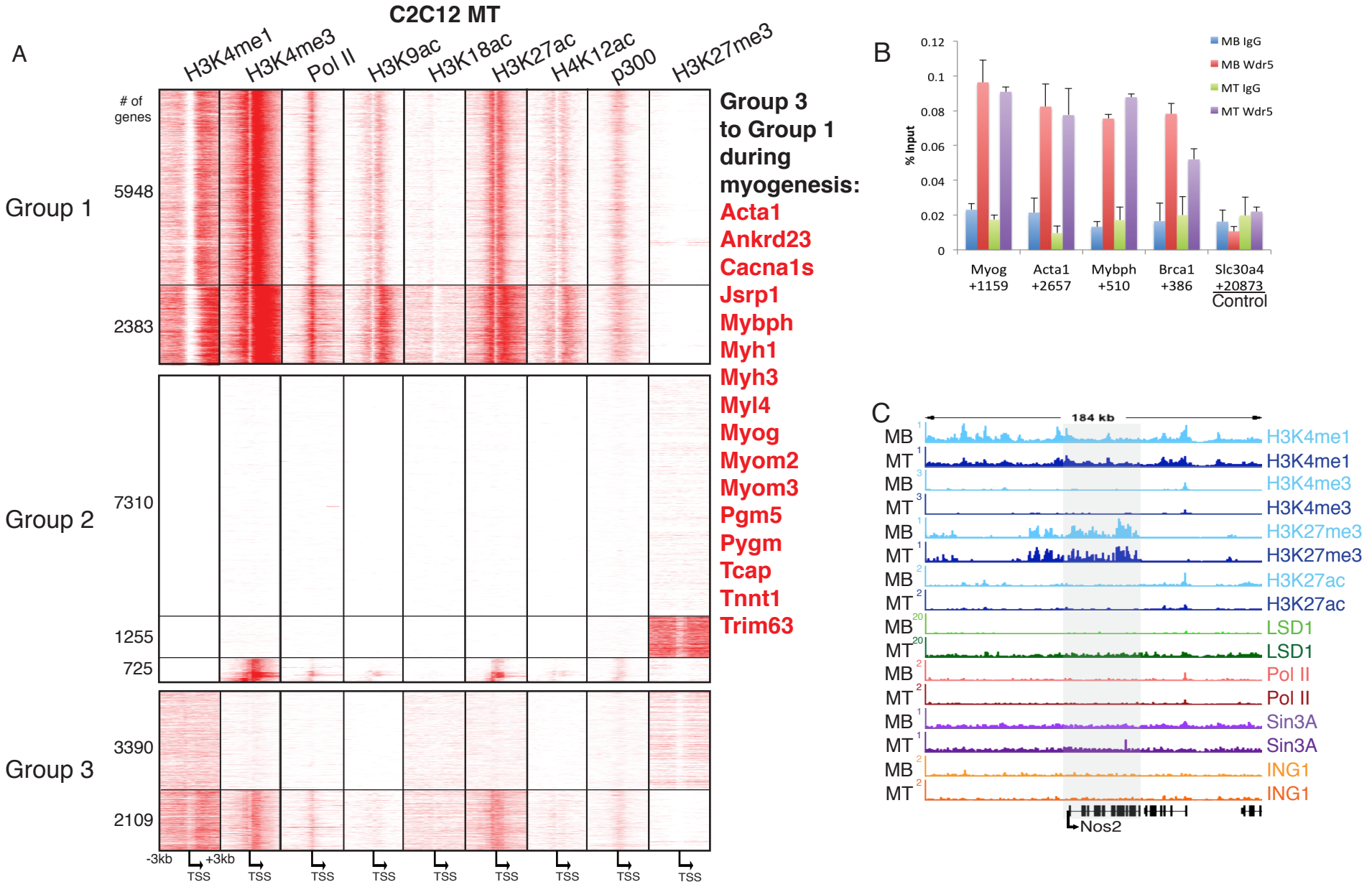
Cheng et al. Figure S2.



