SUPPLEMENTARY INFORMATION

EXTENDED SUPPLEMENTARY METHODS

Cell Culture and Generation of Stable Cells

HMEL-BRAF^{V600E}, PMEL-BRAF^{V600E}, HEK-293T cells were grown in 5% CO_2, at 37^{0}C in DMEM medium with 10% FBS. Cells were routinely tested for mycoplasma infection. Viral production was done in HEK-293T cells using pHIT60 and VSVg. Stable knockdown of GFP or PTEN in early passage (n < 10) HMEL-BRAF^{V600E}, PMEL-BRAF^{V600E} cells was performed using pMKO-shGFP or pMKO-shPTEN vectors (Addgene) to create NTM_H (HMEL-BRAF^{V600E}-shGFP), T_H (HMEL-BRAF^{V600E}-shPTEN), NTM_P (PMEL-BRAF^{V600E}-shGFP) and TM_P (PMEL-BRAF^{V600E}-shPTEN) cells.

ChIP-Seq

Chromatin immunoprecipitation was performed as described earlier (Garber et al., 2012) with optimized shearing conditions and minor modifications for melanocytes. Briefly, cells (5 million per antibody) were cross linked using 1% paraformaldehyde for 10mins at 37^{o}C. Reaction was quenched by 0.125M glycine for 5mins, and cells washed with PBS and stored at -80^{o}C. Next day cells were thawed on ice and lysed with RIPA buffer (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 140mM NaCl, 1% Triton x-100, 0.2%SDS, 0.1% DOC) for 10min on ice. Sonication conditions were optimized for HMEL-BRAF^{V600E} cells and were performed using Branson Sonifier 250 to achieve shear length of 250-500bp. Extracts were then incubated overnight with respective antibody-dynabead mixtures that were incubated separately for 1hr at 4^{o}C earlier. Immunecomplexes were then washed in following order: 5 times with RIPA buffer, twice with RIPA-500 (RIPA with 500mM NaCl), twice with LiCl wash buffer (10mM Tris-HCl pH8.0, 1mM EDTA pH8.0, 250mM LiCl, 0.5% NP-40, 0.1% DOC) and once with TE (10mM Tris-HCl, 1mM

EDTA). Elution and decrosslinking was performed in direct elution buffer (10mM Tris-Cl pH8.0, 5mM EDTA, 300mM NaCl, 0.5% SDS) by incubating immunecomplexes at 65^{o}C overnight. Proteinase K (20mg/ml) and RNaseA treatment was performed and DNA cleaned up using SPRI beads (Beckman-Coulter). Library preparation was done as described in (Garber et al., 2012) using paired end adapters from IDT. Libraries were multiplexed together and sequencing was performed in Hiseq2000 (Illumina). Antibody details are below:

Mark	Company	Catalog Number
H2AK5ac	Abcam	ab45152
H2BK120ac	Active Motif	39119
H2BK15ac	Abcam	ab62335
H2BK5ac	Active Motif	39123
H3	Abcam	ab1791
H3K14ac	Millipore	07-353
H3K18ac	Abcam	ab1191
H3K23ac	Millipore	07-355
H3K27ac	Abcam	ab4729
H3K27me1	Millipore	07-448
H3K27me3	Abcam	ab6002
H3K36ac	Active Motif	39379
H3K36me1	Abcam	ab9048
H3K36me2	Abcam	ab9049
H3K36me3	Abcam	ab9050
H3K4ac	Millipore	07-539
H3K4me1	Abcam	ab8895
H3K4me2	Abcam	ab32356
H3K4me3	Abcam	ab8580
H3K79me1	Abcam	ab2886
H3K79me2	Abcam	ab3594
H3K79me3	Abcam	ab2621
H3K9ac	Abcam	ab4441
H3K9me1	Abcam	ab8896
H3K9me2	Abcam	ab1220
H3K9me3	Abcam	ab8898
H4	Millipore	05-858
H4ac4	Active Motif	39179
H4K12ac	Active Motif	39165
H4K16ac	Millipore	07-329

H4K20me1	Abcam	ab9051
H4K5ac	Millipore	07-327
H4K8ac	Abcam	ab15823
H4K91ac	Abcam	ab4627
5-hmC	Active Motif	39769
H4K20me2	Abcam	ab9052
H4K20me3	Abcam	ab9053

ChIP-Seq Data Analysis:

ChIP-Seq reads were aligned using Bowtie (version 1.0.0) (Langmead et al., 2009) to human genome assembly NCBI Build 37 (UCSC hg19) with the following parameters: -n 1 -m 1 --best –strata (uniquely mapped reads with one mismatch were retained). First 36bp from 5' end of the reads were retained in case read lengths are longer than 36bp for any given histone modifications. To avoid biases due to PCR artifacts, sequencing reads that map to the same genomic location and strand were counted once in the input data.

Peak calling was performed using MACS algorithm (Zhang et al., 2008) with default parameters except a p-value cut-off 10E-8 applied. DiffBind bioconductor package was used to cluster histone marks by using identified peaks with MACS algorithm.

We generated signal tracks at 200bp resolution, by partitioning the genome into non-overlapping bins at that resolution. We calculated signal values over all bins for each histone mark using the following formula:

$$Signal_i = \frac{K_i * 10^9}{L_i * N}$$

where $Signal_i$ is the signal value of a given histone mark at bin *i*, K_i is the raw number of sequencing reads for that mark that span bin *i* after extending each read by 200bp from the start in the direction of the alignment, L_i is the length of bin *i*, and *N* is the total number of sequencing reads for that mark.

ChromHMM Analysis

We used ChromHMM (Ernst and Kellis, 2012) with default parameters to derive genome-wide chromatin state maps for all cell types. We binarized the input data with ChromHMM's BinarizeBed method using a p-value cutoff of 1e-4. We observed that the total number of binary presence calls was very similar between NTM_H and TM_H. However, total number of calls was higher in NTM_P compared to T_P. Thus, to reduce the effect of potential technical confounders, we normalized each chromatin mark in NTM_P and TM_P to have the same number of binary presence calls across these two cell types. To achieve that, we first used the BinarizeBed option of ChromHMM on all datasets from NTM_P and TM_P. Then, we sorted the binary calls for each chromatin mark in each bin according to the number of reads assigned and kept only the top N binary calls, where N is the smaller of the total numbers of binary calls for the corresponding chromatin mark in NTM_P and TM_P. In the case of ties, we dropped randomly binary calls that had the same number of sequencing reads assigned in order to arrive at equal number of binary calls across the two cell lines for the corresponding chromatin mark.

We considered chromatin state models learned jointly on all chromatin marks from NTM_H and TM_H ranging from 10 to 120 states. Two models were considered for additional analysis: 18-state model (with the minimum number of states that had a separate state containing likely artifactual signal locations) and 45-state model (with the minimum number of states that contains a clear poised/bivalent state). We chose to focus on a model with 18 states for our main analysis to balance capturing informative state distinctions while maintaining interpretability and having a manageable number of pairwise state transitions. In particular the model with 18 states was the model with the

minimum number of states that had a separate state containing likely artifactual signal locations. The chromatin state annotations of NTM_H, TM_H, NTM_P and TM_P was produced subsequently by applying this model to the chromatin data from these cell types.

Analysis of Chromatin State Changes

To find important chromatin state changes between non-tumorigenic and tumorigenic cell lines, we intersected the chromatin state annotations of NTM_H and TM_H, and of NTM_P and TM_P, respectively. In each case, we counted the number of 200bp bins that is occupied by each of the 18 by 18 possible chromatin state transitions. To calculate enrichment scores, we divided this number by the expected number of such bins assuming a null model that treats the two chromatin states involved in each transition as independently distributed. Finally, to control for state similarity between each pair of chromatin states *i* and *j*, we divided the enrichment score of transitioning from state *i* in non-tumorigenic cells to state *j* in tumorigenic cells by the enrichment score of transitioning from state *j* in non-tumorigenic cells to state *i* in tumorigenic cells. In order to avoid division by 0 in cases where no overlap was detected between pairs of chromatin states, we added a pseudo-count of 1 bin to each intersection before we computed all enrichments, enrichment ratios and p-values.

Besides our main analysis, we performed the above computations under two other normalization schemes. First, we downsampled randomly the number of sequencing reads for each chromatin mark to the minimum number across NTM_H and TM_H, and across NTM_P and TM_P, respectively. We applied the previously learned 18 states model on the downsampled data and ran the above analysis pipeline on the produced chromatin state annotations. In our second normalization scheme, we downsampled the number of binary calls from ChromHMM's BinarizeBed routine for each chromatin mark

to the minimum number across all four cell types in the same way we did previously for NTM_P and TM_P. Again, we applied the 18 state model to this data and ran the above analysis pipeline.

Analysis of Chromatin State Recovery with Subsets of Marks

For the analysis of the chromatin state recovery with subset of marks relative to using all marks we used the EvalSubset of ChromHMM command(Ernst and Kellis, 2012) (Ernst and Kellis, 2015) applied to the chromatin state annotations of NTM_H and TM_H. For this analysis we separately evaluated for each mark, recovery based on only that mark and using all marks except that mark.

