Supplementary Materials for:

Discovery and characterization of chromatin states for systematic annotation of the human genome.

Jason Ernst[1,2]*, Manolis Kellis*[1,2]
[1]*MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, Massachusetts 02139, USA*
[2]*Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA*

Overview of Supplementary Materials:

C. Properties of chromatin states

D. Predictive power for gene annotation

E. Recovery of chromatin states using different combinations of marks and in additional cell types

Supplementary Notes

1. Capturing Genomic Spatial Context Information

Inspecting the transition matrix of the HMM (see full transition matrix at **Supplementary Fig. 4**) highlighted the value of incorporating spatial information, as the use of state-to-state transition probabilities were highly non-uniform, with a large majority of transition probabilities between states being very small (83% are below 0.005), with only a handful of important state transitions receiving high probabilities for each state (see top transitions at **Supplementary Fig. 5**). By inspecting the transition matrix, several notable findings emerge:

- Upstream promoter states (states 1-3) are most likely to transition to other promoter states, or to active intergenic states (right panel of **Supplementary Fig. 4**), while downstream promoter states (states 9-11) are more likely to transition to other promoter states or to transcribed states, illustrating the transition from active intergenic to upstream promoter, to downstream promoter, and to transcribed states along the body of the gene. (Note that transitions upstream or downstream along the body of a gene contribute equally to the transition matrix, as no transcriptional directionality was imposed in parsing the genome).
- The repressed promoter state (state 4), is the only state to transition to any of the large-scale repressed states, specifically state 45 (right panel of **Supplementary Fig. 4**).
- State 26 which enriches in transcribed regions but dips relative to exons, transitions most frequently to States 24 and 25 which enrich in exons relative to introns (**Supplementary Figure 24a**). Similarly, repressive states 43 and 44 transition frequently to each other, and also show opposite enrichments relative to exons and introns (**Supplementary Figure 24b**). This suggests perhaps an alternation between exonic and intronic states along the body of genes.
- The CTCF island state (State 39) is found most frequently transitioning to other active intergenic states (particularly States 36-38) as well as the H3K27me3 enriched repressive states (particulary State 43), but interestingly not the H3K9me3 associated repressive states. This may suggest that CTCF, which is thought to act an insulator, is playing a role in insulation within more dynamic regions (involving active marks and the repressive H3K27me3 mark), but not in more stably repressed regions (thought to be associated with heterochromatin and H3K9me3).
- The L1/LTR repeat enriched state (state 47), characterized by a dominant H3K9me3 mark, is found to most frequently transition to the broad H3K27me domains (state 43), suggesting the presence of certain H3K9me3-marked repetitive elements within or adjacent to broader H3K27me3 domains, and thus that even though the two marks typically do not overlap, they may be proximal to each other in the genome at least for these repetitive elements as observed by the transition matrix.
- Lastly, we note that the transition matrix helps define large groups of promoter, transcribed, enhancer, repressed, and repetitive states, with significantly higher within-group transitions than outside-group transitions (right panel of **Supplementary Fig. 4**), and also subgroups within each group with most frequent within-subgroup transitions (boxed areas in left panel of **Supplementary Fig. 4**). These groups and subgroups tend to share many additional biological functions (**Supplementary Table 1**), validating the biological interpretability of the learned transition parameters.

These are just some of the many features of the epigenome that can be extracted by close inspection of the transition matrix, highlighting both its importance in guiding the learning of our model, and also its direct interpretability in understanding the chromatin modification landscape.

It is notable that many of these spatial associations persisted over much longer distances than those reported here for neighboring intervals (**Supplementary Fig. 6**). While the most intense peaks of the transition matrix become more diffuse at longer intervals, strong non-random spatial associations are observed at distances of 2kb, 10kb, and 20kb, revealing substantial pairwise dependencies even at long distances, and further highlighting the importance of incorporating spatial information in the study of chromatin.

2. Establishing the number of biologically-relevant chromatin states

We sought to evaluate what number of chromatin states provides an appropriate resolution at which to interpret combinations of chromatin marks and their biological function. While we found distinct functional interpretations for each of 51 chromatin states in the text, additional chromatin marks and additional independent experimental datasets may reveal further meaningful subdivisions, while conversely increasingly finer-grain distinctions provided by additional states may be of decreased biological interest. To address these questions, we took three approaches for studying how distinguishable the 51 chromatin states that we described here are, the extent to which they capture mark co-occurrence patterns, and how frequently they are recovered with varying numbers of states and different initializations for parameter learning.

First, we asked how distinct different chromatin states are from each other in their genome-wide assignments. The probabilistic nature of the multivariate HMM allowed us to directly quantify the likelihood of overlap in the genome-wide assignments of any pair of states. For every location in the genome, we evaluated the posterior probability of each of the 51 states, summed over all possible parses of the genome. We then computed for each state *i* its posterior overlap with each state *j* defined to be a weighted average of the posterior probabilities of state *j* where the weighting is based on the posterior of state *i* in the interval. If the chromatin state assignment of a region is not of high confidence, we would expect many different states to all show similar posterior probability that could be as low as 2% for a truly uncertain assignment given 51 states. Instead, we found that on average the model confidence level in the assignment for 49 of the 51 states was at least 50% and for 28 states at least 75% (**Supplementary Fig. 9**). Thus the states described here are distinct both in their biological enrichments and also in their confident assignments.

Second**,** we evaluated how increasing numbers of states capture the genome-wide dependencies between chromatin marks. If two or more marks work together to define a chromatin state, they should show a strong genome-wide dependencies, namely they should occur more frequently together than one would expect based solely on their total abundance. However, if the chromatin state assignments correctly capture these dependencies and assign regions defined by their combination into the same chromatin state, these marks should then become conditionally independent within those states, namely they should occur together within the state at the frequency dictated by the product of their individual probabilities. Indeed the 51-state model showed pairs of marks occurring as expected by their individual frequencies (**Supplementary Fig. 10**), while models with fewer states showed pairs marks co-occurring more frequently than expected as evidence of un-captured dependencies (**Supplementary Fig. 11**), evidence that the chromatin states defined here have effectively captured pairwise dependencies of chromatin marks by explicitly grouping significant mark combinations in individual states.

Third**,** we evaluated the consistency of chromatin states in models learned at varying complexity and across different initializations, quantified as the correlation of chromatin mark frequencies obtained for corresponding states across different models (see **Online Methods**). We found that the emission parameters of the 51 states described here were highly correlated with states of the highest-scoring 79-state model (**Supplementary Fig. 12**). In general we found that the states recovered in a model of the nested initialization procedure were also consistently recovered in larger models, which was not the case for random initialized models. For instance based on nested initializations, the CTCF island state was recovered in all models with 24 or more states and the simple repeat enriched state was recovered in all models with at least 35 states, while under the highest scoring of three random initializations there were still models with 69 and 41 states respectively that did not recover these states though some randomly initialized models with fewer states did (**Supplementary Fig. 13**). In several cases when considering increasing sized models learned from nested initializations a jump in the best correlation for a specific state in the 79-state model corresponded to a clear jump in the maximal enrichment for specific types of genomic elements such as Zinc Finger Genes, TG simple repeats, and Transcription End Sites (**Supplementary Fig. 14).** Lastly, we confirmed the 51 chromatin states described here were highly representative of the 510 states obtained after training 10 independent random initializations of 51-state models (**Supplementary Fig. 15**), showing the desirable property of high coverage of local maximum state-space variability.

Overall, the 51 chromatin states described here have captured much of the complexity of a 79-state model with significantly fewer states, thus eliminating potentially redundant states. Moreover, the direct comparison of which biological states are recovered at each number of states enables us to select models that capture biologically-meaningful chromatin states we recognize, while including in an unbiased way all chromatin states captured with them. In our case, we selected a 51-state model which was the first to sufficiently capture the end of transcription state (State 27) (**Supplementary Figure 14**).

3. General Validation of Modeling Approach

Upon inspection of the learned model and its parameters, we verified several desirable properties.

- First, we found that the descriptive power of the model was appropriately spent, allocating more states to capture biologically-meaningful complexity in small regions, for example dedicating 11 states (1-11) to capture the subtleties of promoter-associated regions that only cover 1% of the genome, while two states (41 and 43) associated with large-scale repressed regions cover 46% of the genome (**Supplementary Fig. 16**).
- Second, we found that the emission parameters learned showed distinct combinations of chromatin marks, spanning a wide spectrum of combinations (**Supplementary Figs. 2 and 17**).
- Third, we found that the frequency at which the various chromatin marks would be considered detected in the states are highly correlated at the 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} Poisson distribution thresholds (**Supplementary Table 2 and Supplementary Fig. 18**), indicating that the chromatin mark combinations learned are robust across three orders of magnitude in the probability cutoff.
- Fourth, we found that adjusting our thresholds for each mark locally based on the density of nucleosome tags[1] did not affect the state definitions **(Supplementary Fig. 19).** For each mark, we compared the same sequence tag count as before, but we locally adjusted the mean λ used by the Poisson distribution to calculate the read threshold for each mark. Instead of using the genome average for the mark, λ, we used a scaled genome-wide average, based on the local density of nucleosome reads. Intuitively, we increased λ for an interval if there was an enrichment of nucleosome tags found locally, thus requiring more tags for the mark for it to reach the 10^{-4} threshold. More specifically, we computed the nucleosome-adjusted background by scaling the genome-wide average number of tags for each mark in a given interval I by the neighborhood nucleosome read count c_I mapping in a 1kb window centered at the interval I (applying the same tag-shift procedure for the ChIP-seq data described in the methods), divided by c_G, the genome average value of c_I.
- Fifth, we found that chromatin states captured variations in the intensity levels for each chromatin mark in their raw tag count enrichments as well as relative to an IgG control, both of which were highly correlated with the emission parameters of the mark over states **(Supplementary Fig. 20**).
- Lastly, we found even beyond the intensity levels used in making the binary presence/absence decision for each mark, chromatin states capture information on individual mark intensity levels both below and above the tag count thresholds used **(Supplementary Fig. 21)**, likely because our model considers both combinations of marks and spatial information. We separately considered the tag enrichments for those intervals for which the mark was called present and absent, and found that in both cases the tag enrichment is higher in states that have the mark called present at higher frequency and lower outside them, correlating highly with the emission parameters.

4. Transcription Associated States

The transition frequencies between different transcribed states (**Supplementary Fig. 4**) suggest that these states can be divided into four sub-groups and two additional isolated states, with no transition probability greater than 0.03 between states in different subgroups.

- The first subgroup (states 12-16) is characterized by higher frequency of H3K79me2 and H3K79me3 relative to H3K79me1, and is strongly enriched for 5’ proximal region of higher expressed genes (**Figure 3c**). Some of these states also showed enrichment for transcription factors and DNaseI hypersensitive sites making them candidates for being enhancer regions that are also transcribed.

- The second subgroup (states 17-19) is characterized by higher frequency of H3K79me1 relative to H3K79me2 and H3K79me3, and is found on average in 5' proximal region of lower-expression genes.
- The third subgroup (States 20-23) was characterized with lower frequency for H3K79me2 and H3K79me3 and higher frequency of H2BK5me1, H420Kme1, H3K4me1, and various acetylations. These states showed high enrichments for DNaseI hypersensitive regions, GC-rich areas and CpG islands, proximity to both 5' and 3' ends of genes.
- The fourth subgroup (States 24-26) is characterized by relatively high levels of H3K36me3, and is associated with transcribed regions of genes distal to the 5' end (**Figure 3d)**.

