Modeling Spatio-temporal Dynamics of Chromatin Marks

Petko Fiziev[1,2,3], Jason Ernst[1,2,3,4,5,6,*]

[1] Bioinformatics Interdepartmental Program, University of California, Los Angeles

[2] Departments of Biological Chemistry, University of California, Los Angeles

[3] Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA

[4] Computer Science Department, University of California, Los Angeles

[5] Jonsson Comprehensive Cancer Center, University of California, Los Angeles

[6] Molecular Biology Institute, University of California, Los Angeles, California, USA

* To whom correspondence should be addressed. Email: jason.ernst@ucla.edu

ABSTRACT

To model spatial changes of chromatin mark peaks over time we developed ChromTime, a computational method that identifies regions for which peaks either expand or contract significantly or hold steady between time points. We applied our method to chromatin marks in a range of biological systems. Predicted expanding and contracting peaks likely mark regulatory regions associated with transcription factor binding dynamics and gene expression changes. Spatial dynamics of peaks are informative about gene expression changes beyond localized signal changes. In proximity of gene starts, peaks preferentially expand in the same direction as transcription and contract in the opposite direction.

KEYWORDS: epigenomics, time course ChIP-seq, spatial dynamics, histone modifications

BACKGROUND

Genome-wide mapping of histone modifications (HMs) and related chromatin marks using chromatin

immunoprecipitation coupled with high throughput sequencing (ChIP-seq) has emerged as a powerful approach to annotate genomes and study cell states[1–3]. Through the efforts of large consortia projects such as the ENCODE[4], Roadmap Epigenomics[5] and BLUEPRINT[6] as well as individual labs[7–9] multiple different chromatin marks have been mapped across more than a hundred different cell and tissue types. These maps have yielded numerous insights into gene regulation and genetic and epigenetic association with disease[10–14].

While many mapping efforts have largely focused on single or unrelated cell and tissue types[1, 4], a growing number of biological processes have been studied with temporal epigenomic data using time course ChIP-seq assays that map chromatin marks at consecutive stages during particular biological processes. Such datasets have been generated for a wide range of biological settings including T cell development[15], adipogenesis[16], hematopoiesis[17], macrophage differentiation[18], neural differentiation[10], cardiac development[19, 20], somatic cell reprogramming[21–24], embryogenesis[25] and many others[26–34]. The output of these experiments presents a unique opportunity to study the spatio-temporal changes of epigenetic peaks and associated regulatory elements. However, almost all computational methods designed or applied to epigenomic data have been developed based on single or multiple unrelated samples. For example, continuous regions of enrichments of single marks are detected by peak or domain calling methods[35–39]. In cases when multiple HMs are mapped in the same cell type, methods such as ChromHMM[40] and Segway[41] can be used to produce genome-wide chromatin state annotations. In addition, algorithms have been developed for pairwise comparisons of ChIP-seq signal data by differential peak calling[42, 43].

In the context of time course ChIP-seq data, only a few methods have been proposed that consider temporal dependencies between samples. One such method, TreeHMM[44], produces a chromatin state genome annotation similar to ChromHMM and Segway, while taking into account a tree-like structure that captures lineage relationships between the input cell types in order to potentially derive a more consistent annotation across samples. Another method, GATE[27], produces a genome annotation based on

clustering fixed length genomic loci that can be modeled with the same switch from one chromatin state to another over time.

One important limitation of methods for pairwise comparison of ChIP-seq data and time course modeling of it is that they do not directly consider or model spatial changes in the genomic territory occupied by chromatin marks over time. Spatial properties of genomic peaks continuously marked by HMs have gained increasing attention as a potentially important characteristic of chromatin marks. For example, long peaks of H3K27ac have been associated with active cell type specific locus control regions termed super-enhancers or stretch enhancers in a number of cell types[45, 46]. Also, the length of H3K4me3 peaks has been associated with transcriptional elongation and consistency of cell identity genes[47]. In the context of cancer, long H3K4me3 peaks have been linked to transcriptional elongation and enhancer activity at tumor suppressor genes and have been observed to be significantly shortened in tumor cells[48]. Long H3K4me3 domains have been implicated to mark loci involved in psychiatric disorders[49]. Expanded domains of H3K27me3 and H3K9me3 marks have been shown to be characteristic of terminally differentiated cells compared to stem cells[50]. These studies suggest that length of epigenetic peaks is a dynamic feature that can correlate with activity of putative functional elements regulating specific genes. Computational methods that do not explicitly reason about the spatial changes of HMs have significant limitations for studying the dynamics of these properties because they are unable to detect territorial changes that might be associated with redistribution of signal or identify asymmetric directional peak boundary movements.

In this work, we present ChromTime, a novel computational method for detection of expanding, contracting and steady peaks, which can detect patterns of changes in the genomic territory occupied by chromatin mark peaks from time course ChIP-seq data (**Fig 1A**). We applied ChromTime to a diverse set of data from different developmental, differentiation and reprogramming time courses. Expansions and contractions in general mark regulatory regions associated with changes in transcription factor binding and gene expression. ChromTime enables studying the directionality of spatial dynamics of chromatin mark peaks relative to other genomic features, which existing computational approaches do not directly

address. Our results show that the direction of expansions and contractions correlates with direction of transcription near transcription start sites. ChromTime is a general method that can be used to analyze time course ChIP-seq data from a wide range of biological systems to gain insights into the dynamics of gene regulation.

RESULTS

Model for detecting expanding, contracting and steady peaks from temporal ChIP-seq data

We developed a computational method, ChromTime (https://github.com/ernstlab/ChromTime), designed for systematic detection of expansions, contractions and steady peaks from time course ChIP-seq of single chromatin marks (**Fig 1B**). ChromTime takes as input a set of genomic coordinates of aligned sequencing reads from ChIP-seq and, optionally, control experiments over the time course. The method consists of two stages – block finding and dynamics prediction. During the block finding stage, ChromTime determines continuous genomic regions (blocks) that may contain ChIP-seq peaks throughout the time course. To achieve this, ChromTime partitions the genome into fixed length bins and counts the number of ChIP-seq and control reads that map to each bin at each time point. Nearby bins that show significant enrichment are joined into continuous intervals, which subsequently are grouped into blocks if they overlap across time points. As a result, large portions of the genome that are likely to contain background noise at all time points are filtered out, so that peak boundary dynamics are determined within a subset of the genome potentially enriched for the chromatin mark.

During the dynamics prediction stage, for each block ChromTime determines the most likely positions of the peak boundaries at each time point and whether the peak expands, contracts or holds steady at each boundary between consecutive time points. The method uses a probabilistic mixture model to partition the signal within each block at each time point into background and peak components (**Fig 1C**) by reasoning jointly about the data from all time points in the time course. The model assumes that central positions in blocks have higher prior probability to be enriched for ChIP-seq reads on average and thus more likely to

associate with the peak component (see **Methods, Fig S1D**). The number of sequencing reads in bins from each component at each time point is modelled with different negative binomial distributions that can account for the local abundance of control reads. Furthermore, between any two consecutive time points the boundaries of the peaks are assumed to follow one of three possible dynamics: steady, expand or contract. For steady dynamics, the peak boundaries are enforced to have the same genomic position. For expanding and contracting dynamics the number of genomic bins that the peak boundaries move between the two time points is modelled with different negative binomial distributions which depend on the pair of time points and the corresponding dynamic. Each dynamic is also assumed to have a prior probability which captures its genome-wide frequency at each time point.

All model parameters are learned jointly from the whole time course. As a result, ChromTime can adapt to different boundary movements, dynamics frequencies and noise levels across experiments and biological systems. The estimated parameters are used to make a prediction for each block for the most likely positions of the peak boundaries and the corresponding boundary dynamics that had generated the signal within the block. The final output contains predicted peak boundaries annotated and colored by their assigned dynamics, which can be used for downstream analysis with existing tools and visualized in genome browsers (**Fig 2, S2,** https://github.com/ernstlab/ChromTime).

Reproducibility of ChromTime predictions and enrichment for TF binding and DNaseI hypersensitivity sites

To investigate the reproducibility of ChromTime predictions, we applied ChromTime separately to two biological replicate datasets for the H3K4me2 and the H3K(9/14)ac marks in T cell development in mouse[15] and confirmed on average strong enrichment for the same ChromTime annotations co-localizing across replicates (**Fig S3**). We then applied the method to data from pooled replicates for the H3K4me2 and H3K(9,14)ac marks from the mouse T cell development study[15] and to data for the H3K4me3 and H3K27ac marks from a study on stem cell reprogramming in human[21]. To investigate the biological relevance of ChromTime predictions, we intersected their genomic positions with the dynamics

of orthogonal genomic annotations for transcription factor binding[4, 15] and DNaseI hypersensitivity sites[5]. Predicted expanding and contracting peaks enriched for sites bound by important regulators in each biological system as well as sites bound by generic TFs in a cell type specific manner (**Fig 3**). Furthermore, predicted steady peaks associated with TF binding sites and DNaseI hypersensitive sites that are shared between the first and the last time point in each time course, which mark potentially stable regulatory elements.

Predicted spatial dynamics by ChromTime associate better with gene expression changes compared to boundary position changes of peaks called from individual time points in isolation

To investigate whether ChromTime's approach for reasoning jointly about the whole time course increases power to detect associations with gene expression compared to considering boundary differences of peaks at consecutive time points called in isolation, we analyzed gene expression changes of genes with TSSs overlapping ChromTime predictions in relation to posterior probabilities for expansions and contractions compared to boundary differences of peaks called with ChromTime from data from individual time points in isolation. We specifically investigated this in the context of H3K4me2 peaks in mouse T cell development[15] and for H3K4me3 peaks in stem cell reprogramming in human[21] (**Fig 4**). In most cases, ranking block boundaries by their predicted ChromTime posterior probabilities for expansions and contractions associated on average with larger gene expression changes compared to ranking block boundaries directly based on the change in the genomic positions of the boundaries of ChromTime peaks called at individual time points in isolation. These results also held when using peaks from two different peak callers, MACS2[35] and SICER[37] applied on data from individual time points (**Fig S4**).

Spatial dynamics contain information about gene expression changes in consecutive time points not captured by corresponding pairwise HM signal changes

We next investigated whether there is additional information in ChromTime predictions beyond what can be captured by pairwise ChIP-seq signal density changes or by differential peak calls. For this analysis, we focused on H3K4me2 in mouse T cell development[15] and H3K4me3 in human stem cell reprogramming[21]. At each pair of consecutive time points from the first to the last time points with predicted peaks in a block, we computed the change in ChIP-seq signal density within the left most and right most predicted peak boundaries in the block. We associated the signal density changes with gene expression changes at the nearest transcription start site within 50kb of each block and computed the average gene expression change as a function of the ChIP-seq signal density change within blocks (**Fig 5**). We found that locations with the same ChIP-seq signal density change can associate with significantly different average gene expression changes of proximal genes depending on the predicted ChromTime dynamics. Notably, bidirectional expansions, expansions occurring on both sides of a peak, associated with greater average increase in gene expression than unidirectional expansions, those expansions occurring on one side with steady on the other. These unidirectional expansions in turn associated with greater expression change than steady regions, those regions with a steady call on both sides of a peak. We observed a similar relationship for contractions and decrease of gene expression. These results were replicated also after substituting ChIP-seq signal density changes with differential peak scores as outputted by two different differential peak calling methods (SICER[37] and MACS2[35], **Fig S5**). Therefore, ChromTime predictions can provide additional information about gene expression changes beyond what is contained in the corresponding ChIP-seq signal density changes as measured directly or by utilizing differential peak calling procedures.

Spatial dynamics are correlated between multiple histone marks

Previous studies have shown that the locations of different HMs can be correlated[1, 51]. In this context, we tested whether multiple HMs can also exhibit jointly the same type of spatio-temporal dynamics. For this purpose, we compared the genomic locations of predicted expansions, contractions and steady peaks for different HMs within the same time course. We focused on two previously published time courses – stem cell reprogramming in human[21] and adipogenesis in mouse[16], where multiple HMs

were mapped (**Fig 6**). In both datasets, we observed that expansions for H3K4me1, H3K4me2, H3K4me3 and H3K27ac co-localized preferentially and similarly for contractions and steady. In contrast, different spatial dynamics for H3K36me3 and H3K27me3 tended to occupy distinct locations. These results suggest that spatial dynamics of HMs are coordinated at least at a subset of genomic locations.