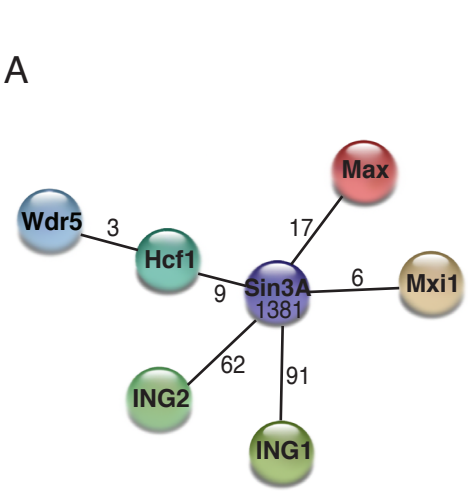
Cheng et al. Figure S3.

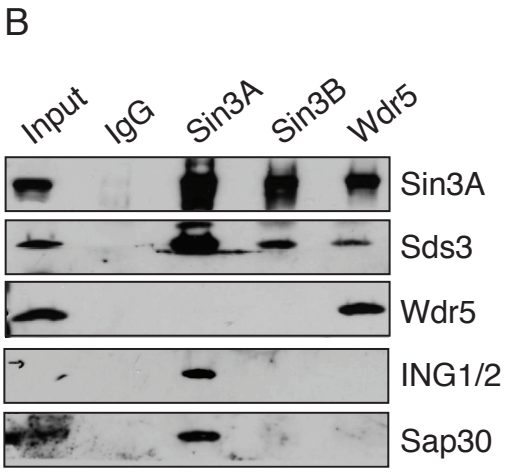


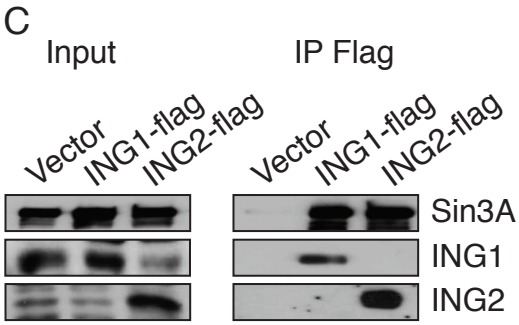
Cheng et al. Figure S4.

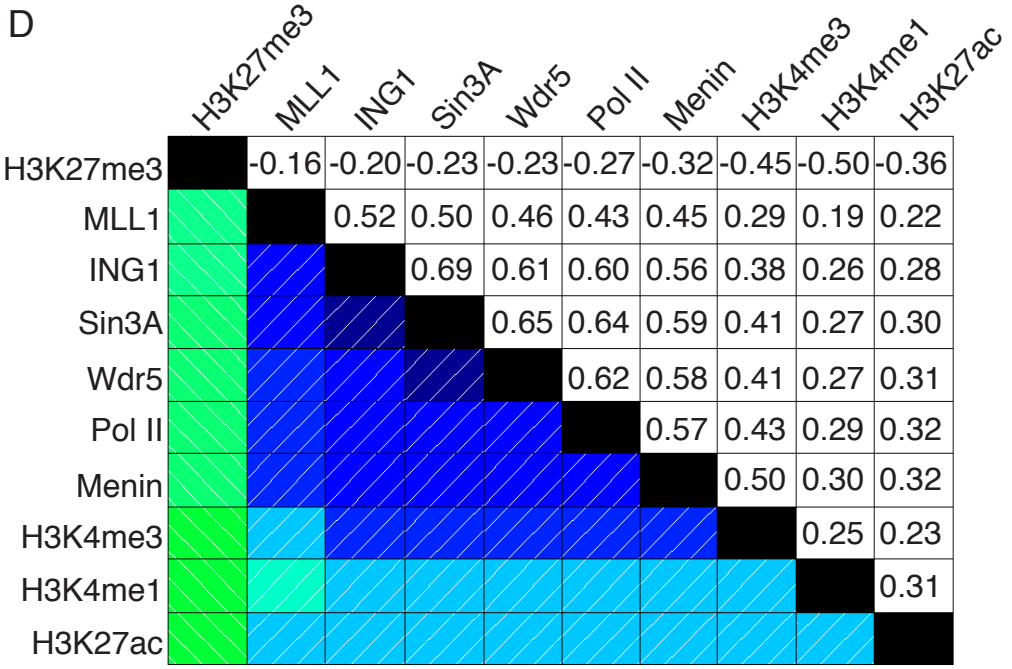


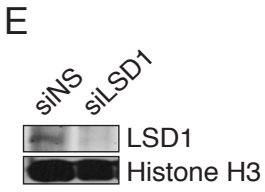
Cheng et al. Figure S5.