Analysis of Individual Mark Enrichments at Promoters and DNaseI hypersensitive sites

Promoter regions were defined as 4kbp regions centered at annotated transcription start sites from RefSeq (as downloaded on March 2014 from UCSC Genome Browser). As for DNaseI hypersensitive sites (DHS), we downloaded a data set with DNaseI peaks for the Melano cell type from the ENCODE project (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeOpenChromDnase/wgEncodeOpenChromDnaseMelanoPk.narrowPeak.gz). To define distal sites, we further excluded peaks whose midpoints are within 4kb of annotated transcription start sites in hg19. To compute the histone mark signal over the remaining sites, we extended them by 2kb from their midpoints in both directions. For each promoter region or distal DHS *i*, chromatin mark in cell type *c*, we calculated the signal strength in RPKM as:

$$\mathrm{RPKM}_{i,c} = \frac{K_{i,c} * 10^9}{L_i * N_c}$$

where $K_{i,c}$ is the number of sequencing reads from that mark in cell type *c* whose center position overlaps with region *i* after extending each read by 200bp in the direction of the alignment, L_i is the length of region *i*, and N_c is the total number of reads for the mark. We then calculated the average fold change of each mark at promoters and DHS separately by summing over all regions for them as:

$$\Delta = \log_2 \left(\frac{\sum_i \mathrm{RPKM}_{i,c_1}}{\sum_i \mathrm{RPKM}_{i,c_2}} \right),$$

Where c_1 and c_2 are NTM_H and TM_H or NTM_P and TM_P respectively.

Differentially Acetylated Promoters

To identify statistically significant differences of acetylations in aggregate at promoters, we compared to a null model in which the non-tumorigenic and tumorigenic label for each acetylation was randomly permuted. Specifically, for both promoters in each cell type separately we calculated an average acetylation level in each region *i* by taking the mean RPKM value across all acetylation marks for the given cell type *c*, denoted by $\mathrm{RPKM(Ac)}_{i,c}$. Next, we calculated the change in the average acetylation levels at region *i*:

$$\Delta(\mathrm{Ac})_i = \log_2 \left(\frac{\mathrm{RPKM(Ac)}_{i,c_1} + 1}{\mathrm{RPKM(Ac)}_{i,c_2} + 1} \right),$$

where c_1 and c_2 are NTM_H and TM_H, or NTM_P and TM_P, respectively. To determine significant changes at a FDR of 1% we used a null model based on 100 randomized pairs of cell types for each system (NTM_H / TM_H and NTM_P / TM_P). Each randomized pair was generated by iterating through all acetylation datasets from NTM_H and TM_H (or NTM_P and TM_P) and randomly switching their labels with probability of 0.5. Based on the randomized data we constructed a background distribution of the $\Delta(\mathrm{Ac})_i$ values across all intervals and randomizations, which we used to calculate two-sided P-values for all

observed $\Delta(\mathrm{Ac})_i$. We then applied the Benjamini–Hochberg procedure on these P-values to derive cutoffs at FDR of 1%.

Promoter State Analysis

Every RefSeq gene was assigned to one chromatin state based on the state call on the gene's TSS in non-tumorigenic (NTM_H, NTM_P) and tumorigenic (TM_H, TM_P) cell types by using the 18 state ChromHMM genome annotation output. STEM software (Ernst and Bar-Joseph, 2006) was used to analyze enriched gene ontology (GO) terms for the genes that are changing their promoter states between non-tumorigenic (NTM_H, NTM_P) and tumorigenic (TM_H, TM_P) cells. Default settings changed to only reporting GO-Biological Process (BP) terms, with equal or below level 5 according to GO taxonomy. STEM software output is processed as following: BP-terms that are enriched for a state-transition with a p-value of less than 10^{-4} and at least 3 genes was assigned for that specific state-transition regarding that particular term is retained. To estimate an overall false discovery rate, we generated random gene sets by keeping the number of genes per state-transition constant but randomly assigning genes from RefSeq annotation table. We did not identify any enriched GO-terms for randomized promoter state-transition pairs with the explained filtering steps. Identified state-transitions and BP-terms were used for heat map generation (Figure 4A).

Pathway Analysis

Pathway Commons analysis on the enriched genomic regions was done using GREAT tool (McLean et al., 2010) (www.great.stanford.edu). For promoter state regions, we used the basal + extension option with -2Kb to +2Kb proximal to TSS and 20Kb extension. For enhancers we used the option of 'single nearest gene" with 1000Kb extension.

RNA-Seq Analysis

Strand specific libraries were constructed using a strand specific method (Levin et al., 2010). Reads were mapped to the human genome (hg19) using Mapsplice algorithm version 2.1.4 (Wang et al., 2010). We first merged the annotations of UCSC gene annotation in Illumina's iGenomes (available at http://cole-trapnell-lab.github.io/cufflinks//igenome_table/index.html) gtf file with the recent long non-coding RNA annotation file (Kelley and Rinn, 2012) using GFFRead tool as part of Cufflinks suite (Trapnell et al., 2013) http://cole-trapnell-lab.github.io/cufflinks. Transcript expression was estimated using Cuffdiff 2.11 with the following option: ''--library-type firststrand'' against the merged annotation file. We then applied Cuffmerge 2.11 based on published protocol (Trapnell et al., 2012) for merging all identified transcripts in each replicates and generated master GTF file used for differential expression analysis. Cuffdiff 2.11 was run with the following options: ''--library-type firststrand, --min-reps-for-js-test 2, –dispersion-method per condition'' and transcripts with less than 0.05 q-values called as differentially expressed (snoRNAs removed from the differentially expressed transcripts list). A transcript was designated as protein coding if it could be assigned to a protein ID using UCSC table browser, rest of the transcripts referred as non-coding. For the up-regulated, down-regulated or unchanged genes, we calculated occurrence of every possible combination of state transitions on TSS, or within -2Kb or +2Kb range. Log2 fold changes calculated based on observed versus expected number of state transitions. Expected number of state transitions was calculated by multiplying all observed transitions within each range (TSS, -2Kb and +2Kb) with the number of up-regulated, down-regulated or unchanged genes then diving with the total RefSeq gene number.

Overlap of average acetylation and transcriptomic data

We systematically overlapped gene expression changes with changes in promoter acetylation to define the nine possible subsets (Figure 5D, S6H): (1) deacetylated-promoters with no corresponding gene-expression changes (LossAc_ConstExp), (2) deacetylated-promoters accompanied with corresponding downregulated gene-expression changes (LossAc_LossExp), (3) deacetylated-promoters accompanied with corresponding upregulated expression changes (LossAc_GainExp), (4) promoters that do not change their acetylation levels but are downregulated at the expression level (ConstAc_LossExp), (5) promoters that do not change their acetylation levels but are upregulated at the expression level (ConstAc_GainExp), (6) acetylation gaining promoters with no corresponding gene-expression changes (GainAc_NoExp), and (7) acetylation gaining promoters accompanied with corresponding upregulated gene-expression changes (GainAc_GainExp). Of the remaining two subsets, one (GainAc_LossExp) was an empty set, while the other set contained only unchanged loci (ConstAc_ConstExp).

DNA Methylation Analysis

We utilized Illumina Infinium HumanMethylation450 BeadChip arrays to profile DNA methylation profiles in NTM_H and TM_H cell lines. The Illumina Infinium HumanMethylation450 BeadChip covers over 450,000 CpG sites in the human genome. We processed the HumanMethylation450 images by Illumina's GenomeStudio Methylation Module software to calculate average beta values for each probes. Later, we used IMA (Illumina Methylation Analyzer) Bioconductor package (Wang et al., 2012) to identify average methylation of CpGs in triplicates of NTM_H and TM_H cells. We removed

the sites with missing beta values and performed quantile normalization and peak correction (Dedeurwaerder et al., 2011).

In addition, we utilized 5-hmCDIP-Seq assay (performed at Active Motif) to identify enriched locations for 5-hydroxymethyl cytosines for non-tumorigenic (NTM_H) and tumorigenic (TM_H) cell lines. Libraries were sequenced as 50bp single-end reads and mapped to the genome using bowtie as mentioned earlier. Peak calling was performed using MACS algorithm with whole cell extract as negative control, and a p-value cut-off of 10^{-10}.

ChIP-String Experiments

We conducted ChIP-string experiments for H2BK5ac, H4K5ac, H3K27ac, H3K4me1, H3K4me3 and H3K27me3 marks in 4 nevi and up to 9 melanoma tumors as well as in NTM_H and TM_H cells on a custom ChIP-string array. These histone marks were chosen to test representative regions from three groups: promoters, enhancers and Polycomb-repressed regions. Since space on the array was limited to 96 probes, we aimed to prioritize marks and regions that are most differential based on the ChIP-seq signal between the tumorigenic and non-tumorigenic cells within each of the three groups. The tested marks were selected based on a combination of prior knowledge about their association with each type of regulatory region and findings in our ChIP-seq data. Initially, we selected H3K27me3 to test Polycomb repressed regions, H3K4me1 to test enhancers, H3K4me3 to test promoters, and H3K27ac to test both enhancers and promoters as these marks are known to correlate with the respective regulatory types. To increase our mark coverage, we further sought to select additional marks that could be tested on the same probes for differential enrichment between NTM and TM cells. By inspecting the top differential regions for pairs of marks, we identified H2BK5ac as a candidate mark that can differentially enrich with H3K27ac, and H4K5ac with H3K4me1.

The 96 probes were split equally in four parts to test regions for differential enrichment of H3K27me3, H3K4me3, H3K27ac together with H2BK5ac, and H3K4me1 together with H4K5ac. In particular, 24 probes were designed for each mark or pair of marks, half of which (12 probes) were chosen to be consistently differentially enriched in both non-tumorigenic cell lines (NTM_H and NTM_P) for the histone mark or pair of marks, and the other half were chosen to be consistently differentially enriched in both tumorigenic cell lines (TM_H and TM_P). This symmetric design allows for a natural positive and negative control of each experiment, because a properly classifiable sample would show positive ChIP-string signal in precisely one of the two groups and no signal in the other group.