Direction of expansions and contractions is correlated with direction of transcription

ChromTime can predict unidirectional expansions and contractions, which enables analysis of directionality of spatial dynamics of peaks, an aspect of chromatin regulation that has not been previously systematically explored. To investigate this, we applied ChromTime on data from 11 previously published studies from a variety of developmental, differentiation and reprogramming processes (**Table 1**) for nine different HMs including narrow and broad marks and for Pol2. We observed that unidirectional expansions and contractions are predicted in most cases on average to be the majority of all expansions and contractions, respectively, at a given pair of consecutive time points (**Fig S6**). One hypothesis for the prevalence of asymmetric boundary movements for the promoter associated chromatin marks is that the direction of boundary movements is associated with the asymmetry of transcription initiation in promoter regions. To test this hypothesis, for each dataset we compared the prevalence of each class of unidirectional dynamics as a function of its distance to the nearest annotated transcription start site (TSS) and the orientation of the corresponding gene (**Fig 7**). Consistent with our hypothesis, for H3K4me3, H3K4me2, H3K(9/14)ac, H3K79me2, and for Pol2, we found that unidirectional expansions that expand into the gene body (i.e. in the same direction as transcription) were substantially more frequently found in proximity of TSSs compared to unidirectional expansions in the opposite direction. Moreover, this difference was not observed for expansions distal from TSSs. Similarly, in most cases for these data unidirectional contractions that contract towards the TSS of the nearest gene (i.e. in the opposite direction of transcription) were substantially more frequently found compared to unidirectional contraction in the opposite direction in proximity of TSSs, whereas their frequencies at distal sites showed much smaller differences. HMs H3K27ac, H3K4me1 and H3K27me3 exhibited the same trend, but to a smaller degree.

DISCUSSION

In this work, we presented ChromTime, a novel computational method for systematic detection of expanding, contracting and steady peaks of chromatin marks from time course ChIP-seq data. ChromTime employs a probabilistic graphical model that directly models changes in the genomic territory occupied by single chromatin marks over time. This approach allowed us to directly encode our modeling assumptions about dependencies between variables in an interpretable and extendable framework.

We applied ChromTime on datasets for broad and narrow HMs and for Pol2 from a variety of developmental, differentiation and reprogramming courses. Our results show that the method can identify sets of expanding and contracting peaks that are biologically relevant to the corresponding systems. In particular, expansions and contractions associate with up- and down-regulation of gene expression and differential transcription factor binding, supporting the biological relevance of ChromTime predictions.

ChromTime gains power by both reasoning jointly about all time points in a time course and by explicitly modeling the peak boundary movements. Supporting this we demonstrated that territorial changes identified by ChromTime had better agreement with gene expression changes compared to considering directly the boundary change of peaks called on data from individual time points in isolation. Additionally, we found that expanding and contracting peaks associated on average with greater change in gene expression compared to peaks with steady boundaries even after controlling for ChIP-seq signal density changes. Some of the power that ChromTime gains from considering spatial information might be explained by its ability to differentiate territorial expansions or contractions which can reflect changes in the number sites of TF binding in close vicinity from changes in ChIP-seq signal within steady peak boundaries. Changes in ChIP-seq signal without territorial expansions or contractions might reflect a change in proportion of cells with the chromatin mark without large changes in activity in any one cell. Additional power can come from the temporal and spatial information that allows the model to effectively smooth over noise in the data enabling more biologically relevant inferences.

ChromTime enables novel analysis of directionality of spatial epigenetic dynamics. In this context, we found that asymmetric unidirectional expansions and contractions for several marks correlate strongly with direction of transcription in promoter proximal regions, which suggests that spatial dynamics at such locations may be related to actions of the transcriptional machinery. One possible explanation for the observed correlation between the direction of spatial dynamics of at least some HMs and transcription can be provided in part by previous studies that have shown that the Pol2 elongation machinery can recruit H3K4-methyltransferases such as members of the SET[52] and MLL[53] families at the promoters of genes. Our findings are consistent with such models where the Pol2 complex itself may be facilitating the attachment and removal of these marks[54].

One current limitation of the ChromTime method is that while the runtime of ChromTime still scales linearly with the number of time points, *T*, the number of observed combinations of dynamics can scale exponentially with *T*. This exponential growth can complicate downstream analyses that directly consider each combination of dynamics, as there will be 3^{T-1} possible sequences of dynamics at each side of a peak. Extensions of the ChromTime model could model the large number of combinations as being instances of a smaller number of more distinct dynamic patterns.

CONCLUSIONS

The availability of time-series HM data provides an opportunity to understand chromatin dynamics in many biological systems. To better leverage the information in these experiments we presented ChromTime, a method to detect expansions, contractions and steady peaks of single chromatin marks between time points from time course ChIP-seq data. Our method gains power by both reasoning about data from all time points in the time course and by explicitly modeling movements of peak boundaries. We showed that ChromTime predictions are reproducible across biological replicates and associate with relevant genomic features such as changes in gene expression and transcription factor binding. We demonstrated that territorial changes of peaks can contain additional information beyond ChIP-seq signal changes with respect to gene expression of proximal genes. ChromTime allows for novel analysis of

directionality of spatial dynamics of chromatin marks. In this context, we showed for multiple chromatin marks that the direction of asymmetric expansions and contractions of peaks correlates strongly with direction of transcription in proximity of transcription start sites. ChromTime is generally applicable to modeling time courses of chromatin marks, and thus should be a useful tool to gaining insights into dynamics of epigenetic gene regulation in a range of biological systems.

METHODS

Overview of the ChromTime method

ChromTime takes as input a set of files in BED format with genomic coordinates of aligned sequencing reads from ChIP-seq experiments for a single mark over the time course and, optionally, from a set of control experiments. ChromTime consists of two stages (**Fig 1B-C**):

1) Detecting genomic intervals (blocks) potentially containing regions of ChIP-seq signal enrichment (peaks)

2) Learning a probabilistic mixture model for boundary dynamics of peaks within blocks throughout the time course and computing the most likely spatial dynamic and peak boundaries for each block throughout the whole time course

Detecting genomic blocks containing regions of ChIP-seq signal enrichment

The aim of this stage is to determine approximately the genomic coordinates of regions with potential ChIP-seq peaks at any time point in the time course. The signal within these blocks will be used as input to build the mixture model in the next stage of ChromTime. ChromTime supports analysis of both narrow marks and broad marks in two different modes. The method partitions the genome into non-overlapping bins of predefined length, BIN_SIZE (200 bp in narrow mode, 500 bp in broad mode by default) and

counts for each bin and time point the number of sequencing reads whose alignment starting positions after shifting by a predefined number of bases (100 bp in the direction of alignment by default) are within its boundaries. Next, each bin at each time point is tested for enrichment based on a Poisson background distribution at a predefined false discovery rate (5% by default). The expected number of reads for a bin at position *p* and time point *t*, $\lambda_{t,p}$, in the Poisson test is computed conservatively as the maximum of:

1) If control reads are provided: for each of windows of size *w*=1,000bp, 5,000bp and 20,000bp the average number of control reads in the window centered at the current bin, normalized by the ratio of total reads in ChIP-seq and control experiments, that is:

$$\lambda_{t,p,w} = \frac{\#\,[\text{Total Foreground Reads}]}{\#\,[\text{Total Control Reads}]}\frac{\text{BIN_SIZE}}{w}\text{Ctrl}_{t,p,w}$$

   where $\text{Ctrl}_{t,p,w}$ is the total number of control reads in each of window of size *w* around the bin at position *p* at time point *t*.

2) The average number of foreground ChIP-seq reads per genomic bin;

3) 1 read

Testing multiple different windows sizes for the background is a strategy we adopted from the MACS2 peak caller[35].

Within each time point, consecutive bins that are significantly enriched are merged into continuous intervals. The intervals are further extended in both directions to include continuous stretches of bins where each bin is significant based on a Poisson background distribution at a weaker P-value threshold (0.15 by default). Extended intervals within a predefined number of non-significant bins, MAX_GAP (3 bins by default), are further joined together. This joining strategy has been previously implemented by other peak callers for single datasets such as SICER[37]. Next, overlapping intervals across time points are grouped into blocks. To capture more of the potential background signal and to increase the likelihood

that central bins within blocks contain higher ChIP-seq signal, the start and end positions of each block are extended additionally by a predefined number of bins, BLOCK_EXTEND, (5 by default) upstream from the left-most coordinate and downstream from the right-most coordinate of the intervals in the block, respectively, or up to the middle point between the current block and its adjacent blocks if they are within BLOCK_EXTEND bins apart. Restricting BLOCK_EXTEND to a relatively limited number of bins helps in keeping the running time of the method within reasonable bounds.

In narrow mode, blocks that contain multiple intervals at the same time point separated by gaps of non-significant bins longer than MAX_GAP are split into sub-blocks at each gap between those intervals. In particular, all gaps within a block are intersected across the time points that have gaps. For each gap intersection, the block is split at the position with the lowest average ChIP-seq signal across all time points. In broad mode, no such splitting is performed in order to avoid excessive peak fragmentation.

Probabilistic mixture model for boundary dynamics of peaks within blocks across the time course

The ChIP-seq signal within the blocks is used as input to build a probabilistic mixture model for the boundary dynamics of the peaks within blocks. One core assumption of the model is that each block contains zero or one peak at each time point. This implies that at each time point, the bins within a block can be partitioned into three continuous intervals: left-flanking background, foreground peak, and right-flanking background. Let $O_{t,p}$ denote the number of observed ChIP-seq reads in a bin at position p at time point t, and $V_{t,p}$ denote the label of that bin (one of Peak or Background). The conditional distribution of $O_{t,p}$ conditioned on $V_{t,p}$ is modeled with different negative binomial distributions depending on the value of $V_{t,p}$:

$$P\left(O_{t,p} = k \middle| V_{t,p} = \mathrm{Peak}\right) = \mathrm{NB}\left(k; \mu_{\mathrm{Peak},t,p}, \delta_t\right) = \frac{\Gamma(k+\delta_t)}{k!\,\Gamma(\delta_t)} \left(\frac{\delta_t}{\mu_{\mathrm{Peak},t,p}+\delta_t}\right)^{\delta_t} \left(\frac{\mu_{\mathrm{Peak},t,p}}{\mu_{\mathrm{Peak},t,p}+\delta_t}\right)^{k}$$

and

$$P\left(O_{t,p} = k \middle| V_{t,p} = \text{Background}\right) = \text{NB}\left(k; \mu_{\text{Background},t,p}, \delta_t\right)$$

$$= \frac{\Gamma(k+\delta_t)}{k!\,\Gamma(\delta_t)}\left(\frac{\delta_t}{\mu_{\text{Background},t,p}+\delta_t}\right)^{\delta_t}\left(\frac{\mu_{\text{Background},t,p}}{\mu_{\text{Background},t,p}+\delta_t}\right)^{k}$$

Similarly to negative binomial regression models[55], ChromTime models the mean of each component through the log link as a linear combination of a two-dimensional vector of covariates, $X_{t,p}$,= (1, $\log\lambda_{t,p}$), which includes a constant term and the logarithm of the expected number of reads as computed in the previous section, $\lambda_{t,p}$:

$$\mu_{\text{Peak},t,p} = \exp\left[\alpha_t + \gamma_t \log\lambda_{t,p}\right]$$

$$\mu_{\text{Background},t,p} = \exp\left[\beta_t + \gamma_t \log\lambda_{t,p}\right]$$

where α_t, β_t and γ_t are time point specific scalar parameters. Negative binomial distributions have been successfully employed in a similar manner to capture the over-dispersion of sequencing reads in ChIP-seq experiments in peak callers for single samples such as ZINBA[56]. Of note however, ChromTime requires that the dispersion parameter δ_t and the coefficient γ_t are shared between the two components at each time point. The first requirement ensures that the distribution with the smaller mean value has higher probabilities compared to the distribution with the larger mean value for the lowest values of the support domain of the negative binomial distribution, and that the opposite holds for the largest values of the support domain (See **Supplementary Methods**). Sharing the dispersion parameter here is analogous to sharing the variance parameter in Gaussian mixture models. The second requirement to share the γ_t parameter ensures that the input signal has equal importance in each component.