States 27-28 are discussed in the main text. **Supplementary Table 1** contains a more complete discussion of differences between states in each subgroup.

5. Active Intergenic Associated States

The eleven active intergenic states, states 29-39 states can be divided into a group of eight states (States 29-36) associated as being either candidate enhancer regions (states 29-33) or being proximal to them (states 34-36) and three additional states (States 37-39) also not associated with pronounced repression of downstream genes as seen with large scale repressed states, states 41-45 (**Supplementary Fig. 22**).

States 34-36 all had lower enrichments for DNaseI hypersensitive sites and transcription factor binding than states 29-33. Of these states, State 34 was the most enriched for being proximal to some of the strongest candidate enhancer states such as States 29 and 30 (**Supplementary Fig. 6**), as well as on average genes of substantially induced expression levels **(Supplementary Fig. 22)**. States 35 and 36 represented acetylation domains most frequently marked by H2AK5ac, H4K91ac, and H3K4ac and were also proximal to candidate enhancer regions and genes with above-average expression.

State 37 represented large domains of low modification frequency that tended to be far from repressed genes. State 38 and 39 would frequently transition to this state (**Supplementary Fig 4**). State 39, corresponds to candidate insulator regions. It showed the highest frequency of CTCF insulator protein binding and enrichment in the associated CTCF motif. It was also enriched for DNaseI hypersensitive sites and transcription factor binding, suggesting potential interactions of CTCF with several other factors either directly in these regions or through looping. State 38 had the highest frequency for H2AZ. State 39 showed strong enrichment for being proximal to two other active intergenic states 31 and 38, both of which had a relatively high frequency for H2AZ (**Supplementary Figs. 2 and 6**), suggesting a potential association between insulators and this histone variants.

6. Predictive power comparison with individual mark intensities and alternative methods

We used the recovery of RefSeq TSS and transcribed regions to gauge the importance of using chromatin mark combinations and spatial genomic information in a *de novo* unsupervised learning approach, compared to individual chromatin mark intensities and alternate methodological approaches.

First, we compared the recovery power of chromatin states and of individual chromatin marks at the binary cutoff dictated by our Poisson threshold for both classes of elements. In both cases, we found that chromatin states consistently surpassed all individual chromatin marks (**Figure 5**), and dramatically so for transcribed regions.

- For promoter regions, H3K4me3 provided a very good predictor, capturing 57% of TSS regions in 1.2% of the genome, lacking however the refinement of 11 distinct promoter classes with distinct positional and functional properties (**Figure 5a**).
- For transcribed intervals, no single mark input surpassed 7% recovery while transcribed states together accounted for nearly 40% of transcripts at <3% false positive rate (**Figure 5b**).

We also assessed the discovery power of chromatin marks at varying thresholds, by considering the signal intensity provided by read counts for each chromatin mark (**Supplementary Fig. 32**):

- For promoter regions, while H3K4me3 performed similar to chromatin states at the binary threshold chosen (Poisson cutoff at 10^{-4}), the full ROC curve at varying signal intensity levels shows that it does not achieve comparable power at higher specificity values (**Supplementary Fig. 32a**). The two other marks most closely associated with promoter regions, Pol2, and H3K9ac, both significantly underperformed chromatin states across the ROC curve.
- For transcribed regions, individual chromatin marks continued to perform significantly below chromatin states, even when varying mark intensity thresholds were considered (**Supplementary Fig. 32b**), again emphasizing the importance of multiple mark combinations and spatial context information.

We next compared chromatin states to two alternative approaches, k-means clustering and logistic regression (**Supplementary Fig. 32c,d**).

- The k-means comparison enabled us to gauge the importance of genomic context information encoded in our transition matrix, by comparing the discovery power of chromatin states to that of a k-means clustering approach with the same input binarization and same number of 51 clusters (**Supplementary Fig. 32c,d**). We found that chromatin states outperformed k-means clustering for both promoter regions (approximately a 20% increase in the true positive rate for a large range of false positive rates), and for transcribed regions (nearly 50% increase in the true positive rate). These results further highlight the importance of spatial context information, especially for long-range features such as transcribed regions.
- We also found that the *de novo* learning of chromatin states did not substantially hurt their performance compared to the supervised learning approach that specifically sought combinations of local chromatin mark signals that maximize prediction of promoter and transcribed locations (**Supplementary Fig. 32c,d**). For promoters, chromatin states performed comparably to a logistic regression supervised learning approach (less than a 4% drop in performance), and for transcribed intervals, they significantly outperformed logistic regression (nearly 25% higher performance despite the *de novo* learning for chromatin states) by being able to take advantage of spatial information despite having no prior knowledge of gene annotations for training.

Method comparison: The k-means clustering was performed using the fastkmeans implementation[2] on the same binarized input used with the HMM, but without any spatial information. The supervised logistic regression predictions were based on the TR-IRLS implementation[3] of logistic regression using the default settings except the cgdeveps parameter was set to 0.0001. The features to the classifier were ln(x+1) transformed values of the raw number of tags mapped to a 200bp interval for each mark, and thus had no spatial information. Results for the classifier are based on five-fold cross validation.

7. Additive and Combinatorial Relationships of Marks

We sought to understand the importance of different marks and mark combinations in defining chromatin states. This revealed both additive and combinatorial relationships between different chromatin marks. Acetylation marks in promoter states seemed to play a largely additive role, with higher levels of diverse acetylation marks consistently associated with higher expression. However, in active intergenic regions, acetylations showed a more complex behavior, with different combinations of acetylation marks (H2BK120ac, H2BK20ac, H2BK5ac, H3K27ac, H4K8ac, H4K91ac) acting as a primary determinant of candidate enhancer states and differences in downstream expression levels (States 29-33). Methylation marks seemed to play more combinatorial roles in defining chromatin states. One such example is found between H3K9me3, H4K20me3 and H3K36me3 which together help define repetitive states 47/48 and the ZNF-enriched state 28 (**Supplementary Fig. 39**). The enrichment of satellite repeat elements varied dramatically with different combinations of the three marks: State 47 (H3K9me3 alone) showed 0.5-fold enrichment, State 48 (H3K9me3 and H4K20me3) showed 63-fold enrichment, and State 28 (H3K9me3, H4K20me3, and H3K36me3) was back at 0.7-fold enrichment, suggesting a complex relationship that cannot be explained by a strictly additive association of any single mark and repeat elements. Similarly, enrichment for ZNF genes goes from 11-fold in State 47 to 112-fold in State 28 with the addition of H3K36me3, even though H3K36me3 is found as a dominant mark in states with a substantially weaker enrichment for ZNF genes (e.g. States 24-26).

8: Chromatin state recovery using subsets of marks

For the recovery of the chromatin state assignments based on the subset of 10 marks example given in **Supplementary Fig. 39**, we found an average of 77.2% sensitivity and 76.9% specificity averaged across the genome though these values varied dramatically across different states, from above 90% for repressed states 40 (defined by a general lack of marks) and 41 (associated with H3K9me3 but lacking most marks) to below 10% for candidate insulator state 39 (unsurprisingly since its major determinant CTCF was not included in the 10 marks surveyed). With the subset of 10 marks, promoter states 1-11 and candidate enhancer states 29-33 showed on average low sensitivity (48% and 35%, respectively), with the exception of repressed promoter state 4 that showed 72% sensitivity as most of its defining marks were profiled. Large state grouping were generally preserved however, with promoter states recognized as such and candidate enhancer states as such, suggesting that it was the subtleties of different promoter and candidate enhancer states that were lost.

We also evaluated the subset of 39 input datasets that excludes CTCF and PolII, to evaluate how much information lies strictly in histone marks and histone variants alone (**Supplementary Fig. 40**). In particular, we asked to what extent the CTCF island state, TSS and promoter states, and transcribed states could be identified without CTCF and Pol2.

- For the first question, we found that the histone marks alone are likely insufficient to demarcate regions of CTCF binding as only 9% of the state recovered in the absence of CTCF/PolII.
- For the second question, we found that promoter states were strikingly well identified in the absence of PolII/CTCF. 95-100% of promoter states remained promoter states even in the absence of PolII/CTCF information. Moreover, individual promoter states 1-11 preserved their exact identity 88-95% of the time without PolII/CTCF as an input.
- Similarly, all transcribed states were assigned to transcribed states in the absence of PolII/CTCF 97-100% of the time, and all but one preserved their specific state identity 93-99% of the time. The one exception was state 27, the transcription end state, which was recovered only 69% of the time, although it remained assigned to a transcribed state 98% of the time.
- We further assessed whether state 27 still peaked at transcription end sites even when PolII/CTCF were not part of the input features, and indeed state 27 was both the most highly enriched state at TES, and conversely TES regions were where state 27 was most enriched. However, the enrichment was reduced from 12.5-fold to 8.75-fold (becoming comparable to states 21 and 23 that showed enrichments of 8.1 and 8.3 respectively). Without Pol2 information, the peak of state 27 on TES remained but became less pronounced, though its enrichment still precipitously dropped after the end of the transcript (**Supplementary Fig. 41**).
- We further assessed the association of state 27 with transcription end sites in CD36 and CD133 cells, using the subset of 10 marks available in those cell types. This analysis confirmed that in both cell types, state 27 peaked strongly at the TES, further confirming its validity (**Supplementary Fig. 41**).

Chromatin state characterization

Promoter states:

State	Shared State Descriptions	State Description with defining marks and candidate biological interpretation	
1	**Promoter Upstream States; Potential enhancer looping.** State 1-3 had high overall frequency of H3K4me1/2/3, H3K9me1, H2AZ, but had a lower frequency for specific methylation marks such as H3K79me2, H3K79me3, H3K27me1, and H4K20me1 than found in other promoter associated states. These states all enriched in the promoter regions particularly upstream of the TSS. These states are associated with high enrichments for open chromatin and transcription factor binding. A portion of this state may also correspond to distal enhancers also having promoter marks possibly due to looping.	**Promoter upstream high expression; Potential enhancer looping**. State 1 relative to states 2 and 3 had higher frequency of all acetylation marks. 51% of this state was located within 2kb of a RefSeq annotated transcription start site (TSS), and when found in promoter regions was more likely to be found upstream of the TSS and associated with higher expressed genes. Relative to States 2 and 3 this state had greater enrichments for open chromatin and experimental transcription factor binding except for the repressive NRSF.	
2		**Promoter upstream medium expression; Potential enhancer looping**. State 2 had an intermediate frequency for detection of all acetylation marks as compared to states 1 and 3. About 41% of this state was located within 2kb of a RefSeq annotated TSS. When found in promoter regions, this state was more likely to be found upstream of the TSS, but there was less of a bias upstream as compared to State 1. The genes downstream of this state had expression levels in between that of State 1 and 3.	
3		**Promoter upstream low expression; Potential enhancer looping**. Relative to states 1 and 2, this state had lower frequency for all acetylations. 52% of this state was located within 2kb of a RefSeq annotated TSS, and when found in promoter regions was almost equally likely to be found upstream and downstream of the TSS. This state was associated with lower expressed genes. Of genes with a TSS most likely in this state, there was an enrichment for cell cycle related genes.	
4	**Repressed Promoter.** State 4 had the greatest frequency for H3K4me3 relative to other marks in the state. It is distinguished from other promoter states in its very low frequency (<= 0.02) of detection for all acetylations, and its relatively higher frequency for H3K27me3. 57% of this state was within 2kb of a TSS. Genes with this state in its promoter region were generally repressed. This was the most enriched state for the repressive NRSF transcription factor. Genes with a TSS in this state enriched for Gene Ontology categories related to embryonic development.		
5	**TSS states.** States 5-7 had high frequency for H3K4me3 and Pol II, but had lower frequency for H3K4me1 and other methylation marks found in other promoter enriched states. These three states had the highest enrichment for TSS of any states.	**TSS low-medium expression; most GC rich**. State 5 compared to states 6 and 7 had a lower frequency for all acetylations. 74% of this state was found within 2kb of a RefSeq annotated transcription start site. This state and genes immediately downstream from it had a lower average expression than state 7 and to a lesser extent state 6. Chromatin is an example of a GO category genes with a TSS in this state had a greater enrichment for than states 6 and 7. Of all states this state had the highest GC content.	
6		**TSS medium expression.** State 6 had a medium frequency of acetylation relative to states 5 and 7. About 78% of this state is located within 2kb of an annotated transcription start site. The average expression level of this state and genes immediately downstream of it was lower than that of state 7, and slightly higher than that of state 5. Response to DNA stimulus is an example of a GO category genes with a TSS in this state had a greater enrichment for than states 5 and 7.	
7		**TSS high expression.** State 7 relative to States 5 and 6 had a higher frequency for all acetylations. This state was 89% within 2kb of a RefSeq annotated TSS. The genes in this state were on average higher expressed than states 5 and 6. RNA processing is an example of a GO category genes with a TSS in this state had a greater enrichment for than states 5 and 6.	
8	**Transcribed Promoter States.** State 8-11 had high frequency H3K4me3 and for some or all of the methylations H3K4me1/2, H3K9me1, H3K79me1/2/3, H3K27me1, H2BK5me1, and H4K20me1. These states were all enriched in the promoter region particularly downstream of the TSS. These four states had higher average expression than the other promoter associated states.	**Transcribed promoter highest expression.** State 8 and 9 had a higher frequency for H3K79me2/3, a lower frequency for most other methylations, and higher average expression than states 10 and 11.	**Transcribed promoter; highest expression, TSS for T-cell activation genes**. State 8 had higher frequency of acetylations than state 9. State 8 was 71.5% within 2kb of a RefSeq annotated TSS. Relative to state 9 this state was found closer to the TSS and had higher enrichments for transcription factor binding and open chromatin. The genes with TSS most likely in this state enriched for cell type specific categories such as T-cell activation.
9			**Transcribed promoter; highest expression, downstream**. State 9 had lower frequency of acetylations than state 8. State 9 was 41% within 2kb of a RefSeq annotated TSS. This state was more likely to be found downstream of the TSS. This state had the highest average expression of any state. Relative to state 8 it was found even further downstream of the TSS and had lower enrichments for transcription factor binding and open chromatin.
10		**Transcribed promoter high expression.** State 10 and 11 had a lower frequency for H3K79me2/3, a higher frequency for most other methylations, and lower average expression than states 8 and 9.	**Transcribed promoter; high expression, near TSS.** State 10 had higher frequency for acetylations as compared to state 11. State 10 was found closer to the TSS than state 11, and had higher relative enrichments for transcription factor binding and open chromatin.
11			**Transcribed promoter; high expression, downstream.** State 11 had lower frequency for acetylations as compared to state 10. State 11 was found further from the TSS than states 10, and had lower relative enrichments for transcription factor binding and open chromatin.

Transcribed States:

12	**Transcribed 5'proximal States.** States 12-16 had low frequency for H3K4me3, higher frequency of H3K79me2/3 relative to H3K79me1, and would frequently transition between each other. As a group these five states tended to be relatively proximal to the 5' end of genes and higher expressed.	**Higher expression.** States 12 and 13 relative to 14-16 had higher frequency for a number of methylation marks including H3K4me1, H3K4me2, H3K9me1, H3K79me1, H3K27me1, H2BK5me1, and H4K20me1 and were more highly expressed.	**Transcribed 5'proximal; higher expression, open chromatin, TF binding, candidate enhancer.** State 12 relative to state 13 had higher frequency for all acetylations and greater enrichment for open chromatin and transcription factor binding.
13			**Transcribed 5'proximal; higher expression, open chromatin, candidate weak enhancer.** State 13 relative to state 12 had lower frequency for all acetylations and open chromatin and transcription factor binding.
14		**High and medium expression.** States 14-16 relative to 12-13 had lower frequency for a number of methylation marks including H3K4me1, H3K4me2, H3K9me1, H3K79me1, H3K27me1, H2BK5me1, and H4K20me1 and were less highly expressed.	**Transcribed 5'proximal; high expression, open chromatin; candidate enhancer.** State 14 relative to states 15 and 16 had higher frequency for acetylations and H3K4me1, and also greater enrichment for open chromatin and transcription factor binding.
15			**Transcribed 5' proximal; high expression.** State 15 had lower acetylation frequencies than state 14 and consistently higher methylation frequencies than state 16. State 15 had lower enrichments for open chromatin and transcription factor binding than state 14, and had higher expression than state 16. The enrichment for Alu repeat elements was between that of state 14 and 16.
16			**Transcribed 5' proximal; medium expression, Alu repeats.** State 16 had lower detected acetylation and methylation frequencies than states 14 and 15. This state had lower average expression, lower enrichments for open chromatin and transcription factor binding, and was the most enriched of all states for Alu repeat elements.
17	**Transcribe less 5' proximal States.** States 17-19 all had low frequency for H3K4me3, higher frequency for H3K79me1 relative to H3K79me2 and H3K79me3 and would frequently transition between each other. These states were also found relatively proximal to the 5' end of genes, but to a lesser extent than States 17-19. These states were more likely associated with lower expressed genes.		**Transcribed less 5'proximal, medium expression; open chromatin; candidate weak enhancer.** State 17 relative to States 18 and 19 had a higher frequency for methylation marks such as H3K4me1/2, H3K9me1, H3K79me1/2/3, H3K27me1, H2BK5me1, H4K20me1, H3K36me3, and acetylations. State 17 relative to States 18 and 19 had higher average expression, greater open chromatin and transcription factor binding enrichments, and fewer Alu repeat elements.
18			**Transcribed less 5' proximal, medium expression.** State 18 had low frequency for acetylations, and relative to States 17 and 19 had an intermediate frequency for methylation marks such as H3K4me1/2, H3K9me1, H3K79me1/2/3, H3K27me1, H2BK5me1, H4K20me1, H3K36me3. State 18 relative to States 17 and 19 had an intermediate average expression level, intermediate open chromatin and transcription factor binding enrichments, and intermediate enrichments for Alu repeat elements.
19			**Transcribed less 5' proximal, lower expression; Alu repeats**. State 19 had low frequency for acetylations, and relative to States 18 and 19 had lower frequency for methylation marks such as H3K4me1/2, H3K9me1, H3K79me1/2/3, H3K27me1, H2BK5me1, H4K20me1, H3K36me3. State 19 relative to States 17 and 18 had a lower average expression level, lower open chromatin and transcription factor binding enrichments, and greater enrichments for Alu repeat elements.
20	**Candidate strong enhancer in transcribed regions**. State 20 had the highest frequency for H3K4me1/2, H3K9me1, and various acetlylations and to a lesser extent other methylations such as H2BK5me1, H4K20me1, H3K79me1, H3K27me1. This state had the greatest enrichment for open chromatin and transcription factor binding among the transcribed states. This state had higher GC levels than all transcribed states except States 21-23. States 21-22 had a higher frequency for this state than H2BK5me1 and H4K20me1, but lower frequency for H3K4me1, H3K4me2, and various acetylations.		
21	**Spliced exons/GC Rich.** States 21-23 all had relatively high frequency for H4K20me1, H2BK5me1, and H3K79me1 relative to most other modifications in the state. These states enriched for regions of the genome that are GC-rich and contained spliced exons		**Spliced exons/GC Rich; open chromatin, TF binding; candidate enhancer.** State 21 relative to States 22 and 23 had higher frequency for H3K4me1, H3K9me1, and acetylations and showed greater enrichments for open chromatin and transcription factor binding. Relative to State 20 it had higher frequency for H2BK5me1 and H4K20me1, and had greater enrichments for GC-rich areas and spliced exons.
22			**Spliced exons/GC Rich.** State 22 relative to State 21 had a lower frequency H3K4me1 and various acetylations, and was less likely to contain open chromatin and detected transcription factor binding. Relative to State 23 this state was on average higher expressed and less likely to contain Alu repeat elements.
23			**Spliced exons/GC Rich; Alu repeats.** State 23 relative to States 21 and 22 had lower frequency of detection of all marks, and had lower average expression and was more likely to contain Alu repeat elements.
24	**Transcribed 5' Distal States.** States 24-26 all share H3K36me3, H3K27me1, and H2BK5me1 as the three most frequent marks in that order and frequently transition with each other. These three states are found more often in genic locations that are distal to the 5' end of the gene.		**Transcribed 5' distal; exons.** State 24 relative to states 25 and 26 had the highest absolute frequency for marks other than H3K36me3 such as H3K27me1 and H2BK5me1. Of these three states this state had the highest average expression and the least relative bias away from 5' ends of genes.
25			**Transcribed Further 5' distal; exons.** State 25 relative to State 24 was found at locations more distal to the 5' end of a gene and relative to State 26 was more likely to be found overlapping exons.
26			**Transcribed 5' distal; Alu repeats.** State 26 relative to states 24 and 25 was less likely to correspond to exons relative to introns, and more likely to overlap Alu repeat elements.
27	**End of Transcription; exons; high expression.** State 27 had the highest frequency for the detection of H3K36me3, H4K20me1, and PolII marks, but low frequency for a number of other marks found with these marks in other states. This state had the highest average expression of the non-promoter associated states, and of any state the greatest enrichment for spliced exons, transcription end sites, and the 3'UTR region of genes.		
28	**ZNF Genes; KAP1 repressed state.** State 28 had the highest frequency for H4K20me3, H3K36me3, and H3K9me3. This state was associated with ZNF genes and KAP1 binding.		