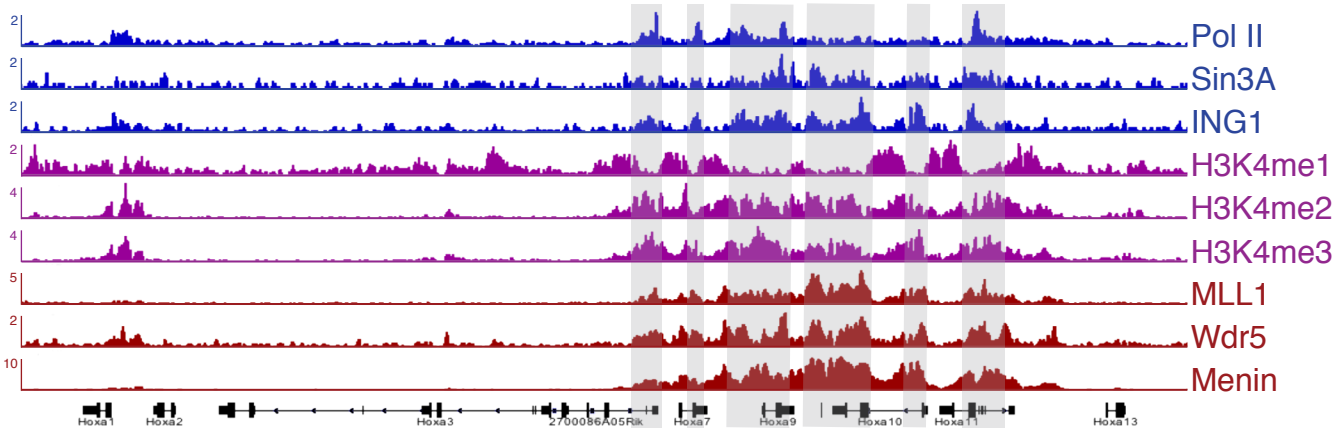




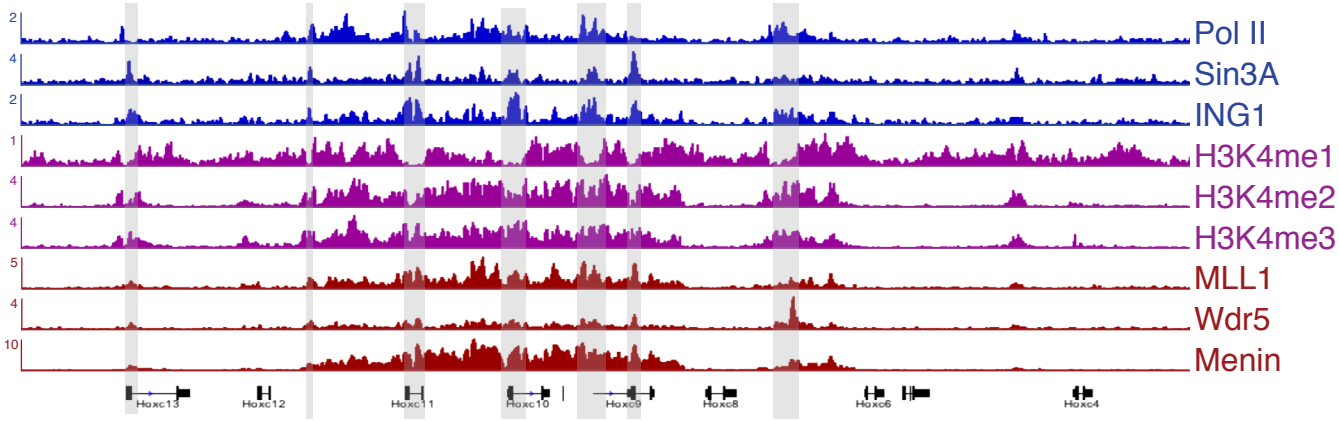




Cheng et al. Figure S6.