Formally, let $B_{L,t}$ and $B_{R,t}$ denote the bin indices relative to the beginning of the block of the first and the last bin in the peak partition at time *t*, respectively, and *N* be the length of the block (i.e. $1 \leq B_{L,t} \leq N + 1$ and $0 \leq B_{R,t} \leq N$, with values of $B_{L,t} = N + 1$ and $B_{R,t}$ =0 corresponding to the special cases of starting a peak after all positions and ending a peak before all positions in a block, respectively). For $B_{L,t}$ and $B_{R,t}$

to denote valid interval boundaries, ChromTime requires that $B_{L,t} \leq B_{R,t} + 1$ at each time point. This can be formally encoded by introducing one auxiliary binary variable for each time point in the model, Z_t, such that:

$$P(Z_t = 1 | B_{L,t} = l, B_{R,t} = r) = \begin{cases} 1 & \text{if } l \leq r + 1 \\ 0 & \text{otherwise} \end{cases}$$

and thus also

$$P(Z_t = 0 | B_{L,t} = l, B_{R,t} = r) = \begin{cases} 0 & \text{if } l \leq r + 1 \\ 1 & \text{otherwise} \end{cases}$$

ChromTime treats all Z_t variables as observed with values equal to 1 for all blocks and time points. Then, the probability of a given partitioning of the ChIP-seq signal at time *t* under the model is:

$$\begin{aligned} P\left(B_{L,t} = l, B_{R,t} = r, Z_t = 1, O_t \middle| X_t\right) \\ &= P\left(Z_t = 1 | B_{L,t} = l, B_{R,t} = r\right) \\ &\times \prod_{p=1}^{l-1} \mathrm{NB}(O_{t,p}; \mu_{\mathrm{Background},t} = \exp[\beta_t + \gamma_t \log \lambda_{t,p}], \delta_t) \\ &\times \prod_{p=l}^{r} \mathrm{NB}\left(O_{t,p}; \mu_{\mathrm{Peak},t} = \exp\left[\alpha_t + \gamma_t \log \lambda_{t,p}\right], \delta_t\right) \\ &\times \prod_{p=r+1}^{N} \mathrm{NB}(O_{t,p}; \mu_{\mathrm{Background},t} = \exp[\beta_t + \gamma_t \log \lambda_{t,p}], \delta_t) \end{aligned}$$

An important special case of the above formulation when $B_{L,t} = B_{R,t} + 1$ corresponds to modelling the whole signal at time point *t* as background, which enables ChromTime to accommodate time points with no peaks.

ChromTime assumes uniform prior probabilities for the left and the right end boundaries at the first time point:

$$P(B_{L,1} = l) = Unif(1, N + 1)$$

and

$$P(B_{R,1} = r) = Unif(0, N)$$

where $Unif(a, b)$ denotes the uniform distribution of integer numbers in the closed interval $[a, b]$.

Between any two time points the ChromTime model allows for one of three possible dynamics at both the left and the right end boundaries of a peak: Steady, Expand or Contract. To capture the change of boundary positions between consecutive time points *t* and *t+1* we define the quantities $J_{L,t} = B_{L,t} - B_{L,t+1}$ and $J_{R,t} = B_{R,t+1} - B_{R,t}$ corresponding to the left and right boundaries respectively. Positive values of $J_{L,t}$ and $J_{R,t}$ indicate the number of bases a peak expanded, whereas negative values indicate the number of bases a peak contracted, and a value of 0 indicates the peak held steady on the left and the right side respectively. ChromTime models $J_{L,t}$ *and* $J_{R,t}$ with a different probability distribution for each of the three dynamics. Let $D_{S,t}$ denote the dynamic between time points *t* and *t+1* on boundary side *S*, where *S* is one of *L* (left side) or *R* (right side). For Steady dynamic, ChromTime uses the Kronecker delta function:

$$P\left(J_{S,t} \middle| D_{S,t} = \text{Steady}\right) = \begin{cases} 1, & \text{if } J_{S,t} = 0 \\ 0, & \text{otherwise} \end{cases}$$

For Expand and Contract, ChromTime employs negative binomial distributions to model the number of genomic bins a peak boundary moves relative to the minimal movement of one bin required for peak expansions and contractions:

$$P\left(J_{S,t} = j \middle| D_{S,t} = \text{Expand}\right) = NB(j - 1; \mu_{\text{Expand,t}}, \delta_{\text{Expand,t}})$$

and

$$P\left(J_{S,t} = j \middle| D_{S,t} = \text{Contract}\right) = NB(-j - 1; \mu_{\text{Conract,t}}, \delta_{\text{Contract,t}})$$

Furthermore, each distribution is parametrized with a mean and dispersion parameter depending on the dynamic and the time point, *t*: $\mu_{\text{Expand,t}}, \delta_{\text{Expand,t}}$ for expansions, and $\mu_{\text{Contract,t}}, \delta_{\text{Contract,t}}$ for contractions.

Of note, in negative binomial distributions the probabilities for negative integers are defined to be 0. Therefore, the above parametrization enforces that boundary movements of negative or zero length (i.e. contracting or steady, respectively) are impossible for expansions and that boundary movements of positive or zero length (i.e. expanding or steady) are impossible for contractions.

The ChromTime model additionally assumes that there is a prior probability to observe each dynamic between time points *t* and *t+1*, $P(D_{S,t} = d) = \pi_{t,d}$, which is the same at each side (left and right). Thus, the joint probability to observe a particular partition of the data into peaks and background and dynamics $d_{L,t}$ and $d_{R,t}$ on the left and right side at two consecutive time points, respectively, can be expressed as:

$$P\left(B_{L,t} = l_t, B_{L,t+1} = l_{t+1}, D_{L,t} = d_{L,t}, B_{R,t} = r_t, B_{R,t+1} = r_{t+1}, D_{R,t} = d_{R,t}, Z_t = 1, Z_{t+1} = 1, O_t, O_{t+1} \mid X_t, X_{t+1}\right) =$$
$$= P\left(B_{L,t} = l_t, B_{R,t} = r_t, Z_t = 1, O_t \middle| X_t\right) \times P\left(J_{L,t} = l_t - l_{t+1} \middle| D_{L,t} = d_{L,t}\right) \times P\left(D_{L,t} = d_{L,t}\right)$$
$$\times P\left(J_{R,t} = r_{t+1} - r_t \middle| D_{R,t} = d_{R,t}\right) \times P\left(D_{R,t} = d_{R,t}\right)$$
$$\times P\left(B_{L,t+1} = l_{t+1}, B_{R,t+1} = r_{t+1}, Z_{t+1} = 1, O_{t+1} \middle| X_{t+1}\right)$$

In a block $\boldsymbol{O}$ with covariates $\boldsymbol{X}$ in a time course with *T* time points, the total probability of a particular sequence of dynamics and boundary positions on the left side ($\mathfrak{D}_L$ and $\mathcal{B}_L$ respectively) and on the right side ($\mathfrak{D}_R$ and $\mathcal{B}_R$ respectively) can be expressed as:

$$P(\mathfrak{D}_L, \mathcal{B}_L, \mathfrak{D}_R, \mathcal{B}_R, \boldsymbol{O} | \boldsymbol{X}) =$$
$$P\left(B_{L,1} = l_1, B_{R,1} = r_1, Z_1 = 1, O_1 \middle| X_1\right) \prod_{t=2}^{T} \Big(P\left(B_{L,t} = l_t, B_{R,t} = r_t, Z_t = 1, O_t \middle| X_t\right)$$
$$\times P\left(J_{L,t-1} = l_{t-1} - l_t \middle| D_{L,t-1} = d_{L,t-1}\right) \times P\left(D_{L,t-1} = d_{L,t-1}\right)$$
$$\times P\left(J_{R,t-1} = r_t - r_{t-1} \middle| D_{R,t-1} = d_{R,t-1}\right) \times P\left(D_{R,t-1} = d_{R,t-1}\right)\Big)$$

where $\mathfrak{D}_S = (d_{S,t} | 1 \leq t \leq T-1)$ are *T-1* dimensional vectors of dynamics labels for each pair of consecutive time points on the left or the right side (*S* = *L* or *R*), respectively, and $\mathcal{B}_S$ are the

corresponding *T* dimensional vectors of boundary positions on each side at each time point. The total probability of the ChIP-seq signal in a block can be expressed as a sum over all possible sequences of dynamics and peak boundary positions that can generate the block across all time points. Thus, the probability of block $\boldsymbol{O}$ is:

$$P(\boldsymbol{O}|\boldsymbol{X}) = \sum_{\mathfrak{D}_L,\mathcal{B}_L,\mathfrak{D}_R,\mathcal{B}_R} P(\mathfrak{D}_L, \mathcal{B}_L, \mathfrak{D}_R, \mathcal{B}_R, \boldsymbol{O}|\boldsymbol{X})$$

where $\mathfrak{D}_L$ and $\mathfrak{D}_R$ each iterate over all possible 3^{T-1} combinations of peak boundary dynamics, and $\mathcal{B}_L$ and $\mathcal{B}_R$ each iterate over all possible ways to place left and right end boundaries across all time points that are consistent with the requirements that $B_{L,t} \leq B_{R,t} + 1$ at each time point. Let $\mathcal{O}$ be the total set of blocks in the data, $\mathcal{X}$ be the set of two-dimensional vectors containing the constant term and the log of the expected number of reads at each position and time point for each block, and *M* be the total number of blocks. Then the total likelihood of all blocks is:

$$P(\mathcal{O}|\mathcal{X}) = \prod_{i=1}^{M} P(\boldsymbol{O}_i|\boldsymbol{X}_i)$$

Model optimization

The total set of parameters of the model consists of:

1) Prior probabilities of each dynamic *d* at each time point *t*: $\pi_{t,d}$
2) Parameters of the negative binomial distributions that model the Peak and the Background components at each time point: α_t, β_t, γ_t and δ_t.
3) Parameters of the negative binomial distributions that model the boundary movements in Expand and Contract dynamics at each time point: $\mu_{\mathrm{Expand,t}}$, $\delta_{\mathrm{Expand,t}}$ and $\mu_{\mathrm{Contract,t}}$, $\delta_{\mathrm{Contract,t}}$, respectively.

The optimal parameter values are attempted to be estimated by Expectation Maximization (EM). In particular, ChromTime finds a local maximum of the expectation of the complete log-likelihood function (for details, see **Supplementary Methods**):

$$LL\left(\theta;\tilde{\theta},\mathcal{O},\mathcal{X}\right)=\sum_{i=1}^{M}\log P(\boldsymbol{O}_i|\boldsymbol{X}_i)=$$

$$=\sum_{i=1}^{M}\sum_{l_1=1}^{N_i+1}\sum_{r_1=l_1-1}^{N_i}P\left(B_{L,1}=l_1,B_{R,1}=r_1|\boldsymbol{O}_i,\boldsymbol{X}_i;\tilde{\theta}\right)\left(\log P\left(B_{L,1}=l_1\right)+\log P\left(B_{R,1}=r_1\right)\right)$$

$$+\sum_{i=1}^{M}\sum_{t=1}^{T}\sum_{p=1}^{N_i}\left(P\left(V_{t,p}=\mathrm{Peak}\middle|\boldsymbol{O}_i,\boldsymbol{X}_i;\tilde{\theta}\right)\log\left(\mathrm{NB}\left(O_{i,t,p};\mu_{\mathrm{Peak},t}=\exp\left[\alpha_t+\gamma_t\log\lambda_{i,t,p}\right],\delta_t\right)\right)\right.$$

$$\left.+P\left(V_{t,p}=\mathrm{Background}\middle|\boldsymbol{O}_i,\boldsymbol{X}_i;\tilde{\theta}\right)\log\left(\mathrm{NB}\left(O_{i,t,p};\mu_{\mathrm{Background},t}=\exp\left[\beta_t+\gamma_t\log\lambda_{i,t,p}\right],\delta_t\right)\right)\right)$$

$$+\sum_{i=1}^{M}\sum_{t=1}^{T-1}\sum_{S\in\{L,R\}}\sum_{d_{S,t}\in\left\{\substack{\mathrm{Expand}\\\mathrm{Contract}}\right\}}\sum_{j=1}^{N_i}P\left(J_{S,t}=(-1)^{\omega(d_{S,t})}j,D_{S,t}=d_{S,t}\middle|\boldsymbol{O}_i,\boldsymbol{X}_i;\tilde{\theta}\right)\log\left(\mathrm{NB}\left(j-1;\mu_{d_{S,t}},\delta_{d_{S,t}}\right)\right)$$

$$+\sum_{i=1}^{M}\sum_{t=1}^{T-1}\sum_{S\in\{L,R\}}\sum_{d_{S,t}\in\left\{\substack{\mathrm{Steady}\\\mathrm{Expand}\\\mathrm{Contract}}\right\}}P\left(D_{S,t}=d_{S,t}\middle|\boldsymbol{O}_i,\boldsymbol{X}_i;\tilde{\theta}\right)\log\left(\pi_{t,d_{S,t}}\right)$$

where

$$\omega\left(d_{S,t}\right)=\begin{cases}1 & \text{if } d_{S,t}=\mathrm{Contract}\\ 0 & \text{otherwise}\end{cases}$$

In the above equation, $\tilde{\theta}$ denotes the set of parameters at the current iteration, and θ denotes the set of parameters to be estimated in the next iteration. Additionally, $P\left(V_{t,p}=\mathrm{Peak}\middle|\boldsymbol{O}_i,\boldsymbol{X}_i;\tilde{\theta}\right)$ and $P\left(V_{t,p}=\mathrm{Background}\middle|\boldsymbol{O}_i,\boldsymbol{X}_i;\tilde{\theta}\right)$ denote the posterior probabilities that the bin of block *i* at position *p* and time point *t* is assigned to the peak and the background component, respectively:

$$P\left(V_{t,p} = \text{Peak} \middle| \boldsymbol{O}_i, \boldsymbol{X}_i; \tilde{\theta}\right) = \sum_{l,r|l \le p \le r} P\left(B_{L,t} = l, B_{R,t} = r, Z_t = 1 \middle| \boldsymbol{O}_i, \boldsymbol{X}_i; \tilde{\theta}\right)$$

$$P\left(V_{t,p} = \text{Background} \middle| \boldsymbol{O}_i, \boldsymbol{X}_i; \tilde{\theta}\right) = 1 - P\left(V_{t,p} = \text{Peak} \middle| \boldsymbol{O}_i, \boldsymbol{X}_i; \tilde{\theta}\right)$$

Parameter initialization

The values of the model parameters before the first EM iteration are initialized as follows:

1) All dynamics priors, $\pi_{t,d}$, are set uniformly to $\frac{1}{3}$;
2) All parameters for the distributions modelling the Peak and Background components, α_t, β_t, γ_t and δ_t, are set to 1;
3) All dispersion and mean parameters for the distributions modelling the boundary movements in Expand and Contract dynamics, $\mu_{\text{Expand,t}}$, $\delta_{\text{Expand,t}}$, $\mu_{\text{Contract,t}}$, and $\delta_{\text{Contract,t}}$, are set to 1.