Active intergenic states:

29	**Candidate strong distal enhancer states.** States 29 and 30 both had relatively high frequency for H3K4me1 and various acetylations compared to other marks in the state. These states showed high enrichments for open chromatin and transcription factor binding enrichments except in some cases for the repressive NRSF. Genes located downstream from states 29 and 30 had among the highest average expression levels of the active intergenic states. Combined this suggests many locations within this state likely represent active enhancers.	**Candidate strong distal enhancer; higher open chromatin; higher target expression.** State 29 relative to State 30 had higher frequency for all acetylations and H3K4me1. Compared with State 30 this state had a higher frequency for detecting acetylations as well as higher open chromatin enrichments, higher enrichments in most transcription factor binding experiments, and a higher GC content.
30		**Candidate strong distal enhancer; high open chromatin; higher target expression**. State 30 had highest frequency relative to other marks in the state for detecting H3K4me1 and specific acetylations, such as H2BK5ac, H3K27ac, and H2BK120ac, though this frequency was lower than in State 29. Compared with State 29 this state had lower frequency for detecting acetylations it showed lower open chromatin, lower enrichments in most transcription factor binding experiments, and a lower GC content.
31	**Intergenic H2AZ with open chromatin/TF binding; Candidate distal enhancer.** State 31 had H2AZ, H4K8ac, H4K5ac, and H3K4me1 as the most frequent marks. This state had higher frequency for H2AZ and H4K8ac than both States 29 and 30. This state showed enrichment for open chromatin and transcription factor binding on the same order of State 29 and State 30, however downstream genes of this state did not display as high an average expression level. Compared to States 29 and 30 this state was even more likely to fall outside of annotated genes. This state enriched for being proximal to the CTCF island state (State 39), but a large portion was also found proximal to other states.	
32	**Candidate weaker distal enhancer.** State 32 had a similar H3K4me1 frequency as found in States 30 and 31, but differed in the relative frequency of specific acetylations with H4K91ac and H2BK20ac being the most frequent in this state. While this state enriched for open chromatin and in some transcription factor binding experiments, the enrichment was lower than what was seen in States 29-31 and 33, and there were also fewer transcription factors enriched for conserved motifs.	
33	**Candidate distal enhancer.** State 33 had a relatively high frequency of H3K4me1 compared to the other marks in this state, and lower acetylation levels than States 29-32. This state showed enrichment for open chromatin lower than in states 29-31, but higher than in 32.	
34	**Proximal to active enhancers; Alu repeats.** State 34 had low frequency for a number of methylations and acetylations. This state was enriched proximal to the intergenic candidate enhancer states. Genes downstream from this state had higher average expression levels than States 31-33. This state had lower enrichments for detected open chromatin, transcription factor binding as compared to States 29-33, while having greater enrichments for Alu repetitive elements.	
35	**Active intergenic regions not enhancer specific.** State 35 had the highest frequency for acetylations such as H2AK5ac, H4K91ac, H3K4ac, and H2BK20ac relative to other marks in the state. These marks were also among the most frequent in State 32, but this state had lower absolute frequencies for these acetylations and also had limited detection of H3K4me1. The enrichment for open chromatin and overall for transcription factor binding was lower than found in States 29-33. In comparison to state 34 this state had lower enrichments for neighboring the stronger intergenic candidate enhancer states. Genes downstream of this state also had a lower average expression than 34.	
36	**Active intergenic further from enhancers; Alu repeats.** State 36 as with state 35 had the highest relatively frequency for the acetylation marks H2AK5ac, H4K91ac, H3K4ac, but at lower absolute levels. In comparison to State 35, this state had even lower frequency for detected open chromatin and transcription factor binding, and higher levels of Alu repetitive elements. This state had lower enrichments for neighboring intergenic candidate enhancers than States 34 or 35.	
37	**Non-repressive intergenic domains; Alu repeats.** State 37 had relatively low level of detection of all marks and represented a large 11.0% of the genome. While the absolute mark frequency was also low in the three larger states: 40, 41, and 43 the transitions transition probabilities for this state was distinct, allowing small absolute differences in mark frequency to become significant over broad domains. Functionally genes near this state had higher average expression than states 40-45, but lower than States 29-36. This state had lower enrichments for being proximal to the intergenic candidate enhancer states than States 29-36, but higher than States 40-45.	
38	**H2AZ specific state.** State 38 had the highest frequency for H2AZ relative to other marks in this state. Relative to State 31 which also had H2AZ as the most frequently detected mark, the acetylations in this state were lower and there was lower enrichment for open chromatin and transcription factor binding. This state was found enriched near CTCF islands states, but a large portion was also found next to other states.	
39	**CTCF Island; Candidate Insulator.** State 39 was most frequently associated with CTCF and to a lesser extent H2AZ. CTCF is an insulator binding protein and while also found in promoter enriched states, the vast majority of CTCF in this state was found distal to promoters.	

Repressed states:

40	**Unmappable.** State 40 had all emission frequencies <0.0010 and a very high self-transition parameter. This state also showed a severe depletion in an IgG control experiment. This state corresponds to large segments of the genome which cannot be interrogated using the ChIP-seq technology, either because the sequence is not available or is duplicated across the genome.
41	**Heterochromatin; Nuclear Lamina; Most A/T rich.** State 41 represented large domains covering 23.3% of the genome with H3K9me3 the most frequently detected mark, but at a low absolute frequency. This was the most A/T rich state, showed strong depletion for promoter regions and genes, genes within this state were repressed, and with State 42 showed the greatest correspondence with the nuclear lamina and the darkest staining chromosomal bands. This state was more gene depleted compared to States 43-45.
42	**Heterochromatin; Nuclear Lamina; ERVL repeats.** State 42 was often flanked by State 41, and as with State 41 was also associated with gene depletion and repression. State 42 relative to 41 was more likely to be marked by H3K27me2, H3K9me2, H4K20me3, and H3R2me2. The enrichments for repetitive elements differed between this state and State 41. This state showed the strongest enrichment for ERVL repetitive elements of any state including State 41, while being less likely to mark an L1 repetitive element than State 41.
43	**Heterochromatin.** State 43 represented large domains covering 22.3% of the genome with H3K27me3 as the most frequently detected mark, but at a low absolute frequency. The genes in this state were on average repressed, but relative to State 41 had less depletion for promoter regions and genes.
44	**Heterochromatin; Nuclear Lamina; Less exon depleted.** State 44 was often flanked by State 43, and as with State 43 was associated with gene repression and less gene depletion compared to states 41 and 42. Compared to State 43 this state was more likely to have detected H3K27me3 with higher frequency and also for other marks such as H3K27me2, H3R2me1, H3K36me1, and/or H3R2me2. State 44 was more likely to contain ERVL repetitive elements and less likely to contain L1 elements than state 43. State 44 also more likely than State 43 to be found near spliced exons.
45	**Specific Repression.** State 45 represented regions of relative higher frequency of detecting H3K27me3. This state contrasts with State 44 both in its more frequent self-transitions as compared to transitions to State 43 as well as the lower frequency for the other marks such as H3K27me2. This state showed enrichment for TSS of genes, though it was less specific to promoter regions as compared to State 4. Genes in this state were repressed genes and the TSS of genes most likely falling in this state enriched for embryonic development related genes.

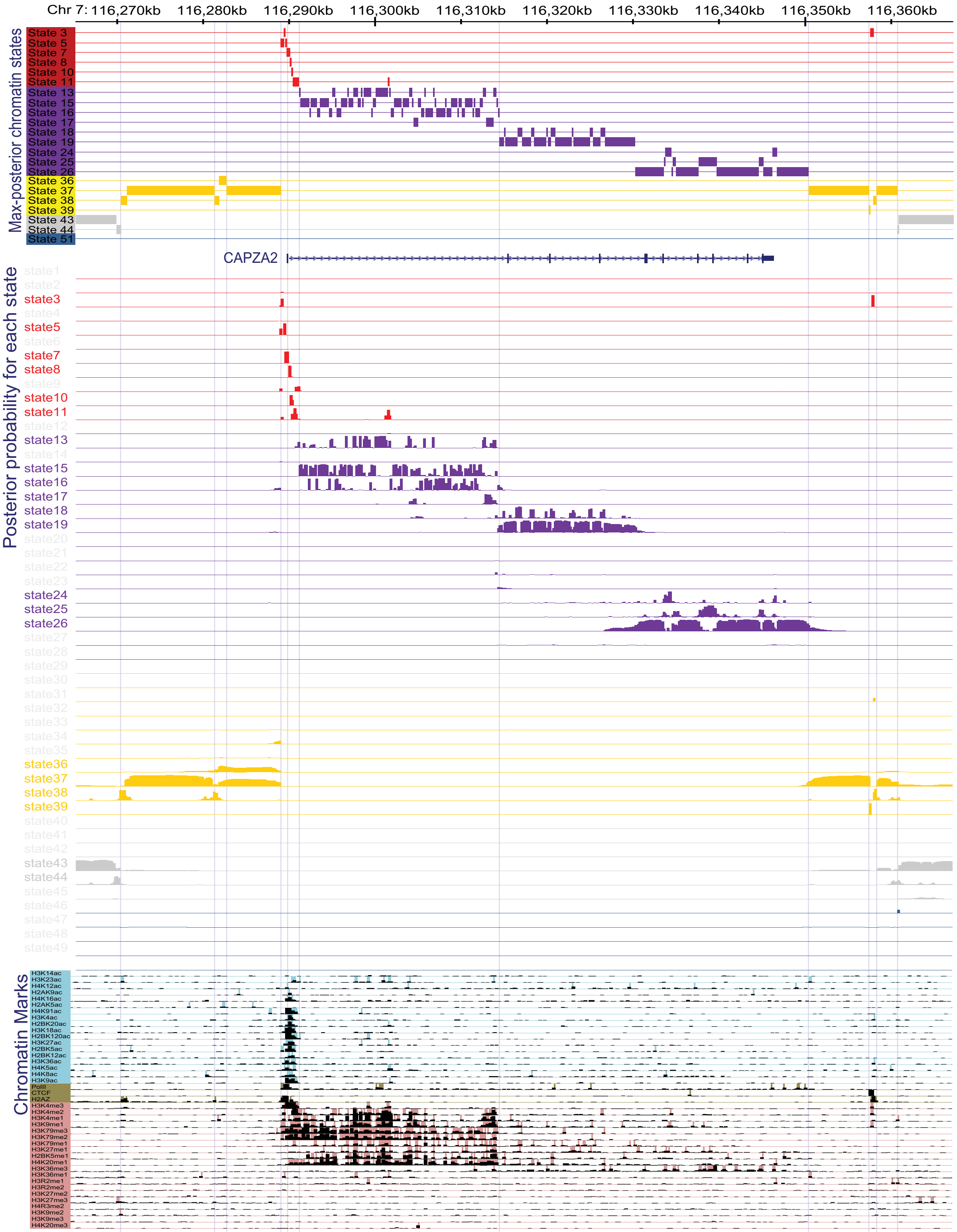
Repetitive states:

46	**Simple repeats (CA)n, (TG)n.** State 46 had the highest frequency for H2BK5me1 and H3R2me1. This state showed strong enrichments for simple repeat elements particularly (CA)n, (TG)n, and (CATG)n repeats.		
47	**L1/LTR Repeats.** State 47 had the highest frequency for H3K9me3. These locations of H3K9me3 are found within or bordering domains of H3K27me3 as evidenced by states 43 and 45 being the majority of non-self transitions, which also differentiates it from the broader H3K9me3 associated state 41. This state shows a specific enrichment for LINE/LTR repeats.		
48	**Satellite Repeats.** States (48-51) were all marked by H4K20me3 and H3K9me3, but did not have the highly specific ZNF signature found in State 28. These four states all strongly enriched for Satellite repeats.	**Satellite Repeat.** Out of States 48-51, only in State 48 was there low frequency of all other marks besides H3K9me3 and H4K20me3 and no substantial bias detected in an IgG control experiment.	
49		**Satellite Repeat; mapping bias.** States 49 -51 showed several additional marks detected, and all enriched based on IgG control. Locations within these states likely reflect sequences in the genome which are non-unique, but are unique in the reference genome.	**Satellite Repeat; moderate mapping bias.** Compared to states 50 and 51 this state had fewer additional marks detected and a lower relative bias in the IgG control.
50			**Satellite Repeat; high mapping bias.** Compared to states 49 and 51 this state had intermediate frequencies for additional marks detected and enrichment in an IgG control.
51			**Satellite Repeat/rRNA; extreme mapping bias.** Compared to states 49 and 50 this state had the highest frequencies for additional marks detected and enrichment in an IgG control. This state in addition to enriching for Satellite repeats also had a notable enrichment for rRNA.