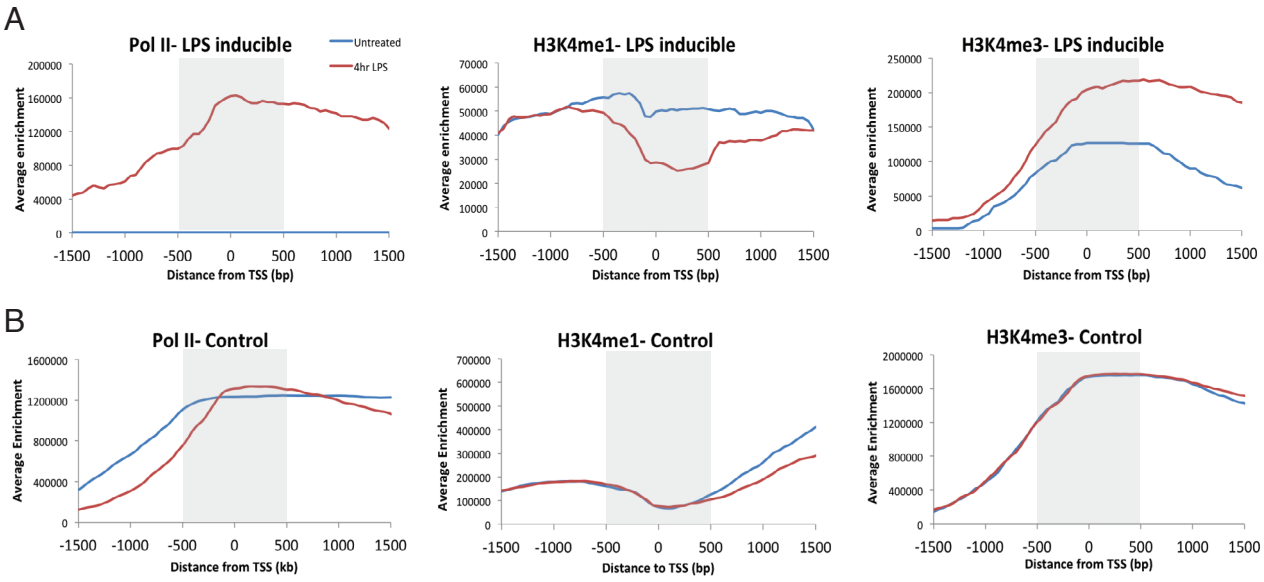


chr6:52,098,956-52,217,434



chr15:102,738,897-102,879,238

Cheng et al. Figure S7.



Supplemental Figure Legends

Figure S1. Localization of MLL3 using ChIP-seq. Related to Figure 1. (a) (*Left*) Primers were designed based on H3K4me1 and H3K4me3 enrichment, determined by ChIP-seq, as indicated. Primers for active (Group 1) genes amplified two locations: regions (A) within an H3K4me3 peak depleted of H3K4me1 or (B) downstream of the H3K4me3 peak that are highly H3K4 mono-methylated. (*Right*) Primers for repressed (Group 3) genes were within a peak of H3K4me1. Number following gene name indicates the 5' ends of the amplified regions relative to the TSS. **(b)** MLL3 ChIP-seq data were plotted on regions 3 kb upstream and downstream of the TSS of Group 1, Group 2, and Group 3 genes determined in Figure 1A. **(c)** H3K4me1, MLL3, and H3K27me3 enrichment by ChIP-seq was plotted for regions 3 kb upstream and downstream of the TSS of all Group 1, Group 2, and Group 3 genes determined in Figure 1A. The x- and y-axes represent the distance from the TSS and mean density (tags/200bp), respectively. **(d)** MLL3 co-localizes with H3K4me1 on Group 1 and 3 genes. Read density profile for MLL3 and H3K4me1 displays distinct peaks across Group 1 genes and broad localization of MLL3 and H3K4me1 across Group 3 genes, as compared to input chromatin. The y-axis corresponds to ChIP-seq signal density.