At each time point, ChromTime requires that the left boundary of each non-zero length peak is placed before its right boundary. This requirement results in non-uniform prior probabilities for the $V_{t,p}$ variables, which model whether a bin at position *p* at time point *t* belongs to the peak or to the background component, making bins in the middle of the block more likely to be in a peak compared to flanking bins on each side. In particular, at any single time point the prior probability for a bin at position p $(1 \le p \le N)$ to be in a peak is:

$$P\left(V_{t,p} = \text{Peak} | Z_t = 1\right) = \frac{\#\ [\text{peak intervals containing bin } p \text{ in a block of length } N]}{\#\ [\text{total peak intervals in a block of length } N]}$$

$$= \frac{p(N-p+1)}{\sum_{i=1}^{N+1}(N-i+2)} = \frac{2p(N-p+1)}{(N+1)(N+2)}$$

which is symmetric with respect to the center of the block and has its maximum of $\frac{N}{2(N+1)} < \frac{1}{2}$ at $p = \frac{N}{2}$. This prior probability plays an important role during the learning stage of ChromTime, because it directs the model during the initial iterations of the EM to correctly associate the Peak component with high ChIP-seq signal and the Background component with low ChIP-seq signal. The assumption that high ChIP-seq

signal will more likely be located in the middle of blocks is motivated by the procedure that determines the block boundaries in the first phase of ChromTime, which naturally produces blocks with high signal in the middle compared to their flanking regions (**Fig S1D**). As result, no further efforts from the initialization or the training procedures are necessary in practice to identify correctly each component.

Expectation step

ChromTime provides an efficient implementation of the expectation step of EM based on a dynamic programming algorithm similar to the Baum-Welch algorithm for hidden Markov models. In brief, at each time point there are $O(N^2)$ ways to place the start and end positions of a peak, resulting in $O(N^4)$ combinations between any pair of consecutive time points. Thus, a standard forward-backward procedure that caches intermediate results can compute all expectations in $O(T*N^4)$ time and $O(T*N^2)$ memory. Since $O(T*N^4)$ time complexity can result in very long running times even for moderate N, ChromTime splits blocks that are longer than a predefined number of bins, MAX_BINS, (30 by default) into two halves (left and right) and estimates all sufficient statistics in each half independently. If block i is longer than MAX_BINS, the split is performed at the position with the highest average ChIP-seq signal across all time points in the block, K_i. This splitting procedure corresponds to imposing an additional constraint on the values of the boundary position variables that $B_{L,t} \leq K_i$ and $B_{R,t} \geq K_i - 1$ at each time point, t, while still having all bins between the left and the right boundaries annotated as peak bins (i.e. $V_{t,p} = \mathrm{Peak}$ for all p such that $B_{L,t} \leq p \leq B_{R,t}$ and $V_{t,p} = \mathrm{Background}$ for all other values of p). This heuristic reduces the time complexity to $O(T*N^2)$ and the memory footprint to $O(T*N)$, thus making the whole EM procedure run in feasible time and space. Since the $O(T*N^4)$ algorithm is applied only to blocks shorter than MAX_BINS bins, the total running time of ChromTime in a dataset of M blocks remains at most quadratic in the length of longer peaks in the data, $O(M*T*N^2)$.

For computational efficiency, if there are more than 10,000 blocks, ChromTime randomly selects 10,000 as input for the EM procedure.

Maximization step

The form of the complete log-likelihood implies that each set of model parameters can be optimized independently by solving for the roots of the respective partial derivatives. The dynamics prior probabilities are updated after each EM iteration as:

$$P\left(D_{L,t} = d\right) = P\left(D_{R,t} = d\right) = \frac{1}{2M}\sum_{i=1}^{M}\sum_{S\in\{L,R\}} P(D_{S,t} = d|\boldsymbol{O}_i, \boldsymbol{X}_i; \tilde{\theta})$$

The parameters of all negative binomial distributions in the model are determined by using an iterative numerical procedure that aims to find the roots of the partial derivatives of the complete log-likelihood function with respect to each parameter. In brief, finding the maximum likelihood estimators for each distribution at each EM iteration is equivalent to fitting weighted negative binomial regressions with the corresponding posterior probabilities as weights. For this purpose, ChromTime re-implements a widely used fitting method for weighted negative binomial regressions from the MASS package[57] in R except that the Iterative Reweighted Least Squares part is substituted with the HYBRD method for root finding from the MINPACK package[58] and the overall procedure is further extended to handle shared parameters. More details are provided in the **Supplementary Methods**.

Computing the most likely spatial dynamic and peak boundaries for each block across the whole time course

After the optimal values for all model parameters are estimated from the data, for each block the most likely positions of the peak boundaries at each time point are calculated. This procedure consists of two steps. First, ChromTime determines for each block all time points with significantly low probability of containing a false positive non-zero peak. Second, conditioned on those time points, ChromTime computes the most likely assignment of the peak boundary variables at each side and each time point.

During the first step, for each block and each time point ChromTime computes the posterior probability that the whole time point is modelled as background, $\varphi_{t,i} = \prod_{p=1}^{N_i} P(V_{t,p} = \text{Background}|\boldsymbol{O}_i, \boldsymbol{X}_i)$. This

probability can be interpreted as the probability of making a false positive non-zero length peak call as estimated by the model at time point *t* in block $\boldsymbol{O}_i$. To determine significant time points with low false positive probability, $\varphi_{t,i}$, ChromTime computes a time point specific threshold, τ_t, at a predefined false discovery rate (0.05 by default) by applying the standard Benjamini-Hochberg procedure[59] on all values of $\varphi_{t,i}$ from time point *t*.

In the second step, for each block ChromTime computes the most likely sequence of assignments of the boundary positions, conditioned on the event that all time points that failed to pass the FDR threshold in the previous step for the block are assigned to having no peaks. In particular, ChromTime executes a dynamic programming algorithm similar to the Viterbi algorithm for hidden Markov models, which uses the following recursive formula to find the most likely position for the peak boundaries at each side, *S* (Left or Right) and enforces that time points that failed the FDR test contain no peaks:

$$\mathrm{DP}_{t,l,r} = \begin{cases} \log P\left(B_{L,1} = l, B_{R,1} = r, O_1 \middle| X_1\right) & , t = 1 \\ \max\limits_{\substack{l_{t-1} \in [1,N+1] \\ r_{t-1} \in \begin{cases} [l_{t-1}-1,N] \text{ if } \varphi_{t,i} \le \tau_t \\ \{l_{t-1}-1\} \text{ otherwise} \end{cases}}} \begin{pmatrix} \mathrm{DP}_{t,l_{t-1},r_{t-1}} + \\ \log P\left(B_{L,t} = l, B_{R,t} = r, O_t \middle| X_t\right) + \\ \log P\left(J_{L,t-1} = l_{t-1} - l \middle| D_{L,t-1} = \mathrm{DYN}(l_{t-1} - l)\right) + \\ \log P\left(D_{L,t-1} = \mathrm{DYN}(l_{t-1} - l)\right) + \\ \log P\left(J_{R,t-1} = r - r_{t-1} \middle| D_{R,t-1} = \mathrm{DYN}(r - r_{t-1})\right) + \\ \log P\left(D_{R,t-1} = \mathrm{DYN}(r - r_{t-1})\right) \end{pmatrix} & , t > 1 \end{cases}$$

$$\text{where } \mathrm{DYN}(j) = \begin{cases} \text{Steady} & \text{if } j = 0 \\ \text{Expand} & \text{if } j \ge 1 \\ \text{Contract} & \text{if } j \le -1 \end{cases}$$

and DP denotes the dynamic programing cube of size $T(N+1)^2$ that stores the log likelihood for the best assignment of the peak boundary variables up to time point *t*. Tracing the DP cube from the highest value on row *T* back to row 1 retrieves the best assignment of the peak end variables. Similarly to the expectation step of the EM phase, for blocks longer than MAX_BINS bins the best Viterbi path is chosen among the splits at the top MAX_BINS positions in the block sorted by their average ChIP-seq signal across all time points. Since MAX_BINS is a predefined constant, the whole procedure has the same time

and space complexity as computing the expectations in the EM phase of ChromTime. The dynamic between any two time points is determined from the direction of the movement of the optimal positions of the corresponding boundaries.

Finally, the software provides an option to parallelize all computations during the EM phase and the Viterbi decoding over multiple CPU cores, which can further reduce the running time by up to a factor equal to the number of allocated cores. In our tests, processing ChIP-seq for the H3K4me2 mark and control input data from 5 time points in mouse T cell development[15] took 3 hours by using 4 CPU cores on a MacBook Pro laptop with 2.7GHz Intel Core i7 and 16 GB RAM.

ChromTime parameters used in this study

In this work, we applied ChromTime in narrow mode on all data for H3K4me2, H3K4me3, H3K27ac and H3K(9,14)ac marks. We applied ChromTime in broad mode on all data for H3K79me2, Pol2, H3K4me1, H3K27me3, H3K9me3 and H3K36me3 marks. All other parameters were set to their default values.

Analyses with External Data

The procedures for analyses with external are described in **Supplementary Methods.**

ABBREVIATIONS

HM: Histone mark

TF: Transcription factor

TSS: Transcription start-site

ChIP-seq: Chromatin immunoprecipitation coupled with high throughput DNA sequencing

FDR: False Discovery Rate

EM: Expectation Maximization

DECLARATIONS

Authors' contributions

J.E. conceived and supervised the project, designed the method, proposed analyses and wrote the manuscript. P.F. designed the method, implemented the software, proposed analyses, performed all analyses, processed all data and wrote the manuscript. Both authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

ChromTime software is freely available at: https://github.com/ernstlab/ChromTime

No new experimental datasets were generated within this study. All ChIP-seq datasets used as input for ChromTime are publicly available from the references listed in **Table 1**. Links to all other datasets (gene expression, TF binding and DNaseI hypersensitivity) used in this study are provided in the **Supplementary Methods**.

Funding

This work was supported by the CIRM Training Grant TG2-01169, the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Training Program (P.F.); NIH grants R01ES024995, U01HG007912, DP1DA044371, an NSF CAREER Award #1254200 and an Alfred P. Sloan Fellowship (J.E.).

Acknowledgements

We are grateful to Constantinos Chronis, Kathrin Plath and members of the Ernst lab for useful discussions.

Competing interests

None

REFERENCES

[1] J. Ernst *et al.*, “Mapping and analysis of chromatin state dynamics in nine human cell types.,” *Nature*, vol. 473, no. 7345, pp. 43–9, May 2011.

[2] A. Barski *et al.*, “High-resolution profiling of histone methylations in the human genome.,” *Cell*, vol. 129, no. 4, pp. 823–37, May 2007.

[3] T. S. Mikkelsen *et al.*, “Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.,” *Nature*, vol. 448, no. 7153, pp. 553–60, Aug. 2007.