Supplementary Table 1: Chromatin state characterization. Defining chromatin marks and biological interpretation summary for each chromatin state, summarizing key findings for reference. When groups of states share marks or candidate biological functions, these are grouped into left-most columns, and their specific differences are discussed in right-most columns.

Mark	Number of Tags (in Millions)	Threshold	% 200 bp Intervals Called '1'	% Tags in Called '1' Interval
H3K4me3	16.85	7	1.28	47.8
H3K36me3	13.57	6	2.34	26.0
H3K4me1	11.32	6	2.59	49.4
H4K20me1	11.02	6	2.80	57.5
H3K27me1	10.05	5	1.53	14.7
H3K9me2	9.78	5	0.60	5.9
H3R2me1	9.56	5	0.67	6.8
H3K9me1	9.31	5	2.56	34.4
H3K27me2	9.07	5	0.65	6.7
H3K27me3	8.97	5	0.86	8.7
H2BK5me1	8.94	5	2.45	41.2
H3K36me1	8.08	5	0.31	3.9
H2AZ	7.54	5	1.37	31.2
H4R3me2	7.36	5	0.18	4.2
H4K16ac	7.06	5	0.42	6.1
H3R2me2	6.52	4	0.67	8.7
H3K9me3	6.35	4	1.29	18.4
H3K79me3	5.93	4	2.39	57.8
H4K20me3	5.72	4	0.80	38.6
H3K4me2	5.45	4	1.74	35.7
H3K79me1	5.14	4	2.27	37.2
H3K79me2	4.71	4	2.11	57.4
H3K36ac	4.37	4	0.74	19.3
H4K8ac	4.28	4	0.90	19.9
H3K18ac	4.25	4	1.36	37.9
PolII	4.15	4	0.96	29.2
H4K5ac	4.12	4	0.99	23.9
H2BK20ac	4.08	4	1.27	33.7
H3K9ac	3.95	4	0.51	19.7
H3K14ac	3.80	4	0.12	2.5
H4K12ac	3.68	4	0.36	7.1
H2BK12ac	3.62	4	0.63	17.8
H3K4ac	3.55	3	1.54	34.1
H2BK120ac	3.44	3	1.63	46.3
H2AK5ac	3.44	3	1.30	21.9
H3K27ac	3.43	3	1.51	51.6
H2BK5ac	3.33	3	1.36	51.8
H4K91ac	3.19	3	1.79	53.0
CTCF	2.95	3	0.68	35.5
H3K23ac	2.53	3	0.51	10.8
H2AK9ac	2.07	3	0.41	11.2

Supplementary Table 2: Sequence tag thresholds for binarization of each chromatin mark input signal. This table indicates for each chromatin mark the number of sequence tags available for that the mark in millions, the threshold number of tags for which that mark was called present (e.g. 7 or more tags for H3K4me3 results in ‘1’), the percentage of 200bp intervals called ‘1’ in the input for that mark, and the percentage of tags which fell in a bin called ‘1’. These thresholds were selected using a Poisson background model, such that a bin was called ‘1’ for a mark if the probability of more sequence tags mapping to the interval was less than $<10^{-4}$.



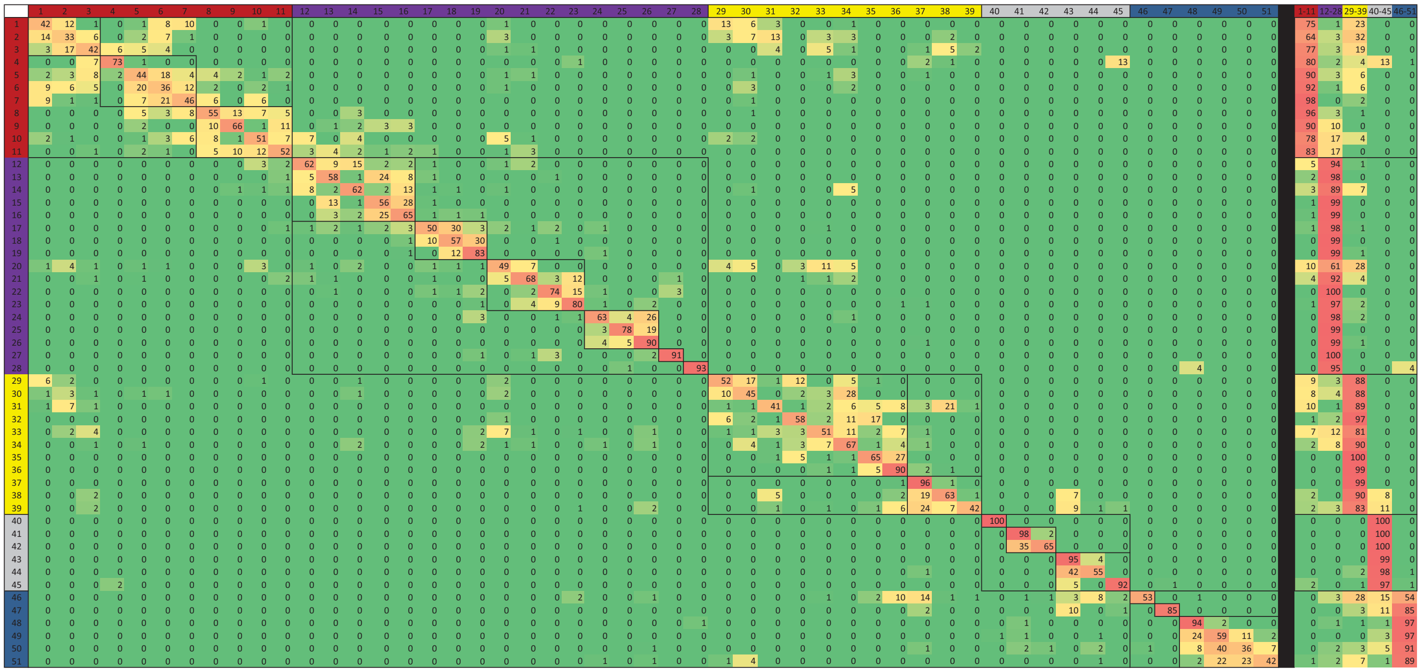
Supplementary Figure 1: Example of posterior probability distributions for all 51 chromatin states. Shown for the same region as Figure 1. Top panel and bottom panels reproduce maximum-posterior-probability chromatin states and input chromatin marks from Figure 1 for comparison. Middle panel shows actual posterior probability values between 0 and 1 for each state at each genomic position (note that these probabilities sum to 1 across states). Figure illustrates that state assignments are largely unambiguous, with the maximum-probability state containing most of the probability density, and nearly all other states containing no posterior probability.