Figure S2. Role of MLL3 in gene repression in myoblasts and MEFs. Related to Figure 2. (a) Enrichment of MLL3 by ChIP is higher across repressed H3K4-mono-methylated (Group 3) genes as compared to active (Group 1) genes or constitutive heterochromatin. **(b)** RT-qPCR analysis to assess depletion of MLL3/4 with siRNAs. **(c)** COMPASS complexes are associated with gene activation and repression. RT-qPCR

analysis of H3K4-methylated genes after ablation of Wdr5. Expression was normalized to the non-specific control (siNS). **(d)** Loss of MLL3/4 has no effect on levels of H3K27me3, H4K20me1 **(e)** or H3K4me3 **(f)**. **(g)** H3K4me1 ChIP-seq (data from Hertz et al., 2012) from *MLL3*$^{-/-}$ and wild-type (WT) MEFs were compared. We identified a cluster of genes that showed decreased enrichment of H3K4me1 in the *MLL3*$^{-/-}$ cells compared to WT cells (cluster 1). The majority of H3K4-mono-methylated genes did not show a decrease in H3K4me1 in *MLL3*$^{-/-}$ MEFs compared to WT. For comparison, we show a second cluster (Cluster 2), which illustrates a representative subset of genes that showed minimal change in H3K4me1 in response to loss of MLL3. Expression data, from Shen et al., 2013 (right column) show expression levels for both clusters in WT cells. These data indicate that genes exhibiting a decrease in H3K4me1 upon ablation of MLL3 are expressed at low levels in WT cells. In contrast, genes that show no change in H3K4me1 exhibit a tendency toward significantly higher expression. **(h)** MLL3 qChIP across the promoter regions of select cluster 1 genes from Figure S2G in wild-type MEFs confirms that these genes are MLL3 targets. **(i)** Expression of cluster 1 genes from Figure S2G in *MLL3*$^{-/-}$ cells normalized to expression in wild-type MEFs. Genes that lose H3K4me1 in the *MLL3*$^{-/-}$ cells show increased expression compared to wild-type cells. ChIP data, expressed as percentage of input, are plotted on the y-axis. Error bars indicate SD. Student's t test was performed to indicate significance: * indicates p-values < 0.05 and ** <0.01, respectively.

Figure S3. H3K4me1 localizes to positioned nucleosomes and promoters with high nucleosome occupancy. Related to Figure 2. (a) Nucleosomes were mapped as described in Supplemental Experimental Procedures. Input chromatin is presented as unprocessed sequence tags (*left*, MNase-seq input). Data were similarly presented for nucleosomes enriched with the indicated antibodies (*top*). Nucleosome enrichment was plotted for regions 500 bp upstream and downstream of the TSS as indicated. Dashed vertical line indicates the TSS in each dataset. Promoters were clustered based on patterns of nucleosomes and sorted by expression with the most highly expressed gene in each cluster at the top. Expression is shown in the right-most column, with the indicated red-green scale (bottom) showing relative expression levels. (**b**) Sin3A and ING1 ChIP-seq data were processed as described in Supplemental Experimental Procedures. Binding site enrichment was plotted for the same regions as in **a**. Asterisks indicate the two nucleosomes downstream from the TSS as determined in **a**. **(c)** Nucleosome enrichment (MNase-Seq Input from **a** was plotted, after dividing genes into Groups 1, 2, or 3 from Figure 1A, for regions 3kb upstream and downstream of the TSS. The right-most panels present representative genes from Groups 1, 2, and 3 with a window of 1.5kb upstream and downstream of the TSS.

Figure S4. Inter-conversion of methylation states during differentiation. Related to Figure 3. (**a)** C2C12 myotube ChIP-seq data were plotted, and genes were sorted as in Figure 1A. Muscle development genes that switch from Group 3 in myoblasts to Group 1 in myotubes are indicated (*right*). **(b)** Persistent enrichment of Wdr5 across muscle genes (*Myog*, *Acta1*, *Mybph*), compared to *Slc30a4* (control), irrespective of H3K4 methylation

status. Background ChIP signal was determined using an antibody against IgG. **(c)** Read density profiles for H3K4me1, H3K4me3, H3K27me3, H3K27ac, LSD1, Pol II, Sin3A, and ING1 across the *Nos2* (indicated in gray) promoter in myoblasts and myotubes. The y-axis corresponds to ChIP-seq signal density. All ChIP data, expressed as percentage of input, are plotted on the y-axis. Error bars indicate SD. Student's t test was performed to indicate significance: * indicates p-values <0.05 and ** <0.01, respectively.