[4] ENCODE_Project_Consortium, “An integrated encyclopedia of DNA elements in the human genome.,” *Nature*, vol. 489, no. 7414, pp. 57–74, Sep. 2012.

[5] Roadmap_Epigenomics_Consortium *et al.*, “Integrative analysis of 111 reference human epigenomes,” *Nature*, vol. 518, no. 7539, pp. 317–330, Feb. 2015.

[6] J. H. A. Martens and H. G. Stunnenberg, “BLUEPRINT: mapping human blood cell epigenomes.,” *Haematologica*, vol. 98, no. 10, pp. 1487–9, Oct. 2013.

[7] F. D. Lay *et al.*, “Reprogramming of the human intestinal epigenome by surgical tissue transposition.,” *Genome Res.*, vol. 24, no. 4, pp. 545–53, Apr. 2014.

[8] P. Fiziev *et al.*, “Systematic Epigenomic Analysis Reveals Chromatin States Associated with Melanoma Progression.,” *Cell Rep.*, vol. 19, no. 4, pp. 875–889, Apr. 2017.

[9] S. Mei *et al.*, “Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data

in human and mouse," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D658–D662, Jan. 2017.

[10] M. J. Ziller *et al.*, "Dissecting neural differentiation regulatory networks through epigenetic footprinting.," *Nature*, vol. 518, no. 7539, pp. 355–359, Dec. 2014.

[11] A. M. Tsankov *et al.*, "Transcription factor binding dynamics during human ES cell differentiation," *Nature*, vol. 518, no. 7539, pp. 344–349, Feb. 2015.

[12] K. K.-H. Farh *et al.*, "Genetic and epigenetic fine mapping of causal autoimmune disease variants," *Nature*, vol. 518, no. 7539, pp. 337–343, Oct. 2014.

[13] E. Gjoneska *et al.*, "Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease," *Nature*, vol. 518, no. 7539, pp. 365–369, Feb. 2015.

[14] A. Gusev *et al.*, "Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases.," *Am. J. Hum. Genet.*, vol. 95, no. 5, pp. 535–52, Nov. 2014.

[15] J. A. Zhang, A. Mortazavi, B. A. Williams, B. J. Wold, and E. V Rothenberg, "Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity.," *Cell*, vol. 149, no. 2, pp. 467–82, Apr. 2012.

[16] T. S. Mikkelsen *et al.*, "Comparative epigenomic analysis of murine and human adipogenesis.," *Cell*, vol. 143, no. 1, pp. 156–69, Oct. 2010.

[17] D. Lara-Astiaso *et al.*, "Chromatin state dynamics during blood formation," *Science (80-.).*, p. science.1256271-, Aug. 2014.

[18] D. K. Goode *et al.*, "Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation," *Dev. Cell*, Feb. 2016.

[19] S. L. Paige *et al.*, "A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development.," *Cell*, vol. 151, no. 1, pp. 221–32, Sep. 2012.

[20] J. A. Wamstad *et al.*, "Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage.," *Cell*, vol. 151, no. 1, pp. 206–20, Sep. 2012.

[21] D. Cacchiarelli *et al.*, "Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of

Induced Pluripotency," *Cell*, vol. 162, no. 2, pp. 412–424, Jul. 2015.

[22] J. Chen *et al.*, "Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming," *Cell Rep.*, Jan. 2016.

[23] R. P. Koche *et al.*, "Reprogramming factor expression initiates widespread targeted chromatin remodeling.," *Cell Stem Cell*, vol. 8, no. 1, pp. 96–105, Jan. 2011.

[24] C. Chronis *et al.*, "Cooperative Binding of Transcription Factors Orchestrates Reprogramming.," *Cell*, vol. 168, no. 3, p. 442–459.e20, Jan. 2017.

[25] O. Bogdanovic *et al.*, "Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis.," *Genome Res.*, vol. 22, no. 10, pp. 2043–53, Oct. 2012.

[26] N. Nègre *et al.*, "A cis-regulatory map of the Drosophila genome.," *Nature*, vol. 471, no. 7339, pp. 527–31, Mar. 2011.

[27] P. Yu *et al.*, "Spatiotemporal clustering of epigenome reveals rules of dynamic gene regulation.," *Genome Res.*, p. gr.144949.112-, Oct. 2012.

[28] P. Arnold *et al.*, "Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting.," *Genome Res.*, p. gr.142661.112-, Sep. 2012.

[29] N. Koike *et al.*, "Transcriptional architecture and chromatin landscape of the core circadian clock in mammals.," *Science*, vol. 338, no. 6105, pp. 349–54, Oct. 2012.

[30] R. Ostuni *et al.*, "Latent Enhancers Activated by Stimulation in Differentiated Cells," *Cell*, vol. 152, no. 1–2, pp. 157–71, Jan. 2013.

[31] S. K. Reilly *et al.*, "Evolutionary changes in promoter and enhancer activity during human corticogenesis," *Science (80-.).*, vol. 347, no. 6226, pp. 1155–1159, Mar. 2015.

[32] A. Weiner *et al.*, "High-Resolution Chromatin Dynamics during a Yeast Stress Response.," *Mol. Cell*, vol. 58, no. 2, pp. 371–386, Mar. 2015.

[33] J. Cotney *et al.*, "The evolution of lineage-specific regulatory activities in the human embryonic

limb.," *Cell*, vol. 154, no. 1, pp. 185–96, Jul. 2013.

[34] J. Zhu *et al.*, "A time-series analysis of altered histone H3 acetylation and gene expression during the course of MMAIII-induced malignant transformation of urinary bladder cells," *Carcinogenesis*, vol. 38, no. 4, pp. 378–390, Apr. 2017.

[35] Y. Zhang *et al.*, "Model-based analysis of ChIP-Seq (MACS).," *Genome Biol.*, vol. 9, no. 9, p. R137, Jan. 2008.

[36] S. Heinz *et al.*, "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.," *Mol. Cell*, vol. 38, no. 4, pp. 576–89, May 2010.

[37] S. Xu, S. Grullon, K. Ge, and W. Peng, "Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells.," *Methods Mol. Biol.*, vol. 1150, pp. 97–111, 2014.

[38] H. Xing *et al.*, "Genome-Wide Localization of Protein-DNA Binding and Histone Modification by a Bayesian Change-Point Method with ChIP-seq Data," *PLoS Comput. Biol.*, vol. 8, no. 7, p. e1002613, Jul. 2012.

[39] A. Harmanci, J. Rozowsky, and M. Gerstein, "MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework.," *Genome Biol.*, vol. 15, no. 10, p. 474, 2014.

[40] J. Ernst and M. Kellis, "ChromHMM: automating chromatin-state discovery and characterization.," *Nat. Methods*, vol. 9, no. 3, pp. 215–6, Mar. 2012.

[41] M. M. Hoffman, O. J. Buske, J. Wang, Z. Weng, J. A. Bilmes, and W. S. Noble, "Unsupervised pattern discovery in human chromatin structure through genomic segmentation.," *Nat. Methods*, vol. 9, no. 5, pp. 473–6, May 2012.

[42] M. Allhoff, K. Seré, H. Chauvistré, Q. Lin, M. Zenke, and I. G. Costa, "Detecting differential peaks in ChIP-seq signals with ODIN.," *Bioinformatics*, vol. 30, no. 24, pp. 3467–75, Dec. 2014.

[43] H. Xu, C.-L. Wei, F. Lin, and W.-K. Sung, “An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data.,” *Bioinformatics*, vol. 24, no. 20, pp. 2344–9, Oct. 2008.

[44] J. Biesinger, Y. Wang, and X. Xie, “Discovering and mapping chromatin states using a tree hidden Markov model,” *BMC Bioinformatics*, vol. 14, no. Suppl 5, p. S4, 2013.

[45] D. Hnisz *et al.*, “Super-enhancers in the control of cell identity and disease.,” *Cell*, vol. 155, no. 4, pp. 934–47, Nov. 2013.

[46] S. C. J. Parker *et al.*, “Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants,” *PNAS*, p. 1317023110-, Oct. 2013.

[47] B. A. Benayoun *et al.*, “H3K4me3 breadth is linked to cell identity and transcriptional consistency.,” *Cell*, vol. 158, no. 3, pp. 673–88, Jul. 2014.

[48] K. Chen *et al.*, “Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes.,” *Nat. Genet.*, vol. 47, no. 10, pp. 1149–57, Oct. 2015.

[49] A. Dincer *et al.*, “Deciphering H3K4me3 broad domains associated with gene-regulatory networks and conserved epigenomic landscapes in the human brain,” *Transl. Psychiatry*, vol. 5, no. 11, p. e679, Nov. 2015.

[50] R. D. Hawkins *et al.*, “Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.,” *Cell Stem Cell*, vol. 6, no. 5, pp. 479–91, May 2010.

[51] Z. Wang *et al.*, “Combinatorial patterns of histone acetylations and methylations in the human genome.,” *Nat. Genet.*, vol. 40, no. 7, pp. 897–903, Jul. 2008.

[52] H. H. Ng, F. Robert, R. A. Young, and K. Struhl, “Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity.,” *Mol. Cell*, vol. 11, no. 3, pp. 709–19, Mar. 2003.

[53] E. Smith, C. Lin, and A. Shilatifard, “The super elongation complex (SEC) and MLL in

development and disease.," *Genes Dev.*, vol. 25, no. 7, pp. 661–72, Apr. 2011.

[54] S. A. Lacadie, M. M. Ibrahim, S. A. Gokhale, and U. Ohler, "Divergent Transcription and Epigenetic Directionality of Human Promoters," *FEBS J.*, Apr. 2016.

[55] A. C. Cameron and P. K. Trivedi, *Regression analysis of count data*, Second edition. Cambridge University Press, 2013.

[56] N. U. Rashid, P. G. Giresi, J. G. Ibrahim, W. Sun, and J. D. Lieb, "ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions," *Genome Biol.*, vol. 12, no. 7, p. R67, 2011.

[57] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, Fourth. New York: Springer, 2002.

[58] Moré, B. Garbow, and K. Hillstrom, "User Guide for MINPACK-1," *ANL-80-74, Argonne Natl. Lab.*, 1980.

[59] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57. WileyRoyal Statistical Society, pp. 289–300, 1995.

[60] Z. Bar-Joseph, D. K. Gifford, and T. S. Jaakkola, "Fast optimal leaf ordering for hierarchical clustering.," *Bioinformatics*, vol. 17 Suppl 1, pp. S22-9, 2001.

FIGURE LEGENDS

Figure 1: **Overview of the ChromTime method. (A)** Examples of locations of peaks of the H3K4me2 mark that exhibit steady, expanding and contracting boundary dynamics, shown from left to right respectively, across five time points during T cell development in mouse[15]. Time points 1, 2 and 3 correspond to in vitro differentiated T cell precursors (FLDN1, FLDN2a, and FLDN2b), whereas time

points 4 and 5 correspond to in vivo purified thymocytes (ThyDN3 and ThyDP). Normalized ChIP-seq signal and called MACS2[35] peaks are shown at each time point. The bottom part of each plot shows the corresponding ChromTime output. A peak upstream of Zfp148 gene remains relatively steady with fluctuations of MACS2 peak boundaries across time. In contrast, a peak at the Skap1/GM11529 promoter expands gradually in both directions relative to its location at the first time point. Conversely, a peak upstream of GPR141 gene contracts towards a narrow genomic interval at the last time point. **(B)** Schematic workflow of the ChromTime method. During the block finding stage, the input ChIP-seq and, optionally, control reads are used to determine genomic positions of blocks of signal enrichments across time points. In the dynamics prediction stage, for each block, peak boundary positions are predicted at each time point and peak boundary dynamics are predicted at each pair of consecutive time points. **(C)** Schematic of the dynamics prediction stage for one block. Boxes represent genomic bins at each time point. The corresponding ChIP-seq signal is depicted as blue bars for each bin whose height represents the number of reads that map to the bin. During the dynamics prediction stage, the ChromTime method learns a probabilistic mixture model from the input data in order to partition each block at each time point into peak and background components. Genomic bins in the peak component (orange) mark ChIP-seq peaks whereas those in the background component (white) mark flanking background signal. The movement of the boundaries on the left and the right side of peaks between consecutive time points are estimated by reasoning jointly about the input data from all time points.