state	H3K14ac	H3K23ac	H4K12ac	H2AK9ac	H4K16ac	H2AK5ac	H4K91ac	H3K4ac	H2BK20ac	H3K18ac	H2BK120ac	H3K27ac	H2BK5ac	H2BK12ac	H3K36ac	H4K5ac	H4K8ac	H3K9ac	PolII	CTCF	H2AZ	H3K4me3	H3K4me2	H3K4me1	H3K9me1	H3K79me3	H3K79me2	H3K79me1	H3K27me1	H2BK5me1	H4K20me1	H3K36me3	H3K36me1	H3R2me1	H3R2me2	H3K27me2	H3K27me3	H4R3me2	H3K9me2	H3K9me3	H4K20me3
1	3.8	23.6	24.2	18.0	37.7	25.5	95.2	94.8	94.3	99.2	99.6	99.7	98.9	79.1	88.6	93.6	86.9	83.6	51.6	15.7	87.5	94.2	93.8	64.2	87.0	3.8	3.3	12.0	19.4	11.6	3.8	0.5	2.6	1.9	2.1	0.2	0.1	0.2	0.5	0.1	1.8
2	2.5	17.5	9.2	3.2	5.9	6.3	44.6	44.4	47.0	73.2	74.1	85.9	71.2	22.1	33.5	61.9	63.3	35.4	18.1	10.9	91.2	86.7	90.4	66.9	78.3	2.4	2.2	7.9	17.6	8.7	2.3	0.6	2.2	1.7	1.5	0.4	0.5	0.2	0.4	0.1	1.4
3	0.5	5.8	1.8	1.0	0.9	1.2	12.3	9.5	8.8	22.6	21.3	22.8	12.1	2.2	4.2	8.4	12.8	7.1	11.2	16.3	77.1	93.9	80.3	45.6	74.2	1.5	1.3	4.6	4.3	7.0	8.8	0.2	1.4	2.1	1.5	0.2	2.1	0.1	0.1	0.1	1.2
4	0.1	0.8	0.1	0.4	0.3	0.2	1.5	0.9	0.7	2.1	2.1	0.5	0.2	0.1	0.1	0.1	0.3	0.3	1.9	6.2	19.0	77.9	20.8	21.1	26.4	0.3	0.0	0.1	0.0	1.3	14.9	0.1	0.2	1.4	0.9	0.0	10.2	0.1	0.1	0.4	1.3
5	0.0	0.2	0.8	1.3	2.0	0.4	26.6	12.6	6.7	15.8	23.1	26.8	24.3	1.5	4.3	0.9	1.4	7.7	53.5	20.6	21.8	87.0	11.2	2.7	15.7	6.3	3.8	2.5	0.0	5.5	14.2	0.0	0.3	0.2	0.6	0.0	0.0	0.0	0.1	0.0	1.5
6	0.1	1.8	3.6	6.9	6.1	1.9	74.5	63.5	53.0	75.7	84.3	89.4	86.7	20.8	41.5	20.5	21.6	62.7	69.2	25.5	61.2	98.3	37.4	7.1	40.3	5.3	2.7	5.6	0.6	6.0	11.6	0.0	0.5	0.4	0.9	0.0	0.0	0.0	0.0	0.1	1.6
7	1.2	8.7	20.6	43.0	53.7	9.8	98.7	98.6	95.7	99.4	99.9	99.9	99.9	76.5	93.3	81.8	76.6	99.2	88.0	26.9	77.1	99.7	38.3	2.2	37.9	32.1	24.9	14.0	2.6	6.5	16.2	0.1	1.0	0.5	1.2	0.0	0.0	0.0	0.1	0.0	1.9
8	1.2	12.7	5.2	11.9	5.6	6.8	56.9	56.1	37.5	52.4	69.8	89.1	85.5	21.3	24.8	16.7	10.3	60.8	62.1	12.0	31.4	96.7	51.7	14.9	45.3	86.3	80.1	23.8	1.7	6.7	42.2	4.5	0.4	1.1	0.9	0.0	0.0	0.0	0.0	0.0	0.5
9	0.5	7.2	3.0	1.1	0.5	2.1	4.0	7.4	2.4	2.5	11.6	35.3	28.0	2.7	2.8	2.0	1.8	8.6	34.7	4.2	4.5	79.2	41.5	23.8	36.1	86.0	82.6	12.0	1.9	6.7	43.5	7.4	0.2	0.6	0.7	0.0	0.0	0.0	0.0	0.2	0.4
10	4.1	24.8	13.1	17.5	24.4	37.0	90.4	88.6	82.0	89.8	95.5	97.0	95.1	54.0	56.4	67.2	45.7	55.7	46.6	10.2	40.8	84.6	92.3	91.4	92.8	67.1	67.2	63.4	29.2	53.8	65.2	4.5	6.8	5.7	3.4	0.1	0.0	0.3	0.1	0.0	1.0
11	1.6	21.0	3.9	3.8	3.2	8.4	28.0	26.1	14.5	22.6	37.6	56.8	47.4	6.0	6.4	13.5	8.8	18.4	30.9	6.9	20.1	92.4	93.5	94.3	94.5	73.8	74.2	55.1	24.0	57.2	79.9	8.9	6.6	4.4	2.5	0.1	0.0	0.2	0.1	0.1	0.7
12	3.6	17.0	8.9	2.2	14.1	34.9	60.3	51.0	38.8	35.6	56.6	53.3	55.3	11.9	11.5	30.0	15.4	1.7	18.0	2.7	0.7	5.4	58.2	96.0	77.8	87.4	87.0	76.3	41.6	79.8	82.2	13.4	3.1	6.4	3.6	0.3	0.0	0.7	0.2	0.1	0.4
13	1.2	10.8	3.6	0.7	2.5	7.4	9.1	6.5	2.6	2.5	6.5	7.7	5.5	0.5	1.0	5.5	3.3	0.3	10.2	1.5	0.0	2.4	56.2	83.7	82.9	92.7	92.6	64.4	38.2	80.0	89.8	12.0	2.4	3.3	2.1	0.3	0.0	0.4	0.2	0.1	0.3
14	0.7	5.3	7.9	1.0	2.4	18.0	19.8	20.7	14.6	7.9	24.5	20.1	21.8	6.7	6.6	11.3	8.6	0.3	6.9	1.7	2.1	1.3	8.0	33.0	16.3	61.8	62.1	37.9	9.7	14.3	18.3	9.7	0.2	1.7	1.2	0.2	0.0	0.1	0.3	0.4	1.0
15	0.2	1.9	2.9	0.3	0.3	1.5	0.8	1.3	0.3	0.2	1.2	1.5	1.2	0.2	0.4	0.7	1.2	0.1	5.0	0.7	0.0	0.2	11.0	17.3	29.3	84.7	82.2	33.1	8.0	26.4	56.2	5.2	0.2	0.7	0.9	0.1	0.0	0.1	0.1	0.6	0.2
16	0.0	0.3	0.4	0.1	0.1	0.5	0.4	0.6	0.2	0.1	0.5	0.4	0.3	0.1	0.2	0.2	0.3	0.0	0.6	0.3	0.1	0.1	1.2	2.8	3.9	29.0	25.2	8.5	0.7	1.3	7.9	1.2	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.4	0.1
17	1.2	9.8	2.8	0.9	2.4	7.8	6.8	6.1	2.3	4.0	3.5	8.4	3.5	0.3	1.0	9.3	6.6	0.5	3.5	1.1	0.4	1.0	52.3	68.9	83.7	22.7	23.5	61.2	48.3	64.7	57.3	21.8	2.8	4.9	2.0	1.0	0.1	0.5	0.5	0.1	0.5
18	0.3	2.6	2.1	0.4	0.5	2.4	1.4	1.6	0.5	0.6	0.9	1.6	0.7	0.2	0.5	1.9	1.9	0.1	1.6	0.7	0.1	0.1	10.4	9.7	29.1	15.0	13.5	34.0	14.9	19.9	21.4	10.6	0.5	1.2	0.9	0.5	0.1	0.1	0.3	0.2	0.2
19	0.1	0.3	0.5	0.2	0.1	0.5	0.1	0.3	0.0	0.0	0.1	0.2	0.1	0.1	0.1	0.1	0.3	0.0	0.4	0.3	0.0	0.0	0.5	0.2	2.0	4.4	2.9	7.3	1.0	1.0	2.3	1.4	0.0	0.2	0.3	0.1	0.0	0.0	0.1	0.4	0.1
20	2.5	10.7	5.4	3.1	9.9	26.2	58.2	48.8	41.7	49.3	54.8	57.1	51.5	13.0	14.1	31.6	21.7	4.0	14.5	6.7	15.6	20.9	56.8	97.1	70.5	5.4	5.6	33.8	31.1	52.6	38.4	7.4	3.4	6.9	3.8	0.5	0.1	0.7	0.3	0.0	1.0
21	0.2	0.8	2.0	1.3	7.2	11.3	32.3	15.4	11.5	5.7	18.6	8.4	12.4	2.1	1.2	3.2	2.3	0.5	15.1	6.5	0.6	4.7	17.0	68.7	33.3	8.7	6.4	37.2	9.6	65.4	87.7	9.2	1.3	7.1	5.3	0.1	0.2	1.4	0.0	0.0	0.8
22	0.1	0.1	1.1	0.6	6.2	2.4	7.8	1.8	0.6	0.1	1.5	0.8	1.5	0.1	0.1	0.7	0.7	0.1	8.5	1.0	0.0	0.0	5.3	8.4	14.6	15.5	9.1	50.0	9.6	77.5	94.1	22.9	0.5	5.6	4.6	0.1	0.0	1.4	0.0	0.1	0.7
23	0.0	0.1	0.1	0.4	2.0	1.6	4.9	1.2	0.5	0.2	0.9	0.2	0.3	0.0	0.1	0.1	0.1	0.0	1.4	1.1	0.0	0.0	0.5	2.6	1.4	0.5	0.1	5.4	1.3	19.4	36.8	2.5	0.1	1.4	1.5	0.1	0.2	0.3	0.0	0.0	0.2
24	0.3	1.8	2.1	0.9	3.2	3.8	4.0	2.3	0.9	0.6	1.1	3.6	1.9	0.1	0.4	3.7	3.9	0.3	2.2	1.0	0.1	0.1	6.0	4.5	17.2	1.3	0.2	15.6	29.8	29.3	7.1	49.5	1.3	4.7	2.2	1.0	0.0	0.6	0.3	0.1	0.3
25	0.1	0.3	0.8	0.5	0.5	0.6	0.3	0.3	0.1	0.0	0.1	0.5	0.3	0.1	0.1	0.4	0.7	0.1	0.8	0.4	0.0	0.1	0.4	0.1	1.4	0.8	0.1	2.0	8.8	2.8	0.4	60.1	0.4	1.1	0.9	0.6	0.1	0.2	0.3	1.7	0.3
26	0.1	0.2	0.6	0.2	0.2	0.2	0.1	0.2	0.0	0.0	0.1	0.3	0.2	0.0	0.1	0.2	0.4	0.0	0.6	0.3	0.0	0.0	0.3	0.1	0.8	0.2	0.0	0.8	2.3	0.9	0.2	4.2	0.1	0.2	0.4	0.1	0.0	0.0	0.1	0.1	0.0
27	0.0	0.5	4.4	0.4	1.3	1.2	1.3	0.7	0.3	0.1	0.7	2.1	2.4	0.1	0.1	1.6	2.7	0.1	21.7	1.4	0.0	0.0	1.1	1.1	3.5	4.6	1.2	9.9	3.0	7.1	31.7	34.0	0.2	0.7	1.1	0.0	0.0	0.1	0.0	0.3	0.1
28	0.0	0.0	0.2	0.3	0.0	0.4	0.1	0.2	0.2	0.0	0.3	0.1	0.1	0.1	0.0	0.1	0.1	0.0	1.3	0.3	0.0	3.3	0.2	0.5	0.5	13.8	3.4	5.7	1.3	3.0	1.5	68.8	0.1	2.7	2.7	0.3	0.2	0.8	0.4	43.0	74.9
29	4.6	8.4	11.1	6.6	20.4	54.7	88.5	88.1	89.6	86.6	95.3	86.3	86.9	68.1	60.2	67.6	42.6	10.4	13.6	3.8	24.2	7.6	24.7	84.4	25.8	4.9	5.6	14.9	17.6	21.7	5.0	2.9	0.9	4.8	3.1	0.4	0.1	0.4	0.8	0.2	4.4
30	1.2	3.6	8.4	2.4	2.6	13.5	24.9	34.5	34.4	24.6	52.1	60.1	64.6	27.4	23.8	20.7	16.7	3.2	9.1	2.9	12.0	6.9	8.8	35.8	6.5	2.8	2.9	4.4	3.7	3.0	1.0	2.8	0.1	0.9	1.0	0.0	0.1	0.1	0.4	0.6	3.4
31	1.7	7.6	4.7	1.7	2.4	6.8	17.7	18.4	21.5	37.9	31.6	35.4	20.1	9.3	13.0	48.3	57.7	3.3	1.2	5.9	69.0	10.5	16.1	41.8	11.0	0.6	0.5	1.2	10.7	2.2	0.0	1.1	1.4	2.0	1.3	1.1	0.5	0.2	0.8	0.2	1.1
32	1.4	1.6	1.0	1.9	8.1	50.1	72.4	57.2	60.9	42.5	57.6	12.1	14.8	23.6	19.5	16.0	5.4	0.3	1.6	1.7	3.6	0.1	2.1	41.4	5.0	1.8	1.7	12.2	10.9	17.1	4.1	4.1	0.9	6.5	3.0	1.3	0.3	0.6	0.5	0.3	3.8
33	1.2	4.6	0.9	0.9	1.2	9.8	12.9	10.3	7.0	12.7	10.5	9.8	5.7	1.3	2.1	6.3	4.1	0.4	2.6	4.3	8.6	1.5	16.3	77.1	23.5	0.7	0.5	5.1	10.3	10.8	3.9	1.8	1.1	2.9	1.6	0.7	0.3	0.2	0.4	0.1	0.4
34	0.2	0.7	2.0	0.5	0.7	5.4	7.2	6.2	4.9	1.2	8.9	5.7	6.2	2.9	2.2	2.3	2.3	0.1	1.5	0.9	1.0	0.2	0.6	6.3	1.0	1.9	1.7	5.2	2.6	2.2	0.6	1.4	0.0	0.5	0.6	0.0	0.0	0.0	0.1	0.2	1.0
35	0.6	1.1	0.1	1.1	1.3	32.1	29.9	20.0	19.6	16.4	11.8	1.1	0.5	3.2	5.4	4.0	1.5	0.2	0.3	1.0	2.0	0.1	0.3	2.6	0.6	0.4	0.2	1.9	2.6	2.2	0.5	1.7	0.9	4.2	2.0	2.6	0.8	0.3	0.5	0.2	0.6
36	0.2	0.3	0.0	0.5	0.2	5.7	2.5	2.1	1.0	1.6	0.8	0.1	0.1	0.3	0.6	0.3	0.1	0.1	0.1	0.3	0.3	0.0	0.1	0.3	0.1	0.2	0.0	0.2	0.4	0.4	0.1	0.4	0.3	1.3	0.8	0.8	0.5	0.1	0.3	0.3	0.1
37	0.1	0.2	0.0	0.2	0.1	0.7	0.1	0.2	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.2	0.2	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.3	0.1	0.5	0.4	0.2	0.2	0.1	0.1	0.2	0.1
38	0.4	2.5	0.6	0.8	0.2	1.9	1.0	1.9	1.4	3.8	2.2	2.4	0.9	0.8	1.4	4.8	9.2	0.4	0.1	0.7	33.0	1.3	1.5	5.9	0.9	0.2	0.1	0.2	1.7	0.3	0.0	0.3	0.8	1.0	0.9	1.0	1.1	0.2	0.6	0.2	0.3
39	0.1	0.4	0.5	0.6	0.1	1.2	0.7	0.8	1.1	0.9	1.0	0.4	0.2	0.3	0.4	1.2	2.0	0.1	4.1	86.2	12.3	1.4	1.6	5.0	1.2	0.2	0.0	0.3	0.6	0.7	0.7	0.2	0.8	2.5	1.2	0.2	0.8	0.1	0.0	0.2	0.6
40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
41	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.5	0.3	0.1	0.1	1.0	2.0	0.4
42	0.1	0.1	0.0	0.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.2	0.1	0.6	0.6	0.1	0.0	0.1	0.1	0.1	0.0	0.0	1.8	0.5	0.0	0.5	1.3	2.0	2.3	8.7	1.2	0.8	5.6	2.3	4.4
43	0.0	0.2	0.0	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.3	0.4	1.0	0.1	0.5	0.5	0.1
44	0.6	1.1	0.0	1.2	0.2	0.7	0.1	0.4	0.0	0.4	0.1	0.1	0.0	0.0	0.2	0.1	0.1	0.3	0.1	0.6	2.0	0.2	0.2	0.3	0.3	0.1	0.0	0.0	0.3	1.1	0.0	1.6	3.3	5.7	3.1	9.0	9.9	0.9	2.5	0.4	1.0
45	0.1	0.7	0.0	0.5	0.0	0.4	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.6	0.6	0.4	0.1	0.4	0.1	0.1	0.0	0.0	0.0	0.1	0.2	0.3	0.5	1.0	0.9	0.4	16.4	0.2	0.3	1.1	0.4
46	1.6	0.7	0.4	1.4	4.2	6.1	5.2	3.6	0.8	4.5	2.4	0.7	0.6	0.4	0.5	0.7	0.4	0.5	0.6	2.0	1.1	2.7	2.1	28.4	5.5	0.4	0.1	1.1	8.4	63.8	16.5	5.9	17.8	48.9	25.0	9.8	7.8	11.9	3.7	1.6	17.0
47	0.0	0.1	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.6	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	1.9	0.1	0.4	0.7	0.3	3.1	0.1	0.7	32.3	1.8
48	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.2	0.1	0.0	0.1	0.0	0.3	0.0	0.1	0.2	0.1	0.0	1.0	0.1	0.5	0.9	0.6	0.3	0.7	0.7	11.9	37.8
49	2.3	0.5	0.6	4.3	9.5	0.2	0.3	0.3	0.2	0.2	0.3	0.1	0.2	0.2	1.4	0.4	0.7	2.4	0.9	3.0	4.3	3.0	0.3	0.5	0.7	1.2	0.1	0.1	10.4	4.4	0.3	16.1	9.0	12.5	38.0	15.3	5.6	49.1	24.5	49.3	85.1
50	21.0	4.2	7.8	16.1	42.7	1.8	1.1	2.0	0.9	0.8	1.1	0.5	0.7	1.5	6.2	2.4	2.9	12.7	14.1	25.3	16.8	36.0	8.4	17.5	14.1	8.3	1.5	1.3	60.0	40.0	9.2	80.0	68.2	70.4	84.9	64.8	28.1	87.5	76.1	79.8	97.1
51	78.3	38.3	65.7	68.2	92.6	20.6	21.4	30.6	27.7	19.7	20.9	12.5	18.3	25.3	54.7	34.7	43.4	65.5	69.6	79.8	72.7	95.1	60.9	73.4	75.8	48.8	24.9	25.3	95.2	85.3	59.0	96.0	96.5	98.4	98.6	90.5	88.5	97.2	96.4	92.5	99.6