To select genomic regions for each mark or pair of marks, we first divided the genome into non-overlapping bins of 200 bp and computed RPKM values for each bin. We then sorted in ascending order all bins by the ratios in their ChIP-seq signal between NTM and TM cells (the smaller of (NTM_H / TM_H) and (NTM_P / TM_P)). For regions tested on pairs of marks, we sorted the bins by the smaller of the ratios of the two marks. We further required that selected bins undergo a transition from one chromatin state to a sufficiently different chromatin state (e.g. bins annotated as promoters in NTM cells transitioning to low signal or to Polycomb repressed in TM cells were allowed, but promoter bins transitioning to other types of promoters were excluded). The chromatin states were defined based on a ChromHMM model learned from a subset of our final ChIP-seq data, which was available at the time the ChIP-String array was commissioned (in the final dataset, a file for H3K4me3 in NTM_P cells was replaced due to a mislabeling issue). Additionally, we required that a binary presence call was made by ChromHMM's BinarizeBed procedure in the cell type the signal was considered enriched in and no binary presence calls were made within 2 kb of the bin for the same mark in the other

cell type. Finally, bins within 2 kb of higher scoring bins were excluded. The probes for the ChIP-string array were designed from the top and bottom parts of the sorted list for either gain or loss of signal, respectively, between NTM and TM cells that pass all of the above criteria. Bins that presented technical problems for probe design were replaced with the next possible bin from the corresponding sorted list. The genomic coordinates of all selected regions for each mark used in the final design of the ChIP-string array are listed in Table S2.

ChIP-String Data Analysis

Raw probes values were first normalized by the same method as the one used by Ram et al. (Ram et al., 2011). Counts for all probes of each sample were then compared to negative controls, and samples in which greater than 90% of probes were at or below background level counts (based on inbuilt negative controls) were omitted from further analysis. Counts derived from each ChIP sample were then normalized as follows. Probes for each individual sample were divided by the median count within the sample, then each probe was divided by the median value of that probe across all samples. The mean and standard deviation were calculated per sample. The mean value per sample was subtracted from each probe and then each probe was divided by the standard deviation. The resulting values were subsequently used for the analysis.

Aggregate Plots

Genome-wide coverage files (bigWig) for each H3K27Ac experiment was generated by using bamCoverage function of deepTools (Ramirez et al., 2016), with Reads Per Kilobase per Million mapped reads (RPKM) normalization. Then, obtained coverage tracks used for aggregate plots of H3K27Ac levels around -/+ 2Kb of de-acetylated promoters with visualization tool – ChAsE (Younesy et al., 2016).

Calculation of IC_{50} Values

Cells were plated in 96-well plates and treated in six replicate wells. The images were obtained and confluence was calculated by the Incucyte machine (Essen biosciences) and associated software. The confluence data was then used for calculation of drug response. Drug-response data was adjusted to a four-parameter log-logistic function using the R package drc. IC_{50} were predicted using the derived model. The area under the curve (AUC) was obtained by numerical integration of cell viability in function of dose (log10 scale) using Bolstad R package.

Mass Spectrometry

Total histones were prepared and subject to mass spectrometry analysis as previously described (Karch et al., 2014) using the LTQ-Orbitrap Velos Pro (Thermo Scientific).

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Cell line based model of melanoma progression and epigenome profiling, Related to Figure 1. (A) Validation of PTEN loss by PTEN shRNA by western blot**. (B-C)** Soft agar colony formation ability of NTM_H, TM_H, NTM_P and TM_P cells. Panel B shows representative image and panel C shows the quantitation of soft-agar colonies. **(D-E)** Matrigel-invasion ability of NTM_H, TM_H, NTM_P and TM_P cells. Panel D shows representative image of invaded cells post Boyden chamber assay and panel E shows the quantitation of invaded cells. **(F)** Log_2ratio between NTM_P and TM_P cells for the average signal strength of each chromatin mark in a window of 2kb around annotated transcription start sites from RefSeq (Blue) and on distal DNaseI hypersensitive sites from 'Melano' cell lines (Red, See Supplementary Methods) from ENCODE. **(G-I)** Measurement of global levels of histone modification marks in NTM_H and TM_H cells. (G-H) Mass Spectrometry based quantitation of various histone marks on histone H3 (G) or histone H4 (H). X-axis shows peptide identity whereas Y-axis shows relative abundance. (I) Western blot analysis for indicated histone marks from acid-extracted histones from NTM_H and TM_H cells.

Figure S1

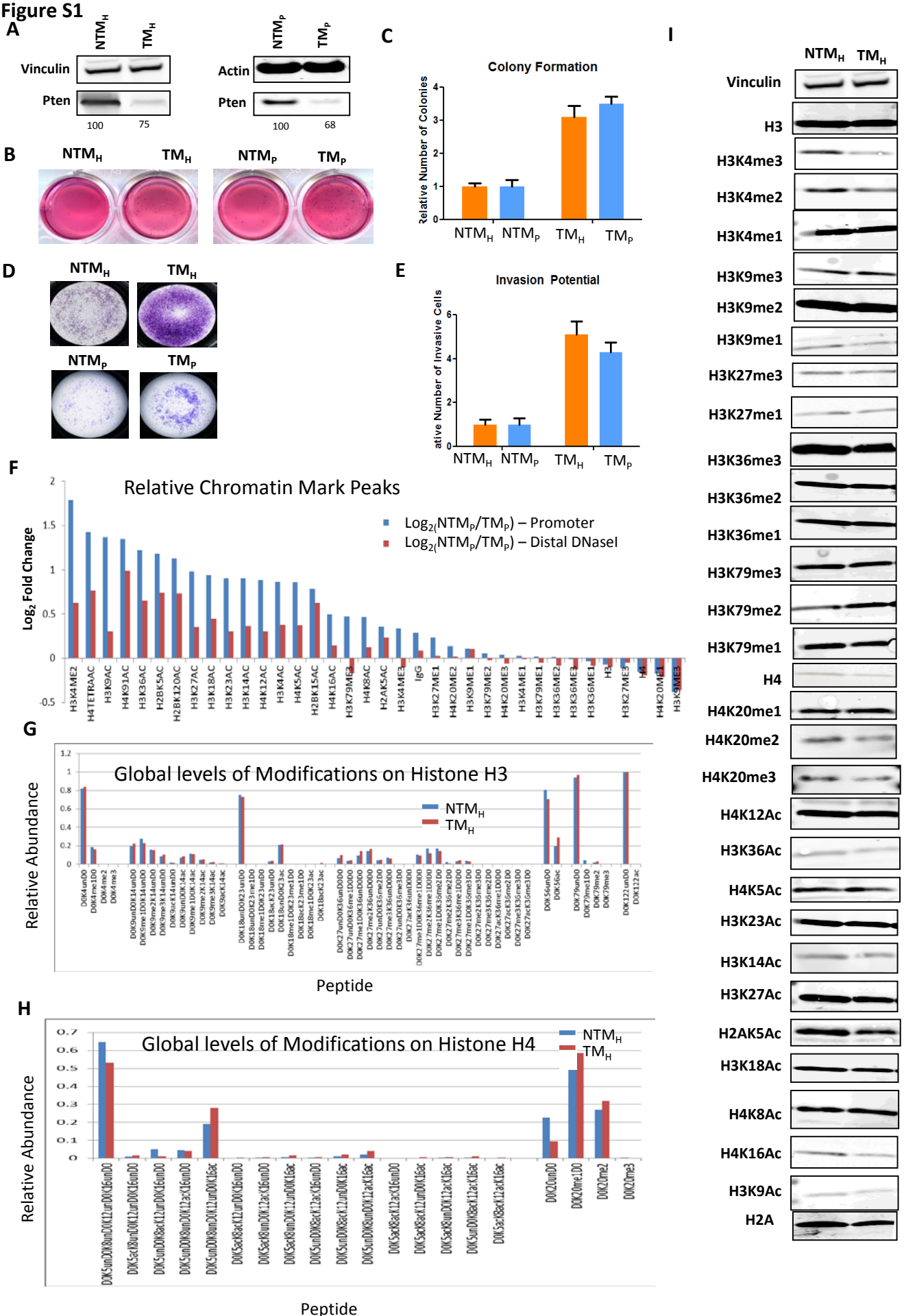

A
NTM_H
TM_H
NTM_P
TM_P
Vinculin
Pten
100
75
Actin
Pten
100
68
B
NTM_H
TM_H
NTM_P
TM_P
C
Colony Formation
Relative Number of Colonies
NTM_H
NTM_P
TM_H
TM_P
D
NTM_H
TM_H
NTM_P
TM_P
E
Invasion Potential
ative Number of Invasive Cells
NTM_H
NTM_P
TM_H
TM_P
F
Relative Chromatin Mark Peaks
Log2(NTMP/TMP) – Promoter
Log2(NTMP/TMP) – Distal DNaseI
Log2 Fold Change
H3K4ME2
H4TETRAAC
H3K9AC
H4K91AC
H3K36AC
H2BK5AC
H2BK120AC
H3K27AC
H3K18AC
H3K23AC
H3K14AC
H4K12AC
H3K4AC
H4K5AC
H2BK15AC
H4K16AC
H3K79ME3
H4K8AC
H2AK5AC
H3K4ME3
IgG
H3K27ME1
H4K20ME2
H3K9ME1
H3K79ME2
H4K20ME3
H3K4ME1
H3K79ME1
H3K36ME2
H3K36ME3
H3K36ME1
H3
H3K27ME3
H4
H4K20ME1
H3K9ME3
G
Global levels of Modifications on Histone H3
NTM_H
TM_H
Relative Abundance
Peptide
H
Global levels of Modifications on Histone H4
NTM_H
TM_H
Relative Abundance
Peptide
I
NTM_H
TM_H
Vinculin
H3
H3K4me3
H3K4me2
H3K4me1
H3K9me3
H3K9me2
H3K9me1
H3K27me3
H3K27me1
H3K36me3
H3K36me2
H3K36me1
H3K79me3
H3K79me2
H3K79me1
H4
H4K20me1
H4K20me2
H4K20me3
H4K12Ac
H3K36Ac
H4K5Ac
H3K23Ac
H3K14Ac
H3K27Ac
H2AK5Ac
H3K18Ac
H4K8Ac
H4K16Ac
H3K9Ac
H2A

Figure S2. Validation of chromatin changes in human tumors, Related to Figure 2.