Figure S5. Sin3A interacts with ING1, ING2, and Wdr5. Related to Figure 4. (a) Mass spectrometric analysis of Sin3A-associated proteins. The peptide counts corresponding to a subset of Sin3A-associated proteins are indicated along each spoke. **(b)** Immunoprecipitations with antibodies against IgG (control), Sin3A, Sin3B, and Wdr5 were performed with nuclear extracts from C2C12 myoblasts. Western blots of immunoprecipitated fractions were probed with indicated antibodies specific for Sin3A, Sds3, Wdr5, ING1/2, and Sap30. Sds3 and Sap30 are known Sin3A-interacting proteins. (**c)** Immunoprecipitations with anti-Flag antibody were performed with extracts of C2C12 myoblasts transfected with vector (control), pBabe ING1-flag, or pBabe ING2-flag. Bound fractions were treated as in (**b**). (**d)** Correlation matrix reflects pairwise correlations of binding peaks between each factor or histone mark in our ChIP-seq data (see Methods). Numbers refer to the Pearson correlation coefficient (dark blue, positive correlation; dark green, negative correlation). **(e)** After ablation of LSD1 with siRNA, myoblast extracts were immuno-blotted for LSD1 and H3 (loading control).

Figure S6. Factor binding and methylation states on *Hox* clusters. Related to Figure 5. Read density profiles for Menin, Wdr5, MLL1, H3K4me3, H3K4me2, H3K4me1,

ING1, Sin3A, and Pol II across the *Hoxa* (*top*) and *Hoxc* (*bottom*) clusters in myoblasts. The y-axis corresponds to ChIP-seq signal density. Gray shaded regions represent regions bound by Sin3A.

Figure S7. Loss of H3K4me1 in response to LPS treatment in macrophages. Related to Figure 6. (a) We extracted data from Ostuni et al., 2013 and identified LPS-inducible genes (n=89) based on enrichment of Pol II after 4 hours LPS treatment. Pol II, H3K4me1, and H3K4me3 ChIP-seq signal was plotted for regions 1.5 kb upstream and downstream of the TSS in untreated (blue) and LPS-treated cells (red). The x-axis represents the distance from the TSS, and the y-axis represents average enrichment. Gray shaded regions represent TSS-proximal H3K4me1. (**b)** Enrichment of H3K4me1 across constitutively active genes (n=627, designated by binding of Pol II in both untreated and LPS-treated cells) is not affected by LPS treatment. Pol II, H3K4me1, and H3K4me3 were plotted as in **a**.

Supplemental Table S1. Myoblast gene groups. Related to Figure 1.

Supplemental Table S2. RNA-seq. Related to Figure 2.

Supplemental Table S3. Myotube gene groups. Related to Figure S3.

Supplemental Table S4. Genes sorted by expression. Related to Figure 4.

Supplemental Table S5. hESC gene groups. Related to Figure 6.

Supplemental Table S6. Primers.

Supplemental Experimental Procedures

Cells and Cell culture

C2C12 cells (Sigma) were cultured and induced to differentiate as described (Blais et al., 2007). After 96 hours (D4) in differentiation medium, myotubes were isolated from undifferentiated cells with diluted trypsin. Differentiation medium consisted of DMEM supplemented with 2% horse serum. Stable cell lines ectopically expressing Flag-tagged ING1 and ING2 were generated by cloning cDNA from Open Biosystems into pBabe-puro. Virus was produced by transfecting constructs into 293T cells. Viral supernatant was harvested and added with polybrene to C2C12 myoblasts. Myoblasts were then selected using puromycin.

Antibodies used in this study

1. Ab: H3K4me1; Source: ab8895, Abcam; Application: qChIP.
2. Ab: H3K4me3; Source: ab8580, Abcam; Application: qChIP.
3. Ab: H3K27me3; Source: ab6002, Abcam; Application: qChIP.
4. Ab: H3K9me3; Source: ab8898, Abcam; Application: qChIP.
5. Ab: H4K20me1; Source: ab9051, Abcam; Application: qChIP.
6. Ab: H4K20me3; Source: ab9053, Abcam; Application: qChIP.
7. Ab: Sin3A; Source: Kind gift from Dr. Donald Ayer; Application: ChIP-seq, qChIP, Western blot.
8. Ab: ING1/2; Source: Kind gift from Dr. Karl Riabowol; Application: Western blot.
9. Ab: ING1; Source: ab22664, Abcam; Application: ChIP-seq, qChIP, Western blot.
10. Ab: ING2; Source: sc-134973, Santa Cruz; Application: qChIP, Western blot.