Supplementary Figure 2: Individual mark frequencies for each chromatin state (Emission Matrix). Each row corresponds to a state and each column corresponds to an input mark. An entry in a cell indicates the emission probability under the model that the mark will be detected in that state, also corresponding to the frequency with which the mark is observed in that state. Frequency values are shown as a percentage (multiplied by 100) to improve figure readability.

State	Most frequent chromatin marks									
	1st	%	2nd	%	3rd	%	4th	%	5th	
1	H3K27ac	100	H2BK120ac	100	H3K18ac	99	H2BK5ac	99	H4K91ac	95
2	H2AZ	91	H3K4me2	90	H3K4me3	87	H3K27ac	86	H3K9me1	78
3	H3K4me3	94	H3K4me2	80	H2AZ	77	H3K9me1	74	H3K4me1	46
4	H3K4me3	78	H3K9me1	26	H3K4me1	21	H3K4me2	21	H2AZ	19
5	H3K4me3	87	PolII	53	H3K27ac	27	H4K91ac	27	H2BK5ac	24
6	H3K4me3	98	H3K27ac	89	H2BK5ac	87	H2BK120ac	84	H3K18ac	76
7	H3K27ac	100	H2BK5ac	100	H2BK120ac	100	H3K4me3	100	H3K18ac	99
8	H3K4me3	97	H3K27ac	89	H3K79me3	86	H2BK5ac	86	H3K79me2	80
9	H3K79me3	86	H3K79me2	83	H3K4me3	79	H4K20me1	43	H3K4me2	41
10	H3K27ac	97	H2BK120ac	95	H2BK5ac	95	H3K9me1	93	H3K4me2	93
11	H3K9me1	95	H3K4me1	94	H3K4me2	94	H3K4me3	92	H4K20me1	80
12	H3K4me1	96	H3K79me3	87	H3K79me2	87	H4K20me1	82	H2BK5me1	80
13	H3K79me3	93	H3K79me2	93	H4K20me1	90	H3K4me1	84	H3K9me1	83
14	H3K79me2	62	H3K79me3	62	H3K79me1	38	H3K4me1	33	H2BK120ac	25
15	H3K79me3	85	H3K79me2	82	H4K20me1	56	H3K79me1	33	H3K9me1	29
16	H3K79me3	29	H3K79me2	25	H3K79me1	8	H4K20me1	8	H3K9me1	4
17	H3K9me1	84	H3K4me1	69	H2BK5me1	65	H3K79me1	61	H4K20me1	57
18	H3K79me1	34	H3K9me1	29	H4K20me1	21	H2BK5me1	20	H3K79me3	15
19	H3K79me1	7	H3K79me3	4	H3K79me2	3	H4K20me1	2	H3K9me1	2
20	H3K4me1	97	H3K9me1	71	H4K91ac	58	H3K27ac	57	H3K4me2	57
21	H4K20me1	88	H3K4me1	69	H2BK5me1	65	H3K79me1	37	H3K9me1	33
22	H4K20me1	94	H2BK5me1	77	H3K79me1	50	H3K36me3	23	H3K79me3	16
23	H4K20me1	37	H2BK5me1	19	H3K79me1	5	H4K91ac	5	H3K4me1	3
24	H3K36me3	50	H3K27me1	30	H2BK5me1	29	H3K9me1	17	H3K79me1	16
25	H3K36me3	60	H3K27me1	9	H2BK5me1	3	H3K79me1	2	H3K9me3	1
26	H3K36me3	4	H3K27me1	2	H2BK5me1	1	H3K9me1	1	H3K79me1	1
27	H3K36me3	34	H4K20me1	32	PolII	22	H3K79me1	10	H2BK5me1	7
28	H4K20me3	75	H3K36me3	69	H3K9me3	43	H3K79me3	14	H3K79me1	6
29	H2BK120ac	95	H2BK20ac	90	H4K91ac	88	H3K4ac	88	H2BK5ac	87
30	H2BK5ac	65	H3K27ac	60	H2BK120ac	52	H3K4me1	36	H3K4ac	34
31	H2AZ	69	H4K8ac	58	H4K5ac	48	H3K4me1	42	H3K18ac	38
32	H4K91ac	72	H2BK20ac	61	H2BK120ac	58	H3K4ac	57	H2AK5ac	50
33	H3K4me1	77	H3K9me1	23	H3K4me2	16	H4K91ac	13	H3K18ac	13
34	H2BK120ac	9	H4K91ac	7	H3K4me1	6	H3K4ac	6	H2BK5ac	6
35	H2AK5ac	32	H4K91ac	30	H3K4ac	20	H2BK20ac	20	H3K18ac	16
36	H2AK5ac	6	H4K91ac	3	H3K4ac	2	H3K18ac	2	H3R2me1	1
37	H2AK5ac	0.7	H3R2me1	0.5	H3R2me2	0.4	H3K36me3	0.3	H3K27me2	0.2
38	H2AZ	33	H4K8ac	9	H3K4me1	6	H4K5ac	5	H3K18ac	4
39	CTCF	86	H2AZ	12	H3K4me1	5	PolII	4	H3R2me1	3
40	all emissions < 0.0010									
41	H3K9me3	2	H3K9me2	1	H3R2me2	0.5	H4K20me3	0.4	H3K27me2	0.3
42	H3K27me2	9	H3K9me2	6	H4K20me3	4	H3K9me3	2	H3R2me2	2
43	H3K27me3	1	H3K9me2	0.5	H3K9me3	0.4	H3K27me2	0.3	H3R2me2	0.3
44	H3K27me3	10	H3K27me2	9	H3R2me1	6	H3K36me1	3	H3R2me2	3
45	H3K27me3	16	H3K9me3	1	H3R2me1	1	H3R2me2	1	H3K23ac	1
46	H2BK5me1	64	H3R2me1	49	H3K4me1	28	H3R2me2	25	H3K36me1	18
47	H3K9me3	32	H3K27me3	3	H3K36me3	2	H4K20me3	2	H3K9me2	1
48	H4K20me3	38	H3K9me3	12	H3K36me3	1	H3R2me2	1	H3K9me2	1
49	H4K20me3	85	H3K9me3	49	H4R3me2	49	H3R2me2	38	H3K9me2	25
50	H4K20me3	97	H4R3me2	88	H3R2me2	85	H3K36me3	80	H3K9me3	80
51	H4K20me3	100	H3R2me2	99	H3R2me1	98	H4R3me2	97	H3K36me1	97