(A-F) Correlation plots between ChIP-Seq and ChIP-String. Plots showing correlations of normalized mark intensity in ChIP-Seq experiment (Y-axis) and ChIP-String experiment (X-axis) in NTM_H and TM_H cells for H2BK5Ac (A), H4K5Ac (B), H3K27me3 (C), H3K27Ac (D), H3K4me1 (E) and H3K4me3 (F). **(G-J)** Boxplots showing average normalized intensity for ChIP-string probes across NTM_H, TM_H, nevi and tumors individually for H2BK5Ac probes high in NTM_H cells (G), H2BK5Ac probes high in TM_H cells (H), H4K5Ac probes high in NTM_H cells (I) and H4K5Ac probes high in T_H cells (J).

Figure S2

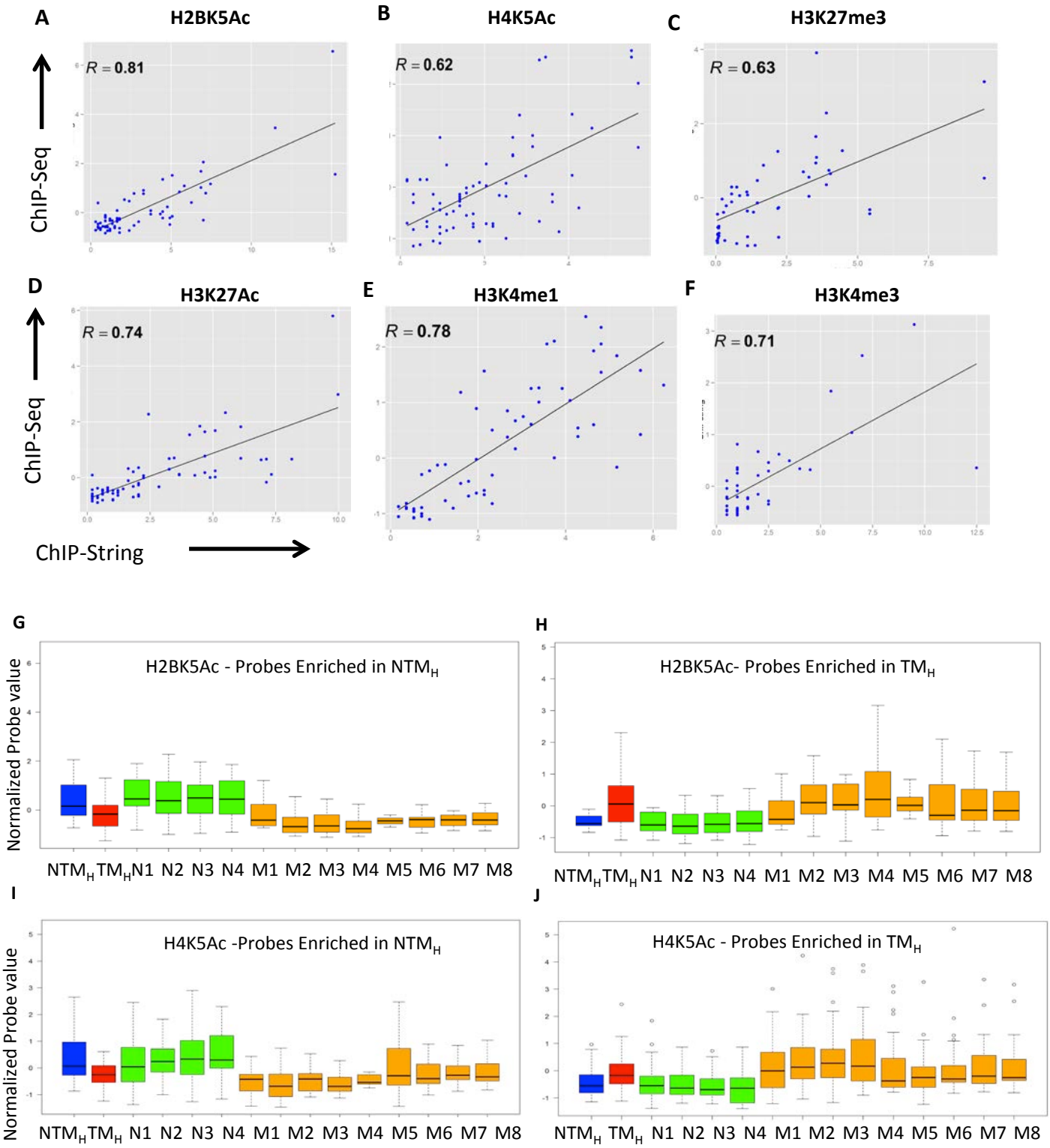

A
H2BK5Ac
R = 0.81
B
H4K5Ac
R = 0.62
C
H3K27me3
R = 0.63
D
H3K27Ac
R = 0.74
E
H3K4me1
R = 0.78
F
H3K4me3
R = 0.71
ChIP-Seq
ChIP-Seq
ChIP-String
G
H2BK5Ac - Probes Enriched in NTM_H
H
H2BK5Ac- Probes Enriched in TM_H
I
H4K5Ac -Probes Enriched in NTM_H
J
H4K5Ac - Probes Enriched in TM_H
Normalized Probe value
Normalized Probe value
NTM_H TM_H N1 N2 N3 N4 M1 M2 M3 M4 M5 M6 M7 M8

Figure S3: Chromatin state profiles, Related to Figure 3 (A) Correlation plot showing pearson correlations of histone modification peaks between the histone marks profiled in our study in NTM_H cells computed based on encoding the presence of a peak with a value of 1 and the absence of a peak with a value of 0. Peak calling was performed using MACS algorithm with default parameters except p-value < $1x10^{-8}$. 'DiffBind' bioconductor package was used to cluster correlation values. **(B)** Transition parameters for 18-state model derived by ChromHMM for NTM_H and TM_H cells. **(C-D)** Overlap of different genomic features (CpG island, RefSeq TSS, RefSeq TES, laminB lads (Guelen et al., 2008), 5-hMeC enriched and 5-MeC enriched regions) with chromatin state calls in NTM_H (C) and TM_H (D) cells. The fold enrichments are calculated as the ratio between observed and expected number of genomic bins for each overlap. The color intensities are normalized within each column between its minimum value (white) and its maximum value (blue). The last column shows the mean DNA methylation level for each chromatin state on the scale from completely unmethylated (white) to fully methylated (red). **(E)** Overlap enrichment of TSS coordinates and then gene body of highly expressed (FPKM >5) and low/not expressed (FPKM <5) genes in NTM_H cells with chromatin states. Next to them is a gene expression positional plot that shows the average gene expression per chromatin state and cell type at a given distance within 50kb of annotated transcription start sites.