11. Ab: Wdr5; Source: Kind gift from Dr. Danny Reinberg; Application: ChIP-seq, qChIP, IP, Western blot.

12. Ab: Sap30; Source: Kind gift from Dr. Donald Ayer; Application: Western blot.

13. Ab: MLL1; Source: A300-086A, Bethyl Laboratories; Application: ChIP-seq.

14. Ab: Menin; Source: A300-105A, Bethyl Laboratories; Application: ChIP-seq.

15. Ab: MLL3; Source: ab71200, Abcam; Application: qChIP.

16. Ab: LSD1; Source: 39186, Active Motif; Application: qChIP, Western blot.

17. Ab: MLL4 serum; Source: Dr. Ali Shilatifard (Hu et al., 2013); Application: ChIP.

ChIP-seq and downstream analysis

ChIP-seq libraries were cluster-amplified and sequenced with the Illumina HiSeq2000 sequencer (51-nucleotide single-ended read) according to the manufacturer's protocols as described in (Asp et al., 2011). For nucleosome mapping, MNase-seq was performed as follows. MNase-digested chromatin was sequenced, and the resulting reads were aligned to the mouse reference genome (assembly mm9) using the Burrows-Wheeler Alignment tool (BWA, ver. 0.7.3) (Li and Durbin, 2009), followed by extraction of uniquely aligned reads and removal of duplicates by SAMtools (Li et al., 2009). Enriched binding sites ("peaks") were determined by the peak-calling algorithm, Qeseq, by analyzing each ChIP library relative to its corresponding input control library, as previously described (Asp et al., 2011). BigWig files were generated by, first, extending the 5' ends of uniquely aligned, non-duplicate ChIP-seq reads by the average DNA fragment length (150bp for histone marks, 250bp for transcription factors) in the 3' direction using BEDtools (Quinlan and Hall, 2010). For each histone mark and transcription factor, the reads were

normalized per million total reads and processed with the UCSC Genome Bioinformatics script bedGraphToBigWig. Bigwig tracks are portrayed using the Integrative Genomics Viewer (IGV) (Robinson et al., 2011, Thorvaldsdóttir et al., 2013).

RNA-seq and downstream analysis

RNA concentrations were measured using a NanoDrop 8000 Spectrophotometer. 1 μg of total RNA was used for the library preparation, using the Illumina TruSeq RNA Sample Prep Kit (polyA selection-based), following the TruSeq RNA Sample Preparation V2 Guide. TruSeq adapters 1, 3, 8, 9, 10, and 11 were used during the ligation step. 15 cycles of PCR were run during the amplification step. The samples were mixed into one pool and run in one 50-nucleotide single-end read Rapid run flow cell with the Illumina HiSeq 2500 sequencer, to generate a total of 288 million reads passing filter.

All raw sequencing reads were mapped to the mouse genome (NCBI37/mm9) using Bowtie aligner (0.12.9) with v2 and m1 parameters. The mapped reads were subsequently sorted and filtered by removing the PCR duplicates with samtools (0.1.19) before further analysis. To generate the visualization files for IGV, each individual filtered bam file was converted to bigwig files with BEDtools (2.17.0) and the UCSC bigwig conversion script.

After removal of PCR duplicates, the reads were then assigned to the Ensemble gene model (Mus_musculus.NCBIM37.67.gff) (Morgan et al., 2009). The raw read counts were normalized using the trimmed mean of M-values normalization method (Robinson and Oshlack, 2010). The common dispersion and statistical significance for genes across sample groups were estimated and calculated using a generalized linear model (Robinson and Oshlack, 2010, Robinson et al., 2010). To obtain an adjusted p-value for each gene,

we applied the FDR method for multiple hypothesis testing to genes that have summed read counts (from all samples) greater than 10.

Assembly of heat maps and clustering of transcription factors and histone marks

For mapping transcription factors and histone marks, we used our Qeseq-filtered enriched tags. To represent each of the enriched tags, their positions were extended by 5 bp in both the 5' and 3' directions. When we employed enriched-peak data from the Encyclopedia of DNA Elements (ENCODE) project (ENCODE Project Consortium et al., 2011), we divided each of the peaks into 10 bp segments. Using the "reads" option in seqMINER, we loaded the 10 bp tags and set the "wiggle" step to 10 to view all tags localized within a 6 kb window surrounding the TSS (promoter regions) of all analyzed Refseq genes. Strand orientation was accounted for, and all analyzed features were therefore presented in the same direction relative to each gene's TSS. The tags were subjected to K-means clustering. Heat map representations of gene expression levels (Liu el al., 2010) were then generated by employing TreeView (v.1.1.6r2) (Saldanha 2004).