Figure 2: **Sample output from ChromTime with contracting peaks.** Genome browser screenshot with sample output of ChromTime for H3K4me2 from the T cell development time course in mouse[15] with 5 time points at the Esam/Vsig2/Nrgn locus. The input ChIP-seq signal and MACS2[35] peaks (black boxes under each signal track) are shown in the upper panel of the screenshot. The ChromTime predicted peaks colored by their boundary dynamics for each block at each time point are shown in the bottom panel. The first peak in each block is colored in dark grey. Each subsequent peak is colored with respect to the predicted dynamic relative to its previous time point. Peaks with steady boundaries on both sides are shown in light grey, and those with at least one contracting boundary are shown in blue. Nearby peaks that touch boundaries are visualized as one peak by the genome browser. Not shown in the figure

are expanding peaks, peaks at single time points and peaks with opposite dynamics (Expand on the left and Contract on the right, or vice versa), which would be colored in red, orange and black, respectively. See **Fig S2** for examples of predicted expanding peaks.

Figure 3: **Enrichments of TF binding events and DNaseI hypersensitivity sites in ChromTime predictions.** For data from mouse T cell development[15] and human stem cell reprogramming[21], the following three sets of ChromTime peaks are examined: 1) peaks present at all time points whose boundaries hold steady on both sides of the peak throughout the whole time course (T1-Tn Steady); 2) non-contracting peaks that expand between at least one pair of consecutive time points and have a peak at the last time point of the time course (Tx-Tn Expand); and 3) non-expanding peaks that contract between at least one pair of consecutive time points and have a peak at the first time point of the time course (T1-Tx Contract). The first column in each table shows the percentage of bases covered by each set out of all bases covered by predicted peaks. The last row shows the baseline percentage of bases covered by each genomic annotation out of all bases covered by ChromTime peaks at any time point for each mark. Percentages are color coded from 0 (white) to 100 (green). Fold enrichments in each column are color coded from 1 (white) to the maximum value in the column (red). **(A)** Fold enrichments of cell type specific and constitutive peaks of the GATA3 transcription factor, which is a master regulator in T cell differentiation, in predicted H3K4me2 peaks during T cell development in mouse[15]. **(B)** As in **(A)** for DNaseI hypersensitive sites, CEBP, POL2 and RAD21 binding in predicted H3K4me3 dynamics during stem cell reprogramming of fibroblast cells in human[21]. **(C)** As in **(A-B)** for DNaseI hypersensitive sites, CEBP, P300 in predicted H3K27ac dynamics in stem cell reprogramming in human[21]. In addition, enrichments are shown for OCT4 and NANOG, which are key transcription factors responsible for maintaining stem cell identity.

Figure 4: Predicted spatial dynamics by ChromTime associate better with gene expression changes compared to boundary position changes of peaks from individual time points called in isolation. For **(A)** H3K4me2 in mouse T cell development[15] and **(B)** H3K4me3 in human stem cell reprogramming[21], the plots show differences of the average gene expression ranks of block boundaries

ranked based on the length of boundary movement of ChromTime peaks called at individual time points in isolation and of block boundaries ranked based on ChromTime posterior probabilities computed by using data from the whole time course (see **Supplementary Methods**) for expansions (left plots) and contractions (right plots) overlapping annotated TSSs. To rank block boundaries based on the boundary movements of ChromTime peaks called in isolation, block boundaries were sorted by the difference in the boundary positions of peaks called in isolation between consecutive time points in descending order for expansions and in ascending order for contractions. Positive values in the plots correspond to ranks for which rankings based on ChromTime posterior probabilities associate on average with larger gene expression changes compared to rankings based on boundary movements of peaks called in isolation between consecutive time points, and vice versa for negative values. The black lines and the shaded regions show the difference of the average ranks for a set of randomized rankings of the block boundaries and their 95% confidence intervals, respectively.

Figure 5: **Spatial dynamics can contain additional information about gene expression changes beyond ChIP-seq signal density changes.** Gene expression change is plotted as function of ChIP-seq signal density change after loess smoothing for each predicted ChromTime dynamic for **(A)** H3K4me2 dynamics in T cell development in mouse[15]; and **(B)** in H3K4me3 dynamics in stem cell reprogramming in human[21]. Peaks of each type of dynamics were pooled from all time points for this analysis. Peaks with asymmetric dynamics E/S and S/E were pooled together in the “E-S” group. Similarly C/S and S/C peaks were pooled in the “C-S” group. In both systems, peaks with the same signal density change can associate with different gene expression changes depending on the predicted spatial dynamic. Shaded regions represent 95% confidence intervals.

Figure 6: Spatial dynamics of multiple different HMs co-localize within a time course. Hierarchical clustering with optimal leaf ordering[60] of the geometric average fold enrichments taken across all time points of the overlap of every pair of predicted spatial dynamics for mapped HMs in **(A)** mouse adipogenesis[16] and **(B)** human stem cell reprogramming[21]. At each pair of time points, “Expand” and “Contract” dynamics are defined as all peaks that are predicted as either unidirectional or bidirectional

expansions and contractions, respectively, whereas “Steady” dynamics are defined as all peaks that have predicted steady boundaries at both sides. Peaks with opposing dynamics at each side (i.e. E/C and C/E) were excluded from this analysis. In both datasets, expansions, contractions and steady peaks of H3K4me1, H3K4me2, H3K4me3 and H3K27ac tend to cluster together within each of the three classes, whereas spatial dynamics of H3K27me3 and H3K36me3 peaks tend to occupy different locations. All enrichments were capped at 50.

Figure 7: **Direction of asymmetric dynamics correlates with direction of transcription. (A) (i)** Left panel shows a schematic representation of unidirectional expansions that expand in the same direction as transcription and in the opposite direction of transcription. The adjacent plots show, for each mark, the average $\log_2$ ratio between the fraction of unidirectional expansions that expand in the same directions as transcription of the nearest gene and the fraction of unidirectional expansions that expand in the opposite direction of transcription of the nearest gene for blocks that are within 1 kb of annotated TSSs and for more distal blocks. Positive values correspond to enrichment of unidirectional expansions in the same direction as transcription. For marks mapped in at least six time courses, a black line is plotted representing the average across all data sets and significant differences are denoted with asterisks based on a two-tailed Mann-Whitney test at a p-value threshold of 0.05. **(ii)** Left panel shows analogous schematic for unidirectional contractions. Likewise, adjacent plots show, for each mark, the average ratio between the fraction of unidirectional contractions that contract in the opposite direction of transcription of the nearest TSS and unidirectional contractions that contract in the same direction as transcription of the nearest TSS. **(B)** Left panel shows an example of unidirectional expansions between pairs of time points that expand in the same direction as transcription at the Hs6st1 gene of the H3K4me2 mark in the T cell development dataset[15]. Right panel shows an example of unidirectional contractions between pairs of time points that contract in the opposite direction of transcription at the DNMT3B gene of the H3K4me2 mark in the T cell development dataset. The predicted ChromTime peaks colored by their boundary dynamics are shown under the signal track for each time point.

Table 1: Datasets used for analysis with ChromTime

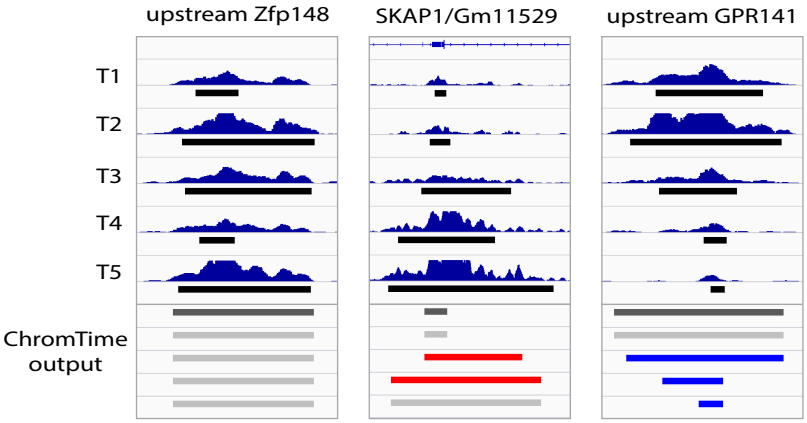
System	Histone Marks	Species	# Time Points	Reference
Adipogenesis	H3K4me2 H3K4me3 H3K27ac H3K4me1 H3K36me3 H3K27me3	Mouse Human	4	[16]
Blood Formation	H3K4me2 H3K4me3 H3K27ac H3K4me1	Mouse	5-7	[17]
Cardiac Development	H3K4me3 H3K27ac H3K4me1 H3K27me3 H3K36me3 Pol2	Mouse	4	[20]
Cardiac Development	H3K4me3 H3K27me3 H3K36me3	Human	5	[19]
Embryogenesis	H3K4me3 H3K27ac H3K4me1	Zebrafish	4	[25]
Macrophage Differentiation	H3K4me3	Mouse	5	[18]

	H3K9ac H3K27ac H3K27me3			
Neural Differentiation	H3K4me3 H3K27ac H3K27me3 H3K4me1	Human	5	[10]
Stem Cell Reprogramming	H3K4me2 H3K4me3 H3K27ac H3K4me1 H3K27me3 H3K36me3	Human	4-6	[21]
Stem Cell Reprogramming	H3K4me2 H3K4me3 H3K9ac H3K27ac H3K27me3 H3K36me3 H3K4me1 H3K79me2 H3K9me3	Mouse	4	[24]
Stem Cell Reprogramming	H3K4me3 H3K27ac H3K4me1 H3K27me3 Pol2	Mouse	9	[22]
T Cell Development	H3K4me2	Mouse	5	[15]

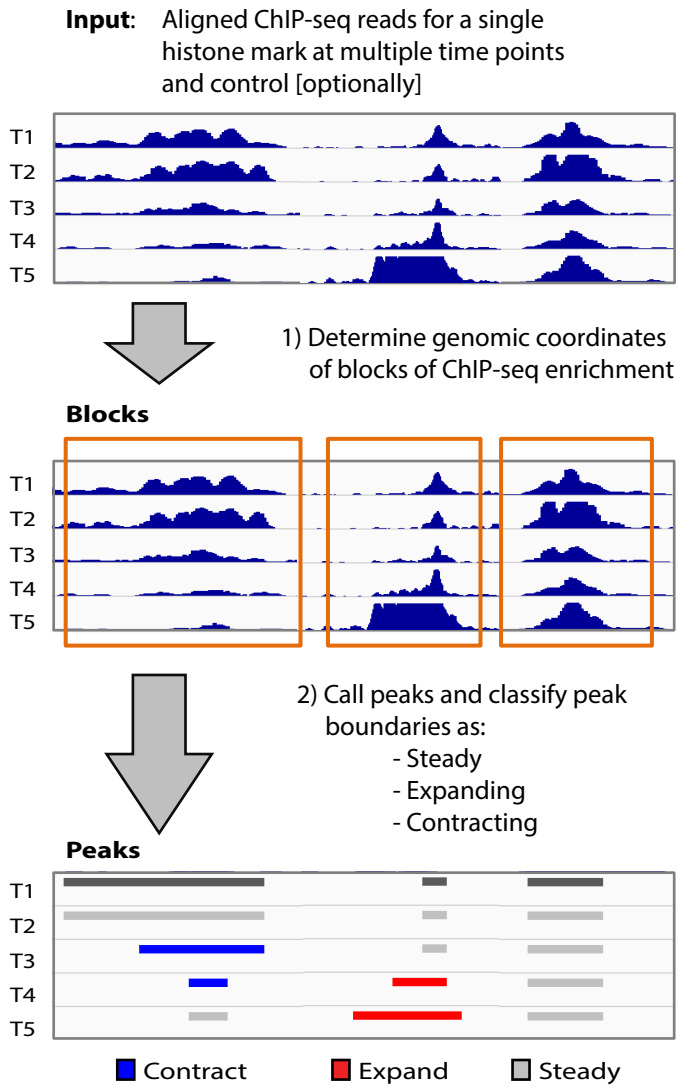
	H3K(9,14)ac H3K27me3			

Figure 1

A.


upstream Zfp148
SKAP1/Gm11529
upstream GPR141
T1
T2
T3
T4
T5
ChromTime output

B.


Input: Aligned ChIP-seq reads for a single histone mark at multiple time points and control [optionally]
T1
T2
T3
T4
T5
1) Determine genomic coordinates of blocks of ChIP-seq enrichment
Blocks
2) Call peaks and classify peak boundaries as:
- Steady
- Expanding
- Contracting
Peaks
Contract
Expand
Steady

C.