Figure S3

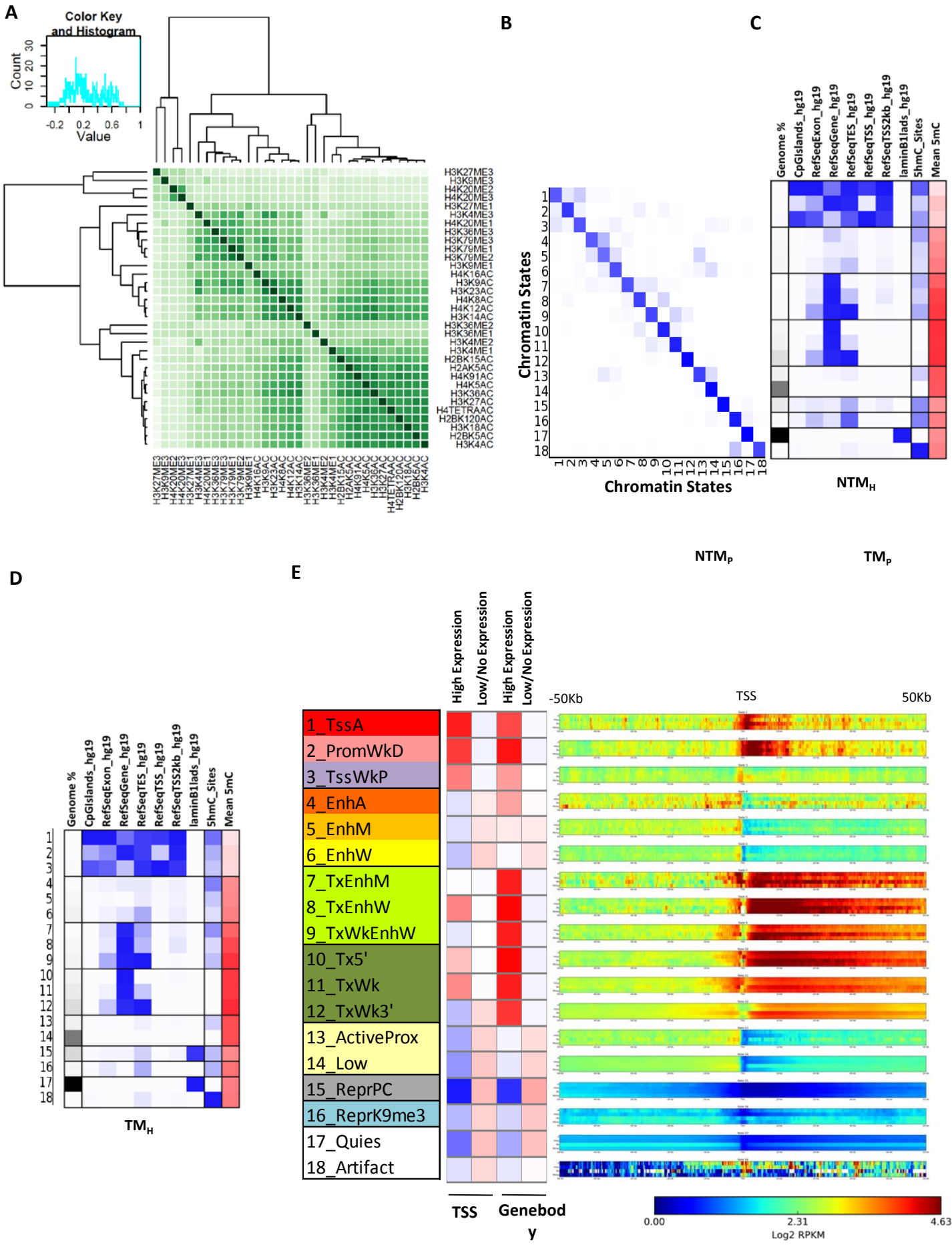
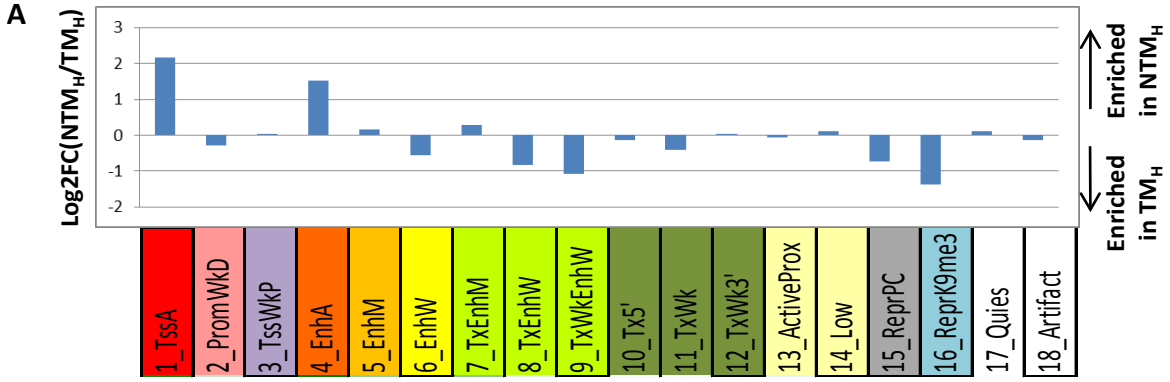
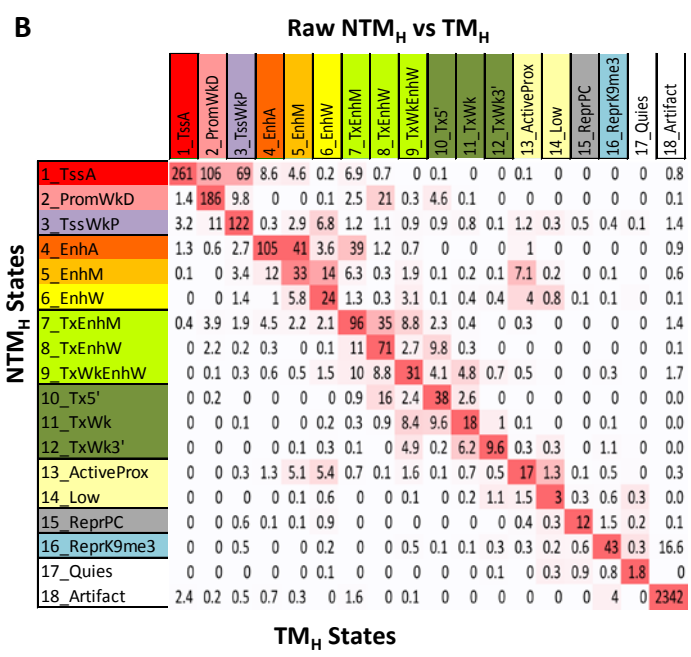

A
Color Key and Histogram
Count
Value
B
Chromatin States
C
Genome %
CpGIslands_hg19
RefSeqExon_hg19
RefSeqGene_hg19
RefSeqTES_hg19
RefSeqTSS_hg19
RefSeqTSS2kb_hg19
laminB1lads_hg19
5hmC_Sites
Mean 5mC
NTM_H
NTM_P
TM_P
D
TM_H
E
High Expression
Low/No Expression
1_TssA
2_PromWkD
3_TssWkP
4_EnhA
5_EnhM
6_EnhW
7_TxEnhM
8_TxEnhW
9_TxWkEnhW
10_Tx5'
11_TxWk
12_TxWk3'
13_ActiveProx
14_Low
15_ReprPC
16_ReprK9me3
17_Quies
18_Artifact
-50Kb
TSS
50Kb
TSS
Genebody
0.00
2.31
4.63
Log2 RPKM

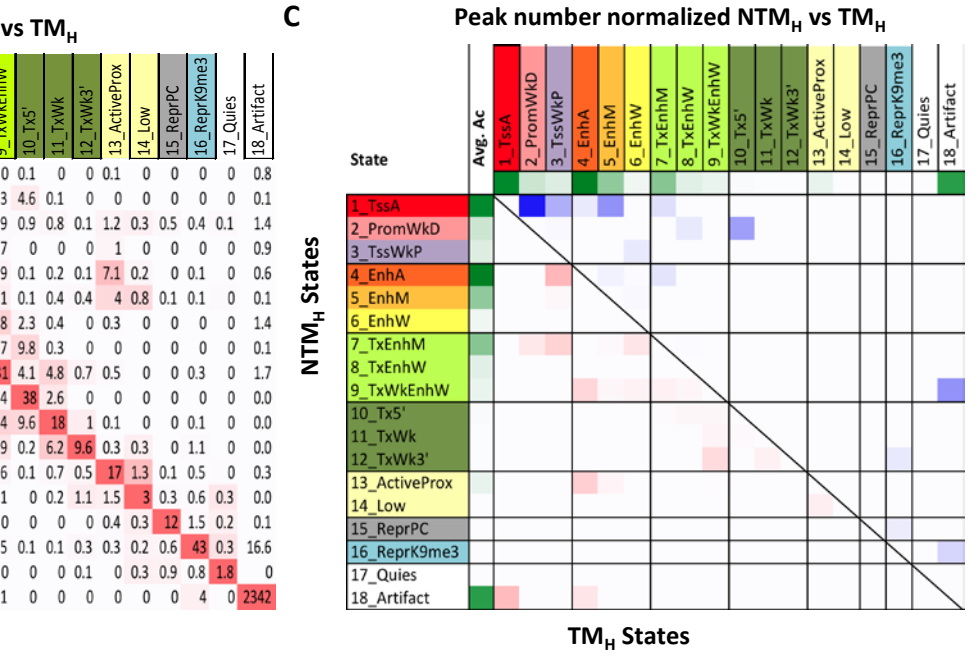
Figure S4. Chromatin state transitions between non-tumorigenic and tumorigenic cells, Related to Figure 3. (A) Log_2 ratios between the total number of genomic bins occupied by each chromatin state in non-tumorigenic (NTM_H) and tumorigenic (TM_H) cells. **(B)** Heat map showing state transitions in NTM_H and TM_H cells with raw enrichment scores. **(C-E)** Heat maps showing under different normalization schemes fold enrichment of transitions of chromatin states in non-tumorigenic to tumorigenic cells controlling for the overall state size and similarity (**see Supplementary Methods**). The color intensities above the main diagonal range from white (relative enrichment <1) to blue (relative enrichment > 20), thus indicating chromatin state transitions that lose acetylation marks from non-tumorigenic to tumorigenic cells within the same category are more enriched compared to the reverse chromatin state transition (i.e. from tumorigenic to non-tumorigenic). Similarly, the colors below the main diagonal range from white (relative enrichment < 1) to red (relative enrichment > 20), thus indicating the lack of chromatin state transitions that gain acetylation marks from non-tumorigenic to tumorigenic cells within each category that are more enriched compared to the reverse chromatin state transition (i.e. from tumorigenic to non-tumorigenic). **(C)** Relative enrichments for NTM_H vs. TM_H with binary peak calls normalized to the same number. **(D)** Relative enrichments for NTM_H vs. TM_H with sequencing reads downsampled to same number. **(E)** Relative enrichments for NTM_P vs. TM_P with binary peak calls normalized to the same number.