Average enriched distribution of histone marks and transcription factors

To study (1) the enrichment of H3K4me1, MLL3, and H3K27me3, (2) the dynamics of LSD1, Sin3A, and LSD1 during myogenesis and, (3) the pattern of histone marks and transcription factors on Sin3A target genes, we first divided genes into defined groups according to the enrichment of the indicated factors across TSS regions (6 kb window surrounding the TSS). Tags were annotated to transcription start sites, and the accumulation of total enrichment values per 50 bp or 200bp bins were analyzed by

implementing in-house Perl scripts. Next, the total enrichment values for all annotated tags were identified by Qeseq as peaks; total enrichment values were then divided by the number of genes in each group. Therefore, plots indicate average enrichment per gene.

Co-localization of binding sites of transcription factors and histone marks

Enriched ChIP-seq regions at promoters (6 kb window surrounding the TSS) for each of the transcription factors and histone marks were combined together to generate a unified track consisting of all merged enriched regions. The presence or absence of each transcription factor and histone mark for each merged, enriched region in the unified track was tabulated, yielding a binary co-localization vector for each of the factors or histone marks. Pearson correlation coefficients for each pair of co-localization vectors were calculated and used as a similarity measure to cluster the transcription factors and histone marks using the default distance definition and clustering algorithm in the Corrgram function of R.

Nucleosome mapping

To visualize nucleosome positions, we applied the alignment correction as described in Li et al., 2012. Briefly, the 5' ends of each read were shifted towards its 3' end by half the estimated nucleosome length (78bp), which was then extended by 50 bp in both the 5' and 3' directions to generate a 100 bp shifted sequence. The newly shifted sequence, representing the nucleosome position, was used in all downstream analyses for mapping nucleosomes and histone marks (Asp et al., 2011). For all the mm9 Refseq genes, we used seqMINER to extract the sequences in a 1 kb or 3 kb window surrounding the TSS

and generated heat maps as well as mean tag density distributions (Ye et al., 2011). Using the “reads” option, we loaded 100 bp reads, set the wiggle step to 10, and performed all subsequent steps to generate heat maps as described above.

Supplemental References

Asp P, Blum R, Vethantham V, Parisi F, Micsinai M, Cheng J, Bowman C, Kluger Y, Dynlacht BD. (2011) Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. Proc Natl Acad Sci USA. *108*, 149-158.

Blais A, van Oevelen CJ, Margueron R, Acosta-Alvear D, Dynlacht BD (2007) Retinoblastoma tumor suppressor protein-dependent methylation of histone H3 lysine 27 is associated with irreversible cell cycle exit. J Cell Biol *179*, 1399–1412.

ENCODE Project Consortium, Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, Birney E et al. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. *9*, e1001046.

Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A. (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. *26*, 2604-2620.

Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. (2013) The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. Mol Cell Biol. *33*, 4745-54.

Li H and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. *25*, 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics. *25*, 2078-2079.

Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH. (2012) Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation.

Cell. *151*, 1608-1616.

Liu Y, Chu A, Chakroun I, Islam U, Blais A. (2010) Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation. Nucleic Acids Res. 38, 6857-6871.

Morgan M, Anders S, Lawrence M, Aboyoun P, Pagès H, Gentleman R. (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics. *25*, 2607-2608.

Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Quinlan AR and Hall IM. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. *26*, 841-842.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011) Integrative genomics viewer. Nat Biotechnol. *29*, 24-26.

Robinson MD, Oshlack A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. *26*, 139-140.

Saldanha AJ. (2004) Java Treeview--extensible visualization of microarray data. Bioinformatics. *20*, 3246-3248.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B. (2012) A map of the cis-regulatory sequences in the mouse genome. Nature. *488*, 116-20.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. (2013) Integrative Genomics Viewer

(IGV): high-performance genomics data visualization and exploration. Brief Bioinform. *14*, 178-192.

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