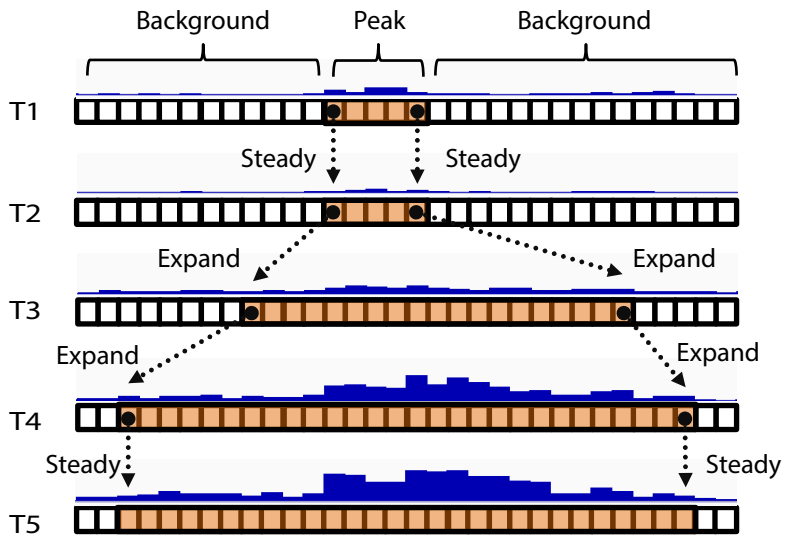

Background
Peak
Background
T1
Steady
Steady
T2
Expand
Expand
T3
Expand
Expand
T4
Steady
Steady
T5

Figure 2

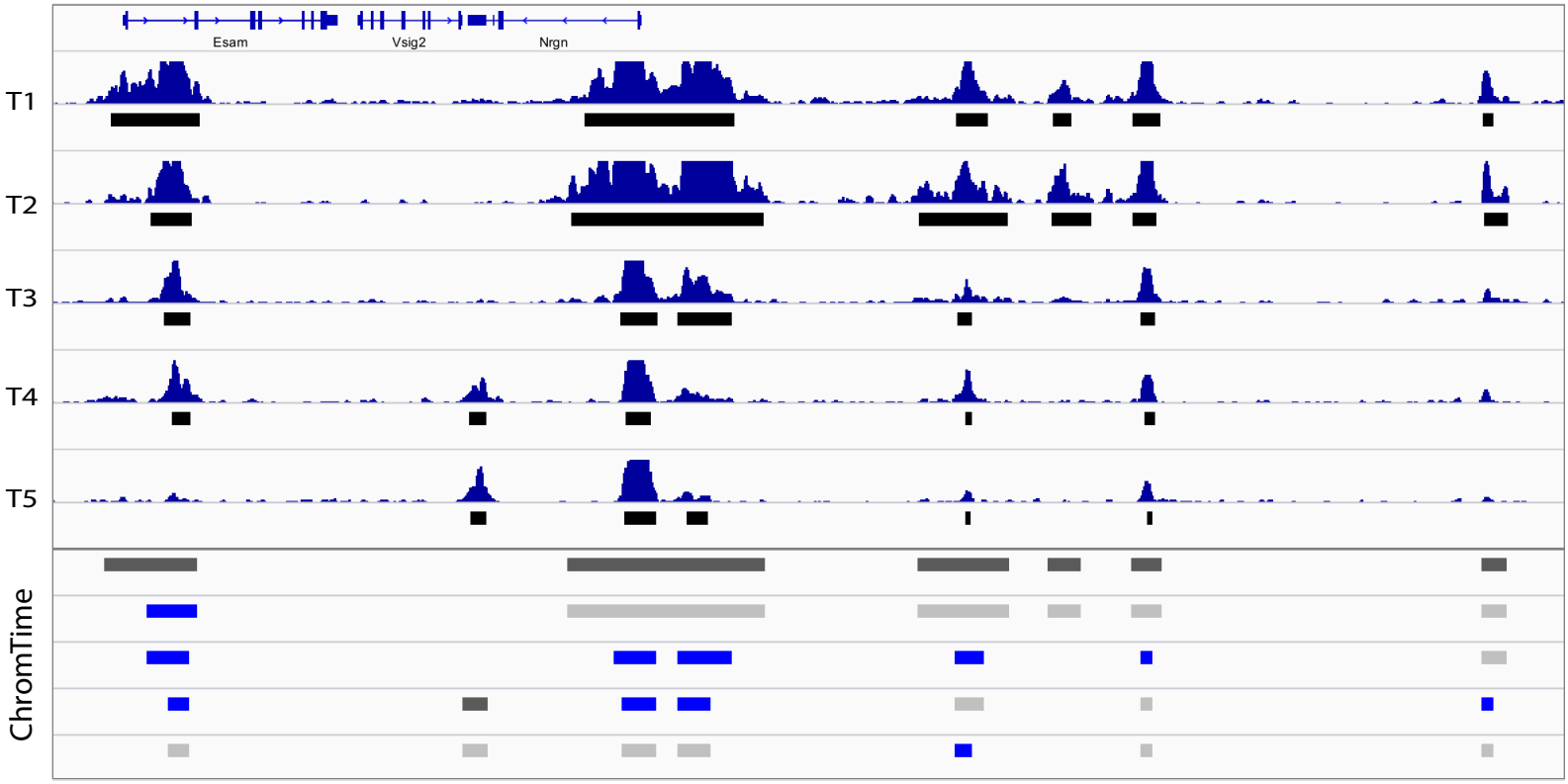


Figure 3

A.

Dynamic	% all block bases	FLDN1-ThyDP shared GATA3	ThyDP-specific GATA3	FLDN1-specific GATA3
T1-Tn Steady	28	1.2	1.1	0.5
Tx-Tn Expand	13	0.6	3.4	0.2
T1-Tx Contract	55	1	0.3	1.5
Base %		0.1	0.2	0.1

B.

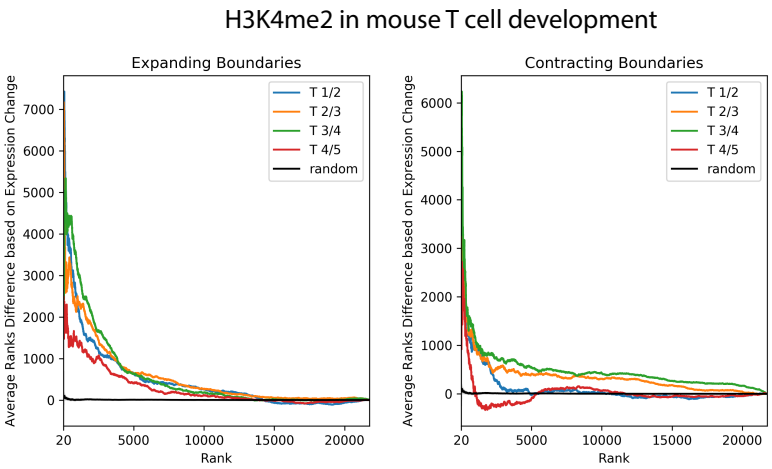
Dynamic	% all block bases	H1-IMR90 shared DNaseI	H1-IMR90 shared CEBPB	H1-IMR90 shared POL2	H1-IMR90 shared RAD21	H1-specific DNaseI	H1-specific CEBPB	H1-specific POL2	H1-specific RAD21	IMR90-specific CEBPB	IMR90-specific POL2	IMR90-specific RAD21	IMR90-specific DNaseI
T1-Tn Steady	26	1.5	1.5	1.8	1.3	0.6	0.9	1.2	0.9	1.2	1.2	1.2	0.8
Tx-Tn Expand	38	0.7	0.7	0.4	0.9	1.9	1.6	1.3	1.6	0.4	0.4	0.5	0.3
T1-Tx Contract	22	1.2	1.1	1.3	1	0.3	0.4	0.5	0.4	1.9	1.9	1.8	2.4
Base %		18	0.7	4.5	1.4	11	0.3	3	0.9	1.7	4.2	1	3.1

C.

Dynamic	% all block bases	H1-IMR90 shared DNaseI	H1-IMR90 shared CEBPB	H1-IMR90 shared P300	H1-specific DNaseI	H1-specific CEBPB	H1-specific P300	H1 NANOG	H1 OCT4	IMR90-specific DNaseI	IMR90-specific CEBPB	IMR90-specific P300
T1-Tn Steady	2	3.6	3.7	1.4	0.8	1.2	1.3	1.6	1.3	0.6	1.2	1.7
Tx-Tn Expand	14	1.4	1.4	1	3	3.5	2	3.6	3.8	0.2	0.3	0.3
T1-Tx Contract	26	0.5	0.7	1	0.2	0.2	0.3	0.2	0.2	2.3	2	2.2
Base %		10	0.5	31	4.9	0.2	7.9	0.4	0.3	7.8	2.5	1.5

Figure 4

A.


H3K4me2 in mouse T cell development
Expanding Boundaries
Contracting Boundaries
Average Ranks Difference based on Expression Change
Rank
T 1/2
T 2/3
T 3/4
T 4/5
random

B.

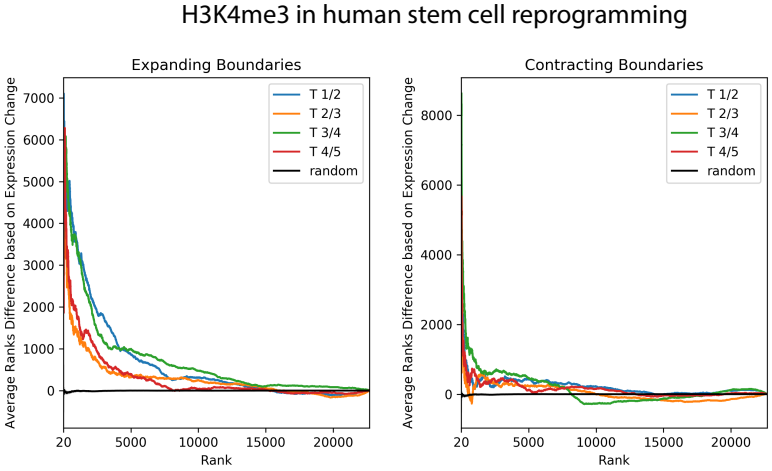

H3K4me3 in human stem cell reprogramming
Expanding Boundaries
Contracting Boundaries
Average Ranks Difference based on Expression Change
Rank
T 1/2
T 2/3
T 3/4
T 4/5
random