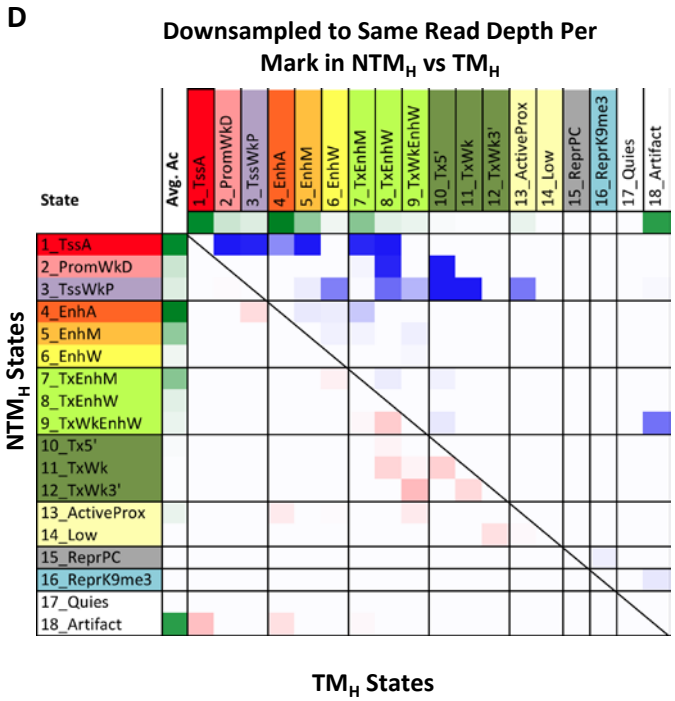
Figure S4





NTM_H States \ TM_H States	1_TssA	2_PromWkD	3_TssWkP	4_EnhA	5_EnhM	6_EnhW	7_TxEnhM	8_TxEnhW	9_TxWkEnhW	10_Tx5'	11_TxWk	12_TxWk3'	13_ActiveProx	14_Low	15_ReprPC	16_ReprK9me3	17_Quies	18_Artifact
1_TssA	261	106	69	8.6	4.6	0.2	6.9	0.7	0	0.1	0	0	0.1	0	0	0	0	0.8
2_PromWkD	1.4	186	9.8	0	0	0.1	2.5	21	0.3	4.6	0.1	0	0	0	0	0	0	0.1
3_TssWkP	3.2	11	122	0.3	2.9	6.8	1.2	1.1	0.9	0.9	0.8	0.1	1.2	0.3	0.5	0.4	0.1	1.4
4_EnhA	1.3	0.6	2.7	105	41	3.6	39	1.2	0.7	0	0	0	1	0	0	0	0	0.9
5_EnhM	0.1	0	3.4	12	33	14	6.3	0.3	1.9	0.1	0.2	0.1	7.1	0.2	0	0.1	0	0.6
6_EnhW	0	0	1.4	1	5.8	24	1.3	0.3	3.1	0.1	0.4	0.4	4	0.8	0.1	0.1	0	0.1
7_TxEnhM	0.4	3.9	1.9	4.5	2.2	2.1	96	35	8.8	2.3	0.4	0	0.3	0	0	0	0	1.4
8_TxEnhW	0	2.2	0.2	0.3	0	0.1	11	71	2.7	9.8	0.3	0	0	0	0	0	0	0.1
9_TxWkEnhW	0	0.1	0.3	0.6	0.5	1.5	10	8.8	31	4.1	4.8	0.7	0.5	0	0	0.3	0	1.7
10_Tx5'	0	0.2	0	0	0	0	0.9	16	2.4	38	2.6	0	0	0	0	0	0	0.0
11_TxWk	0	0	0.1	0	0	0.2	0.3	0.9	8.4	9.6	18	1	0.1	0	0	0.1	0	0.0
12_TxWk3'	0	0	0	0	0.1	0.3	0.1	0	4.9	0.2	6.2	9.6	0.3	0.3	0	1.1	0	0.0
13_ActiveProx	0	0	0.3	1.3	5.1	5.4	0.7	0.1	1.6	0.1	0.7	0.5	17	1.3	0.1	0.5	0	0.3
14_Low	0	0	0	0	0.1	0.6	0	0	0.1	0	0.2	1.1	1.5	3	0.3	0.6	0.3	0.0
15_ReprPC	0	0	0.6	0.1	0.1	0.9	0	0	0	0	0	0	0.4	0.3	12	1.5	0.2	0.1
16_ReprK9me3	0	0	0.5	0	0	0.2	0	0	0.5	0.1	0.1	0.3	0.3	0.2	0.6	43	0.3	16.6
17_Quies	0	0	0	0	0	0.1	0	0	0	0	0	0.1	0	0.3	0.9	0.8	1.8	0
18_Artifact	2.4	0.2	0.5	0.7	0.3	0	1.6	0	0.1	0	0	0	0	0	0	4	0	2342





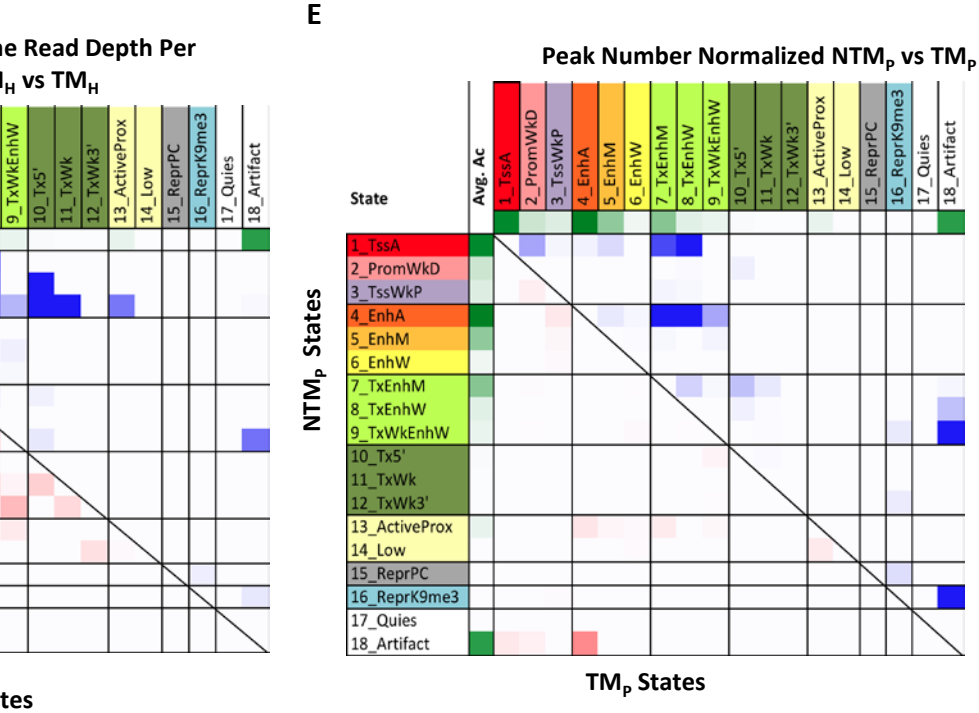
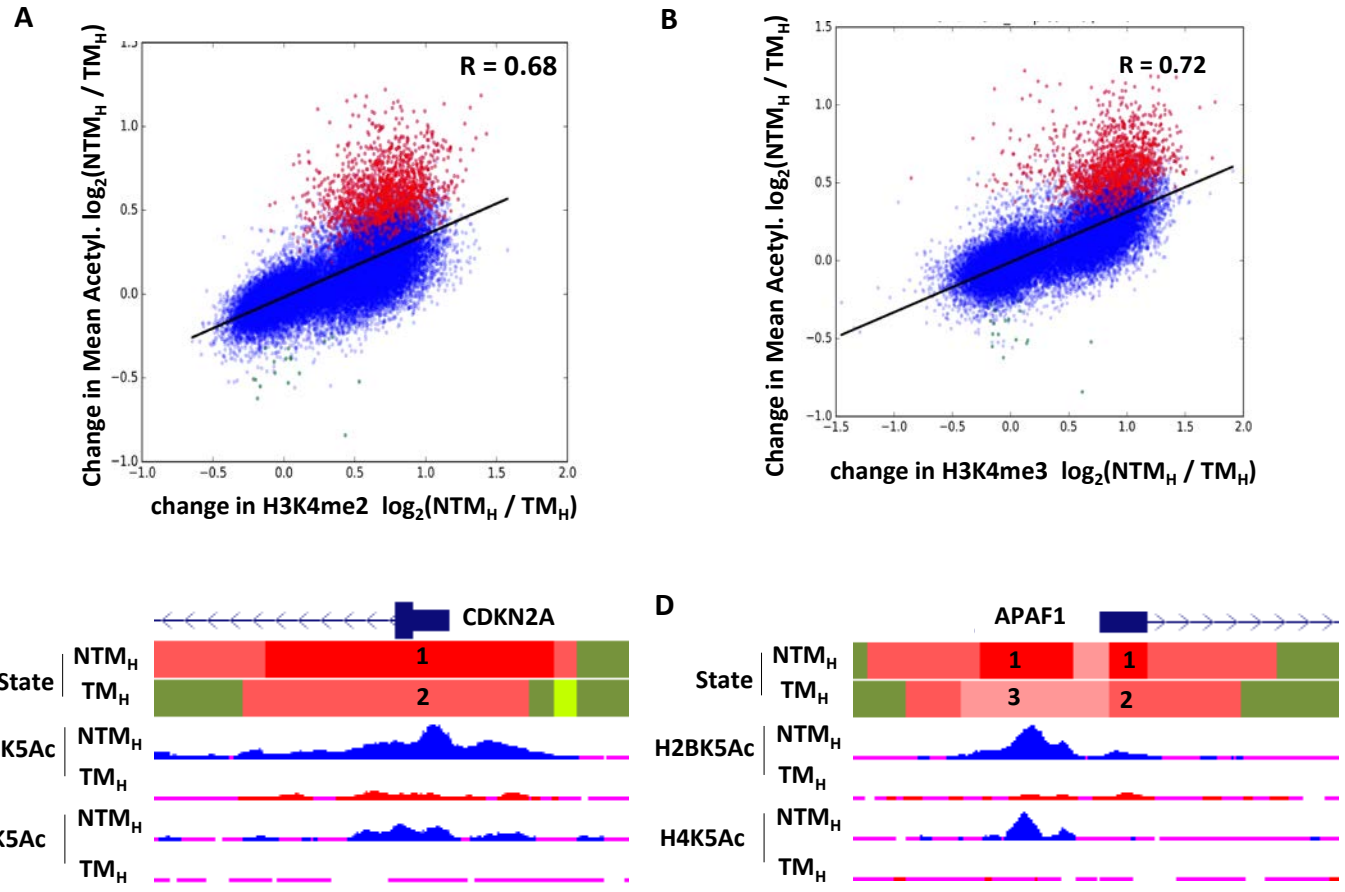


Figure S5. Chromatin state changes mark specific cancer pathways, Related to Figure 4. (A-B) The graphs show correlation between log_2 fold change in mean acetylation and (A) H3K4me2 or (B) H3K4me3. Log_2 fold change in mean acetylation (Y-axis) for a particular RefSeq promoter was plotted against log_2 fold change in H3K4me2 or H3K4me3 signal (X-axis) in that promoter for NTM_H vs. TM_H. Log_2 fold changes were calculated as $log_2((1 + \text{signal in } NTM_H) / (1 + \text{signal in } TM_H))$. Overall these changes in H3K4me2/3 correlate highly (R = 0.68 for H3K4me2 and 0.72 for H3K4me3) with alteration in mean acetylation suggesting that these marks function as coregulators. Points in red indicate promoters that were called as significantly deacetylated in TM_H at FDR of 1% by the permutation test in our analysis. **(C-D)** UCSC genome browser track for chromatin state and histone acetylations H2BK5Ac and H4K5Ac on genomic loci encompassing CDKN2A (A) and APAF1 (B) in NTM_H and TM_H cells. **(C-D)** Top enriched pathways (pathway commons) associated with genes closest to enhancers displaying state transitions from State 4_EnhA in non-tumorigenic cells (NTM_H) to States 7_TxEnhM and 5_EnhM in tumorigenic (TM_H) cells.

Figure S5



E

State 4_EnhA (NTM_H) to 7_TxEnhM (TM_H)

Pathway	Hyper FDR
Integrin family cell surface interactions	2.54E-29
VEGF and VEGFR signaling network	1.65E-29
Beta1 integrin cell surface interactions	1.73E-29
ErbB receptor signaling network	2.22E-29
LKB1 signaling events	2.86E-29
Signaling events mediated by VEGFR1 and VEGFR2	2.66E-29
Glypican 1 network	3.48E-29
Proteoglycan syndecan-mediated signaling events	3.55E-29
PDGF receptor signaling network	3.29E-29
IFN-gamma pathway	4.51E-29

F

State 4_EnhA (NTM_H) to 5_EnhM(TM_H)

Pathway	HyperFDR
Beta1 integrin cell surface interactions	8.48E-31
Integrin family cell surface interactions	4.98E-31
Proteoglycan syndecan-mediated signaling events	7.29E-31
Glypican pathway	2.83E-30
ErbB receptor signaling network	7.80E-30
VEGF and VEGFR signaling network	8.29E-30
TRAIL signaling pathway	1.16E-29
Glypican 1 network	1.34E-29
LKB1 signaling events	1.58E-29
PAR1-mediated thrombin signaling events	2.15E-29

Figure S6. Comparative analysis of chromatin changes with RNA expression changes, Related to Figure 5. (A) Relative enrichment (in log space) of number of steady genes that do not change expression between NTM_H and TM_H cells for all pairs of chromatin state transitions in their promoters. **(B)** Difference in enrichment of downregulated genes and upregulated genes on all pairs of chromatin state transitions between NTM_H and TM_H cells. **(C-E)** Relative enrichment of number of downregulated genes (C) or upregulated genes (D) or steady genes that do not change expression between NTM_P and TM_P cells (E) for all pairs of chromatin state transitions. **(F)** Difference in enrichment of downregulated genes and upregulated genes on all pairs of chromatin state transitions between NTM_P and TM_P cells. **(G)** Percent of genes showing down- (orange) or up- (blue) regulated gene expression change in top 100, 250, 500, 1000 and 2000 genes with differential levels in either direction between NTM_H and TM_H cells. Similarly, percent of the promoters showing gain (red) or loss (green) of acetylation in top 100, 250, 500, 1000 and 2000 promoter regions with differential levels between NTM_H and TM_H cells. **(H)** Scatter plot displays $\log_2$(fold change + 1) for acetylation and gene expression changes between NTM_H and TM_H. The color scheme is same as that in Figure 5D.

Figure S6

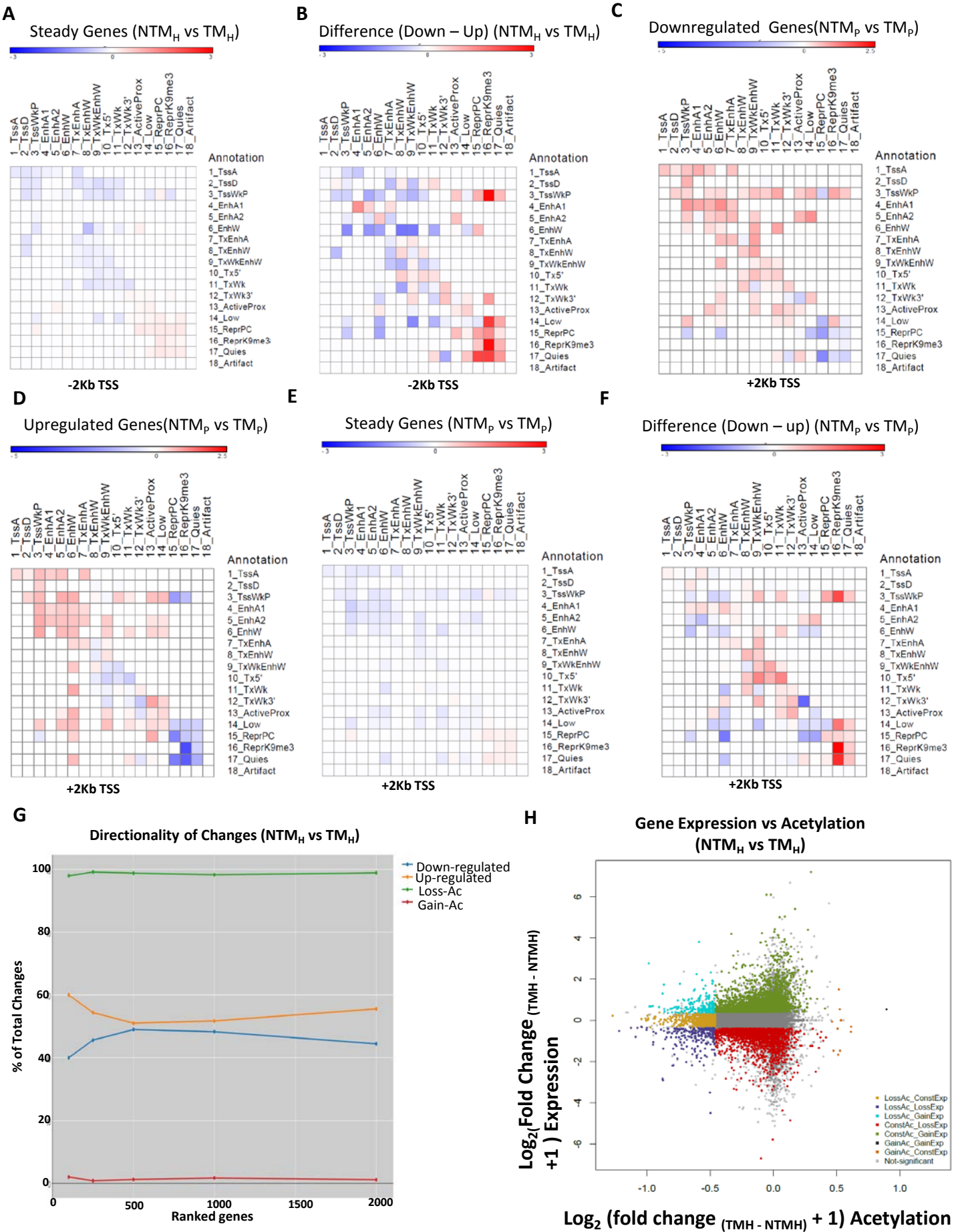
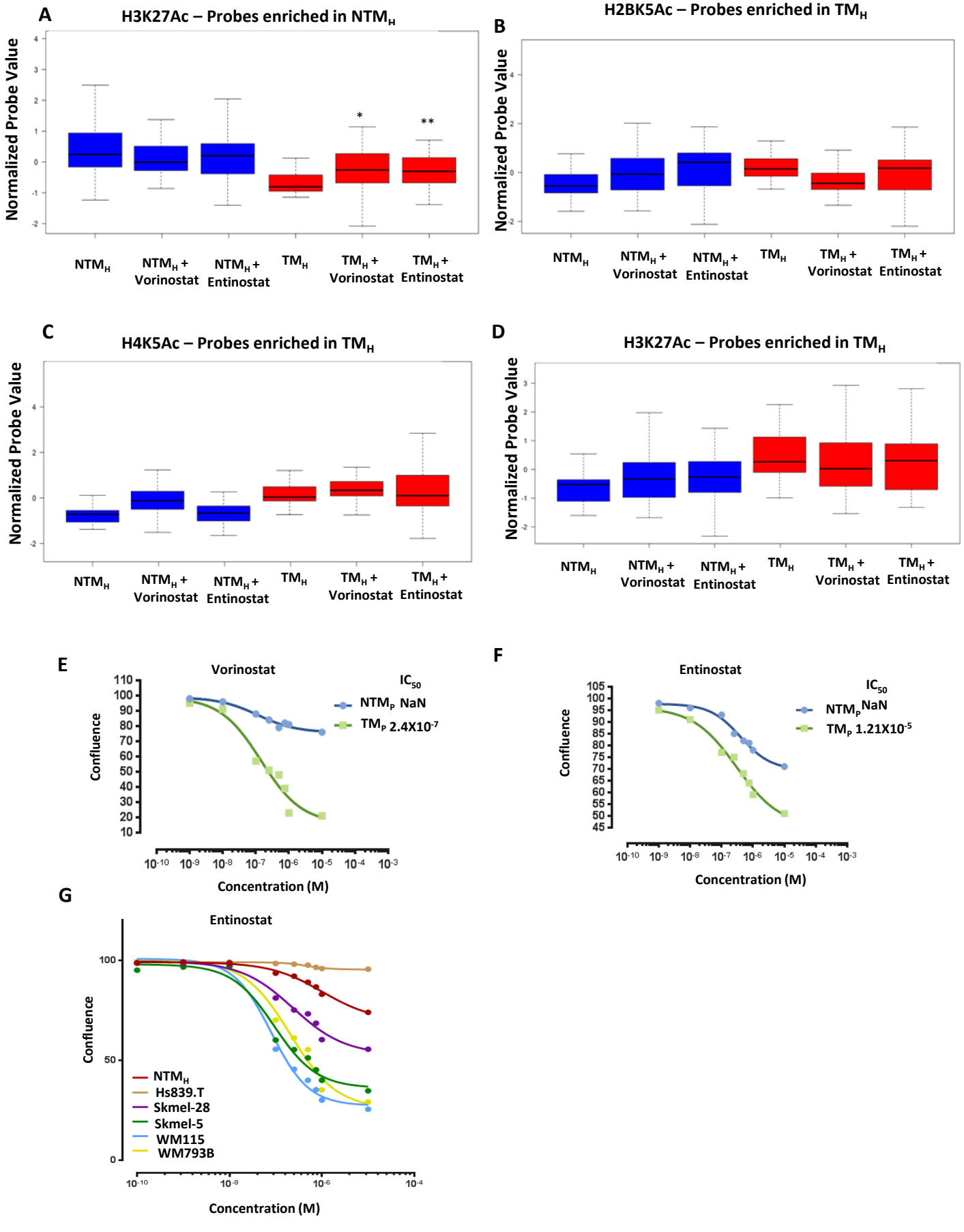