Figure 5

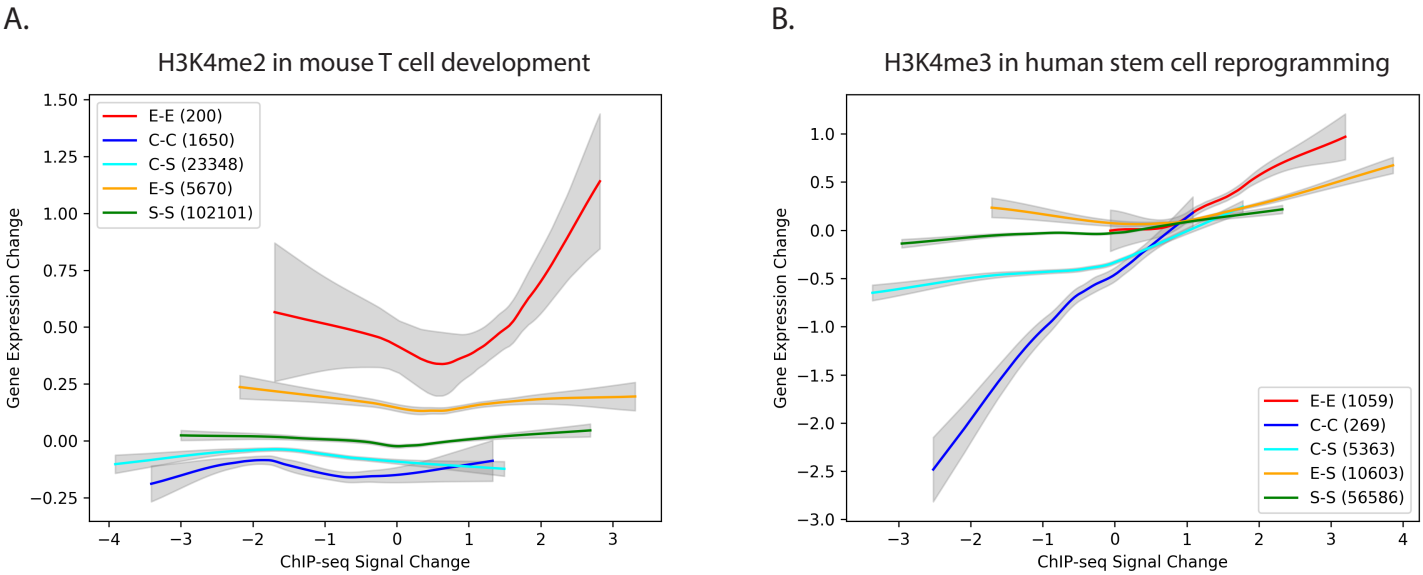
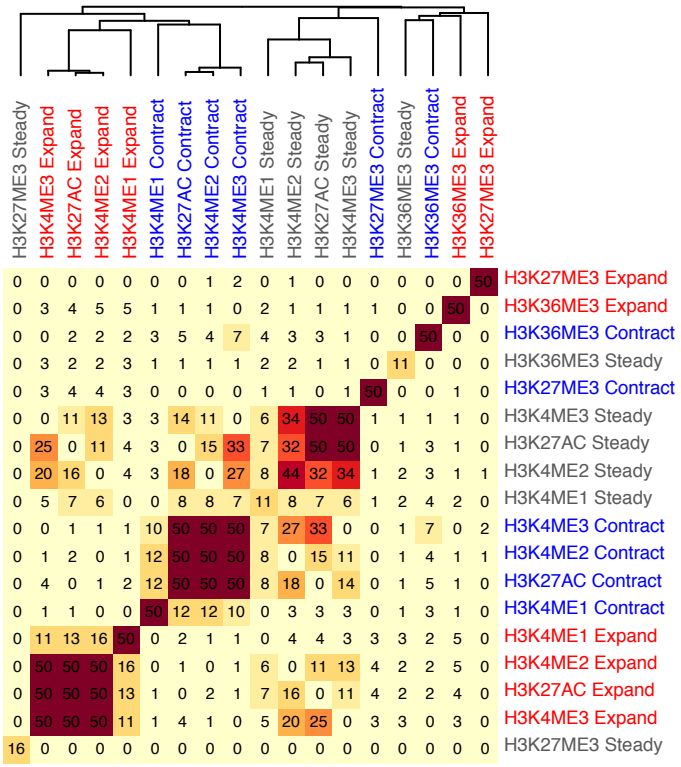
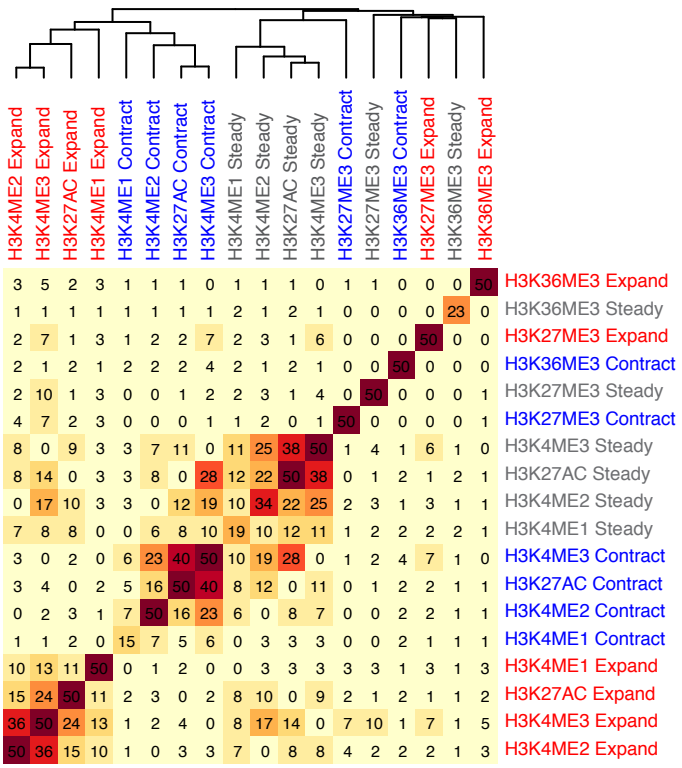

A.
H3K4me2 in mouse T cell development
E-E (200)
C-C (1650)
C-S (23348)
E-S (5670)
S-S (102101)
Gene Expression Change
ChIP-seq Signal Change
B.
H3K4me3 in human stem cell reprogramming
E-E (1059)
C-C (269)
C-S (5363)
E-S (10603)
S-S (56586)
Gene Expression Change
ChIP-seq Signal Change

Figure 6

A.


H3K27ME3 Steady
H3K4ME3 Expand
H3K27AC Expand
H3K4ME2 Expand
H3K4ME1 Expand
H3K4ME1 Contract
H3K27AC Contract
H3K4ME2 Contract
H3K4ME3 Contract
H3K4ME1 Steady
H3K4ME2 Steady
H3K27AC Steady
H3K4ME3 Steady
H3K27ME3 Contract
H3K36ME3 Steady
H3K36ME3 Contract
H3K36ME3 Expand
H3K27ME3 Expand
H3K27ME3 Expand 0 0 0 0 0 0 0 1 2 0 1 0 0 0 0 0 50
H3K36ME3 Expand 0 3 4 5 5 1 1 1 0 2 1 1 1 1 0 0 50 0
H3K36ME3 Contract 0 0 2 2 2 3 5 4 7 4 3 3 1 0 0 50 0 0
H3K36ME3 Steady 0 3 2 2 3 1 1 1 1 2 2 1 1 0 11 0 0 0
H3K27ME3 Contract 0 3 4 4 3 0 0 0 0 1 1 0 1 50 0 0 1 0
H3K4ME3 Steady 0 0 11 13 3 3 14 11 0 6 34 50 50 1 1 1 1 0
H3K27AC Steady 0 25 0 11 4 3 0 15 33 7 32 50 50 0 1 3 1 0
H3K4ME2 Steady 0 20 16 0 4 3 18 0 27 8 44 32 34 1 2 3 1 1
H3K4ME1 Steady 0 5 7 6 0 0 8 8 7 11 8 7 6 1 2 4 2 0
H3K4ME3 Contract 0 0 1 1 1 10 50 50 50 7 27 33 0 0 1 7 0 2
H3K4ME2 Contract 0 1 2 0 1 12 50 50 50 8 0 15 11 0 1 4 1 1
H3K27AC Contract 0 4 0 1 2 12 50 50 50 8 18 0 14 0 1 5 1 0
H3K4ME1 Contract 0 1 1 0 0 50 12 12 10 0 3 3 3 0 1 3 1 0
H3K4ME1 Expand 0 11 13 16 50 0 2 1 1 0 4 4 3 3 3 2 5 0
H3K4ME2 Expand 0 50 50 50 16 0 1 0 1 6 0 11 13 4 2 2 5 0
H3K27AC Expand 0 50 50 50 13 1 0 2 1 7 16 0 11 4 2 2 4 0
H3K4ME3 Expand 0 50 50 50 11 1 4 1 0 5 20 25 0 3 3 0 3 0
H3K27ME3 Steady 16 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

B.


H3K4ME2 Expand
H3K4ME3 Expand
H3K27AC Expand
H3K4ME1 Expand
H3K4ME1 Contract
H3K4ME2 Contract
H3K27AC Contract
H3K4ME3 Contract
H3K4ME1 Steady
H3K4ME2 Steady
H3K27AC Steady
H3K4ME3 Steady
H3K27ME3 Contract
H3K27ME3 Steady
H3K36ME3 Contract
H3K27ME3 Expand
H3K36ME3 Steady
H3K36ME3 Expand
H3K36ME3 Expand 3 5 2 3 1 1 1 0 1 1 1 0 1 1 0 0 0 50
H3K36ME3 Steady 1 1 1 1 1 1 1 1 2 1 2 1 0 0 0 0 23 0
H3K27ME3 Expand 2 7 1 3 1 2 2 7 2 3 1 6 0 0 0 50 0 0
H3K36ME3 Contract 2 1 2 1 2 2 2 4 2 1 2 1 0 0 50 0 0 0
H3K27ME3 Steady 2 10 1 3 0 0 1 2 2 3 1 4 0 50 0 0 0 1
H3K27ME3 Contract 4 7 2 3 0 0 0 1 1 2 0 1 50 0 0 0 0 1
H3K4ME3 Steady 8 0 9 3 3 7 11 0 11 25 38 50 1 4 1 6 1 0
H3K27AC Steady 8 14 0 3 3 8 0 28 12 22 50 38 0 1 2 1 2 1
H3K4ME2 Steady 0 17 10 3 3 0 12 19 10 34 22 25 2 3 1 3 1 1
H3K4ME1 Steady 7 8 8 0 0 6 8 10 19 10 12 11 1 2 2 2 2 1
H3K4ME3 Contract 3 0 2 0 6 23 40 50 10 19 28 0 1 2 4 7 1 0
H3K27AC Contract 3 4 0 2 5 16 50 40 8 12 0 11 0 1 2 2 1 1
H3K4ME2 Contract 0 2 3 1 7 50 16 23 6 0 8 7 0 0 2 2 1 1
H3K4ME1 Contract 1 1 2 0 15 7 5 6 0 3 3 3 0 0 2 1 1 1
H3K4ME1 Expand 10 13 11 50 0 1 2 0 0 3 3 3 3 3 1 3 1 3
H3K27AC Expand 15 24 50 11 2 3 0 2 8 10 0 9 2 1 2 1 1 2
H3K4ME3 Expand 36 50 24 13 1 2 4 0 8 17 14 0 7 10 1 7 1 5
H3K4ME2 Expand 50 36 15 10 1 0 3 3 7 0 8 8 4 2 2 2 1 3

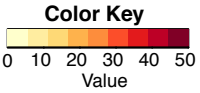

Color Key
0 10 20 30 40 50
Value

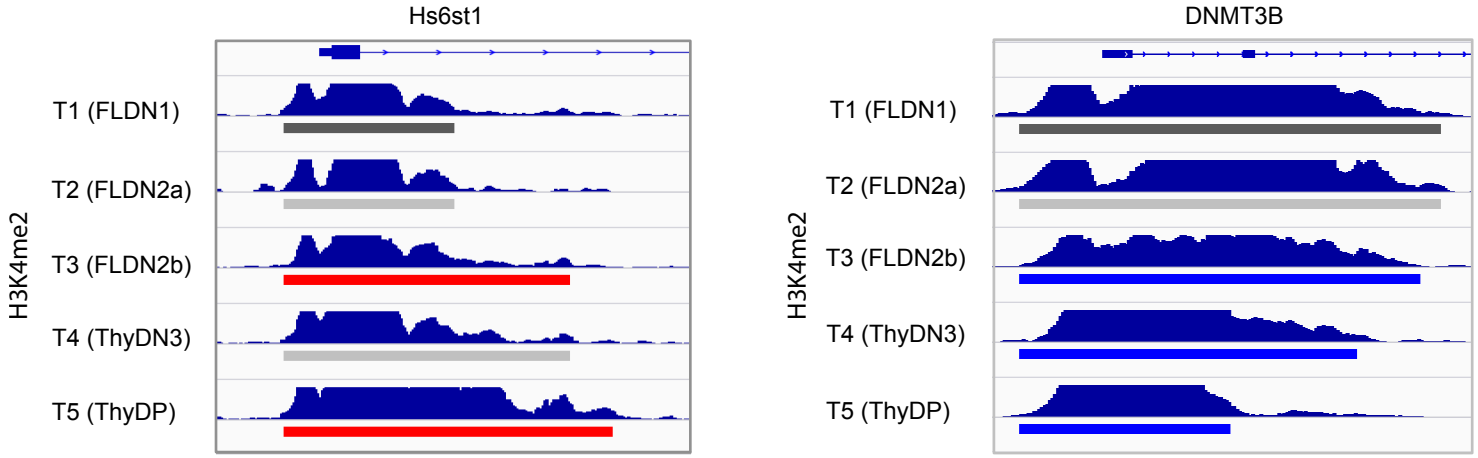
Figure 7

A.

i)

Unidirectional Expansions
Transcription
TSS
S|E Steady Expand
E|S Expand Steady
H3K79ME2
H3K4ME3
Pol2
H3K4ME2
H3K(9/14)AC
H3K4ME1
H3K27AC
H3K27ME3
H3K9ME3
H3K36ME3
Log2 % S|E / % E|S
TSS +-1kb
Distal
ii)
Unidirectional Contractions
Transcription
TSS
S|C Steady Contract
C|S Contract Steady
Log2 % S|C / % C|S
Dataset
Mouse Reprogramming (Chronis et al 2017)
Human Adipogenesis
Human Neural Differentiation
Mouse Cardiac Development
Human Cardiac Development
Mouse Reprogramming (Chen et al 2016)
Human Reprogramming
Mouse Adipogenesis
Mouse Macrophage Differentiation
Mouse Hematopoiesis
Mouse T-cell Development
Zebrafish Embryogenesis

B.


Hs6st1
DNMT3B
H3K4me2
T1 (FLDN1)
T2 (FLDN2a)
T3 (FLDN2b)
T4 (ThyDN3)
T5 (ThyDP)