A
Steady Genes (NTM$_H$ vs TM$_H$)
B
Difference (Down – Up) (NTM$_H$ vs TM$_H$)
C
Downregulated Genes(NTM$_P$ vs TM$_P$)
D
Upregulated Genes(NTM$_P$ vs TM$_P$)
E
Steady Genes (NTM$_P$ vs TM$_P$)
F
Difference (Down – up) (NTM$_P$ vs TM$_P$)
Annotation
1_TssA
2_TssD
3_TssWkP
4_EnhA1
5_EnhA2
6_EnhW
7_TxEnhA
8_TxEnhW
9_TxWkEnhW
10_Tx5'
11_TxWk
12_TxWk3'
13_ActiveProx
14_Low
15_ReprPC
16_ReprK9me3
17_Quies
18_Artifact
-2Kb TSS
+2Kb TSS
G
Directionality of Changes (NTM$_H$ vs TM$_H$)
Down-regulated
Up-regulated
Loss-Ac
Gain-Ac
% of Total Changes
Ranked genes
H
Gene Expression vs Acetylation
(NTM$_H$ vs TM$_H$)
Log$_2$(Fold Change $_{(TMH - NTMH)}$ +1) Expression
Log$_2$ (fold change $_{(TMH - NTMH)}$ + 1) Acetylation
LossAc_ConstExp
LossAc_LossExp
LossAc_GainExp
ConstAc_LossExp
ConstAc_GainExp
GainAc_GainExp
GainAc_ConstExp
Not-significant

Figure S7. Acetylation status and proliferation changes in response to HDAC inhibitors, Related to Figure 7. (A) Boxplots showing average normalized intensity for H3K27Ac levels on ChIP-string probes (that are enriched in NTM_H cells in ChIP-seq studies) across NTM_H and TM_H cells that were either untreated or treated with vorinostat (200nM) or entinostat (300nM) for 72 hrs. **(B-D)** Boxplots showing average normalized intensity for (B), H2BK5Ac, (C), H4K5Ac, or (D) H3K27Ac levels on ChIP-string probes that are enriched in T_H cells across NTM_H and TM_H cells treated with vorinostat (200nM) or entinostat (300nM) for 72hrs. Asterisk (*) represents $p<0.05$ and double asterisk (**) represents $p<0.001$ (Wilcoxon Rank test) when comparisons are made to TM_H. **(E-F)** Growth curves for NTM_H and TM_H cells grown under various concentrations of (E) vorinostat or (F) entinostat. IC_{50} values are also shown. NaN refers to 'not a number'. **(G)** Growth curves for melanoma cell lines grown under various concentrations of entinostat. IC_{50} values are in Figure 7G.

Figure S7


A
H3K27Ac – Probes enriched in NTM$_H$
Normalized Probe Value
NTM$_H$
NTM$_H$ + Vorinostat
NTM$_H$ + Entinostat
TM$_H$
TM$_H$ + Vorinostat
TM$_H$ + Entinostat
B
H2BK5Ac – Probes enriched in TM$_H$
C
H4K5Ac – Probes enriched in TM$_H$
D
H3K27Ac – Probes enriched in TM$_H$
E
Vorinostat
Confluence
IC$_{50}$
NTM$_P$ NaN
TM$_P$ 2.4X10^{-7}
Concentration (M)
F
Entinostat
NTM$_P$ NaN
TM$_P$ 1.21X10^{-5}
G
Entinostat
NTM$_H$
Hs839.T
Skmel-28
Skmel-5
WM115
WM793B

SUPPLEMENTARY TABLES

Table S1. Details of sequencing data generated in this study, Related to Figure 1. Total read numbers for each mark in each of the 4 cell types used in this study, NTM_H, TM_H, NTM_P and TM_P.

Table S2. Details of probe locations used for ChIP-String, Related to Figure 2. Genomic location (hg19) of the 96-probes used in the nanostring codeset.

Table S3. Details of the nevi and tumor samples, Related to Figure 2. Clinical and genetic data for the nevi and tumor samples that were used for validation of histone modification levels used in Figure 2.

Table S4. Chromatin state recovery with subsets of marks. Related to Figure 3. Top panel shows fraction of state assignments recovered of the state of the row with only the mark of the column compared to using all the marks in NTM_H and TM_H cells. We observed cases of high recovery (>60%) of acetylated enhancer or promoter states with a single acetylation mark. Bottom panel shows fraction of state recovery with all marks except the mark of the column in NTM_H and TM_H cells compared to using all marks. We observed four cases of low recovery (<60%) of a chromatin states when all marks except one mark were included highlighting the existence of many locations uniquely marked by one mark. These four cases of low recovery were the chromatin states 6_EnhW, 10_Tx5', 15_ReprPC, and 16_ReprK9me3 when excluding the H3K4me1, H3K79me2, H3K27me3, and H3K9me3 marks respectively (Table S4).

Table S5. GO terms for all state changes from NTM_H to TM_H, Related to Figure 4.

Lists of GO-terms for the regions that belong to all significant chromatin state changes between NTM_H and TM_H cells.

Table S6: Pathways enriched for top 2 promoter and enhancer state transitions from NTM_H to TM_H, Related to Figure 4.

Table S7: Pathway analysis for different groups in overlap of gene expression and average acetylation, Related to Figure 5.

REFERENCES

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. Epigenomics *3*, 771-784.

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. BMC bioinformatics *7*, 191.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nature methods *9*, 215-216.

Ernst, J., and Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nature biotechnology *33*, 364-376.
Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z.*, et al.* (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Molecular cell *47*, 810-822.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W.*, et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.
Karch, K.R., Zee, B.M., and Garcia, B.A. (2014). High resolution is not a strict requirement for characterization and quantification of histone post-translational modifications. J Proteome Res *13*, 6152-6159.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome biology *13*, R107.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nature methods *7*, 709-715.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology *28*, 495-501.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M.*, et al.* (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell *147*, 1628-1639.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research *44*, W160-165.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature biotechnology *31*, 46-53.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols *7*, 562-578.

Wang, D., Yan, L., Hu, Q., Sucheston, L.E., Higgins, M.J., Ambrosone, C.B., Johnson, C.S., Smiraglia, D.J., and Liu, S. (2012). IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics *28*, 729-730.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M.*, et al.* (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic acids research *38*, e178.
Younesy, H., Nielsen, C.B., Lorincz, M.C., Jones, S.J., Karimi, M.M., and Moller, T. (2016). ChAsE: chromatin analysis and exploration tool. Bioinformatics.

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