Supplementary Figure 3: Top five most-frequently-detected chromatin marks for each state. Number in each cell indicates the frequency of the mark in that state (multiplied by 100 to improve readability). Marks are colored according to their similarity in chromatin state enrichments in order to visually reveal groups of states defined by similar mark combinations, and also differences between states within each group.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	1-11	12-28	29-39	40-45	46-51
1	42	12	1	0	1	8	10	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	13	6	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75	1	23	0	0
2	14	33	6	0	2	7	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	3	7	13	0	3	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	64	3	32	0	0
3	3	17	42	6	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	4	0	5	1	0	0	1	5	2	0	0	0	0	0	0	0	0	0	0	0	0	77	3	19	0	0
4	0	0	7	73	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	13	0	0	0	0	0	0	80	2	4	13	1
5	2	3	8	2	44	18	4	4	2	1	2	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	3	6	0	0
6	9	6	5	0	20	36	12	2	0	2	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92	1	6	0	0
7	9	1	1	0	7	21	46	6	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	2	0	0
8	0	0	0	0	5	3	8	55	13	7	5	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	3	1	0	0
9	0	0	0	0	2	0	0	10	66	1	11	0	1	2	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	10	0	0	0
10	2	1	0	0	1	3	6	8	1	51	7	7	0	4	0	0	0	0	0	5	1	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	17	4	0	0
11	0	0	1	0	2	1	0	5	10	12	52	3	4	2	1	2	1	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83	17	0	0	0
12	0	0	0	0	0	0	0	0	0	3	2	62	9	15	2	2	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	94	1	0	0
13	0	0	0	0	0	0	0	0	0	0	1	5	58	1	24	8	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	98	0	0	0
14	0	0	0	0	0	0	0	0	1	1	1	8	2	62	2	13	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	89	7	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	13	1	56	28	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	99	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	3	2	25	65	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	99	0	0	0
17	0	0	0	0	0	0	0	0	0	0	1	1	2	1	2	3	50	30	3	2	1	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	98	1	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10	57	30	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	12	83	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	1	0	0
20	1	4	1	0	1	1	0	0	0	3	0	1	0	2	0	0	1	1	1	49	7	0	0	0	0	0	0	0	4	5	0	3	11	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	61	28	0	0
21	0	0	1	0	1	0	0	0	0	0	2	1	1	0	0	0	1	0	0	5	68	3	12	0	0	0	1	0	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	92	4	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	2	0	2	74	15	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	9	80	1	0	2	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	97	2	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	1	63	4	26	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	2	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	78	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	5	90	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	1	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	0	0	0	2	91	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	95	0	0	4
29	6	2	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	52	17	1	12	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	3	88	0	0
30	1	3	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	10	45	0	2	3	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	4	88	0	0
31	1	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	41	1	2	6	5	8	3	21	1	0	0	0	0	0	0	0	0	0	0	0	0	10	1	89	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	6	2	1	58	2	11	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	97	0	0
33	0	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7	1	0	1	0	0	1	0	0	1	1	3	3	51	11	2	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	12	81	0	0
34	0	0	1	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	2	1	1	0	0	1	0	1	0	0	0	4	1	3	7	67	1	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	8	90	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	1	1	65	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	5	90	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	96	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0
38	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	2	19	63	1	0	0	0	7	0	0	0	0	0	0	0	0	2	0	90	8	0
39	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	1	0	1	0	1	6	24	7	42	0	0	0	9	1	1	0	0	0	0	0	0	2	3	83	11	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	2	0	0	0	0	0	0	0	0	0	0	0	0	100	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	65	0	0	0	0	0	0	0	0	0	0	0	0	100	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95	4	0	0	0	0	0	0	0	0	0	0	99	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	42	55	0	0	0	0	0	0	0	0	0	2	98	1
45	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	92	0	1	0	0	0	0	2	0	1	97	1
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	1	0	2	10	14	1	1	0	1	1	3	8	2	53	0	1	0	0	0	0	3	28	15	54
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	10	0	1	0	85	0	0	0	0	0	0	3	11	85
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	94	2	0	0	0	1	1	1	97
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	24	59	11	2	0	0	0	3	97
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	1	0	2	0	1	0	8	40	36	7	0	2	3	5	91
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	22	23	42	1	2	7	1	89

Supplementary Figure 4: State-to-state and state-to-group transition probabilities (Transition Matrix). To the left of the thick black line is the full set of transition probabilities between states multiplied by 100. The rows are the state the transition is from and the column the state the transitioning is to. The boxes show the groups and sub-groups of states described in the text. To the right of the thick black line is the total probability of transitioning from a state to any state in the indicated group of states (these are not separate model parameters, but rather a summarization of the table by summing the transition probabilities to each group of states).

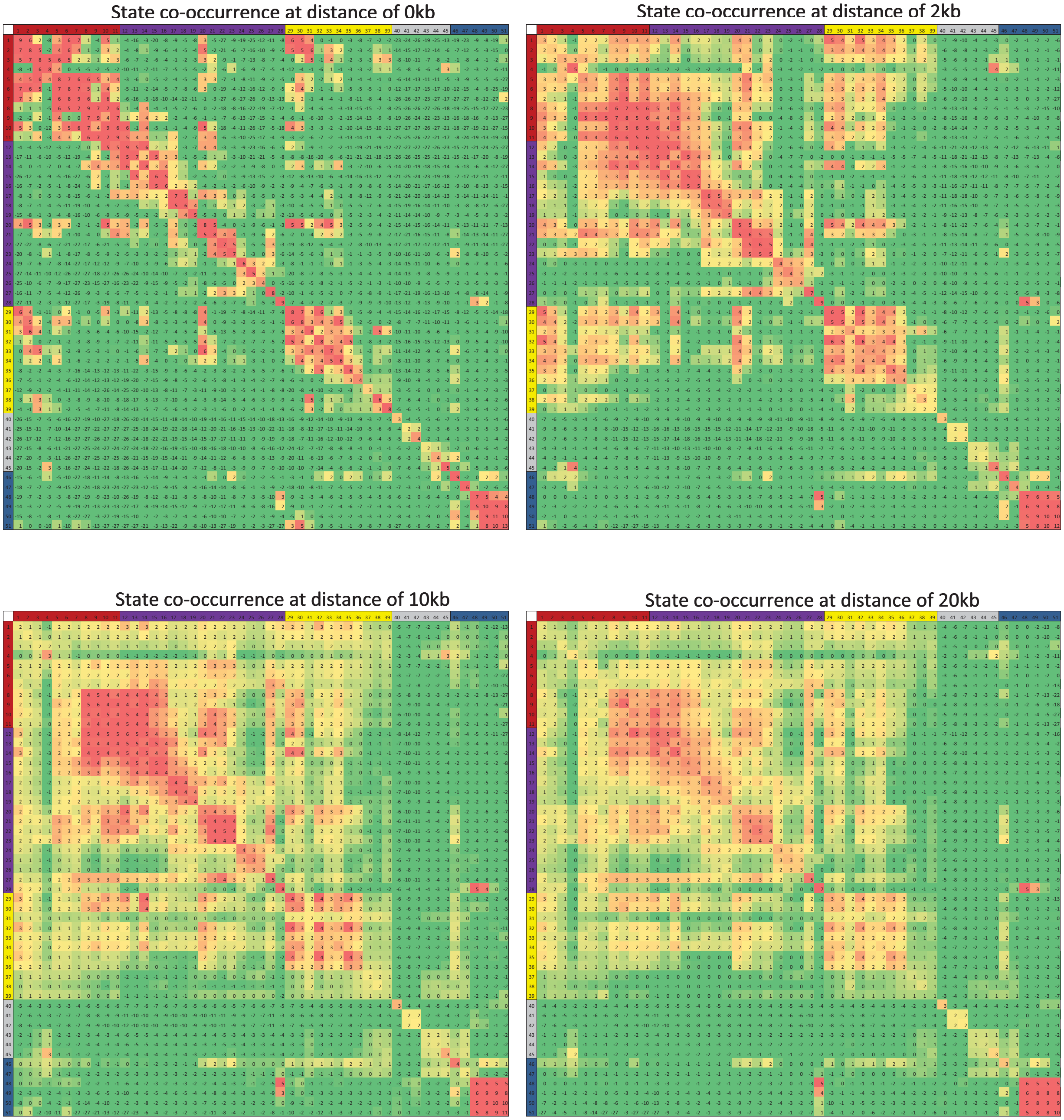
States Transitioning To

State	Non-self transitions >=0.015 in decreasing order										
1	29	2	7	6	30	31					
2	1	31	6	30	3	29	33	34	20	5	38
3	2	4	38	5	33	6	31	1	39		
4	45	3	37								
5	6	3	7	8	2	34	9	1	4	11	
6	5	7	1	2	3	30	8	10	34		
7	6	1	5	8	10						
8	9	7	10	5	11	6	14				
9	11	8	16	15	5	14					
10	8	12	11	7	20	14	6	29	1	30	
11	10	9	8	13	12	21	14	5	16		
12	14	13	10	21	16	15	11				
13	15	16	12								
14	16	12	34	15	13						
15	16	13									
16	15	13	14								
17	18	16	19	13	20	22	15				
18	19	17									
19	18										
20	33	21	30	34	29	2	10	32	14		
21	23	20	22	34	11						
22	23	27	19	21							
23	22	21	26								
24	26	25	19								
25	26	24									
26	25	24									
27	22	26									
28	48										
29	30	32	1	34	2	20					
30	34	29	2	33	20	32					
31	38	36	2	34	35	37	33				
32	35	34	29	33	30						
33	34	36	20	3	31	32	2	35	19		
34	33	30	36	32	14	19					
35	36	32									
36	35	37									
37											
38	37	43	31	3	36						
39	37	43	38	36	26	3					
40											
41	42										
42	41										
43	44										
44	43										
45	43	4									
46	37	36	44	43	23	35	45				
47	43	37									
48	49										
49	48	50	51								
50	49	48	51	44	41						
51	50	49	30	48							

Transition Probability

State	Self-Transition	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	Other
1	42	13	12	10	8	6	3						6
2	33	14	13	7	7	6	3	3	3	3	2	2	4
3	42	17	6	5	5	5	4	4	3	2			8
4	73	13	7	2									6
5	44	18	8	4	4	3	3	2	2	2	2		8
6	36	20	12	9	6	5	3	2	2	2			4
7	46	21	9	7	6	6							4
8	55	13	8	7	5	5	3	3					1
9	66	11	10	3	3	2	2						3
10	51	8	7	7	6	5	4	3	2	2	2		4
11	52	12	10	5	4	3	3	2	2	2			6
12	62	15	9	3	2	2	2	2					3
13	58	24	8	5									5
14	62	13	8	5	2	2							8
15	56	28	13										3
16	65	25	3	2									4
17	50	30	3	3	2	2	2	2					6
18	57	30	10										3
19	83	12											5
20	49	11	7	5	5	4	4	3	3	2			9
21	68	12	5	3	2	2							8
22	74	15	3	2	2								4
23	80	9	4	2									5
24	63	26	4	3									4
25	78	19	3										1
26	90	5	4										1
27	91	3	2										4
28	93	4											3
29	52	17	12	6	5	2	2						5
30	45	28	10	3	3	2	2						8
31	41	21	8	7	6	5	3	2					7
32	58	17	11	6	2	2							4
33	51	11	7	7	4	3	3	2	2	2			8
34	67	7	4	4	3	2	2						11
35	65	27	5										3
36	90	5	2										3
37	96												4
38	63	19	7	5	2	2							2
39	42	24	9	7	6	2	2						8
40	100												0
41	98	2											0
42	65	35											0
43	95	4											1
44	55	42											3
45	92	5	2										2
46	53	14	10	8	3	2	2	2					7
47	85	10	2										3
48	94	2											4
49	59	24	11	2									4
50	36	40	8	7	2	2							6
51	42	23	22	4	2								7

Supplementary Figure 5: High-probability transitions for each state. Left: For each state all non self-transitions that are greater than 0.015 are indicated in decreasing order. Right: Corresponding table of probabilities for self-transition (first column), and non-self transition probabilities for each transition indicated in the left table. Right-most column shows remaining transition probability for all other states (probability <0.015).



State co-occurrence at distance of 0kb

State co-occurrence at distance of 2kb

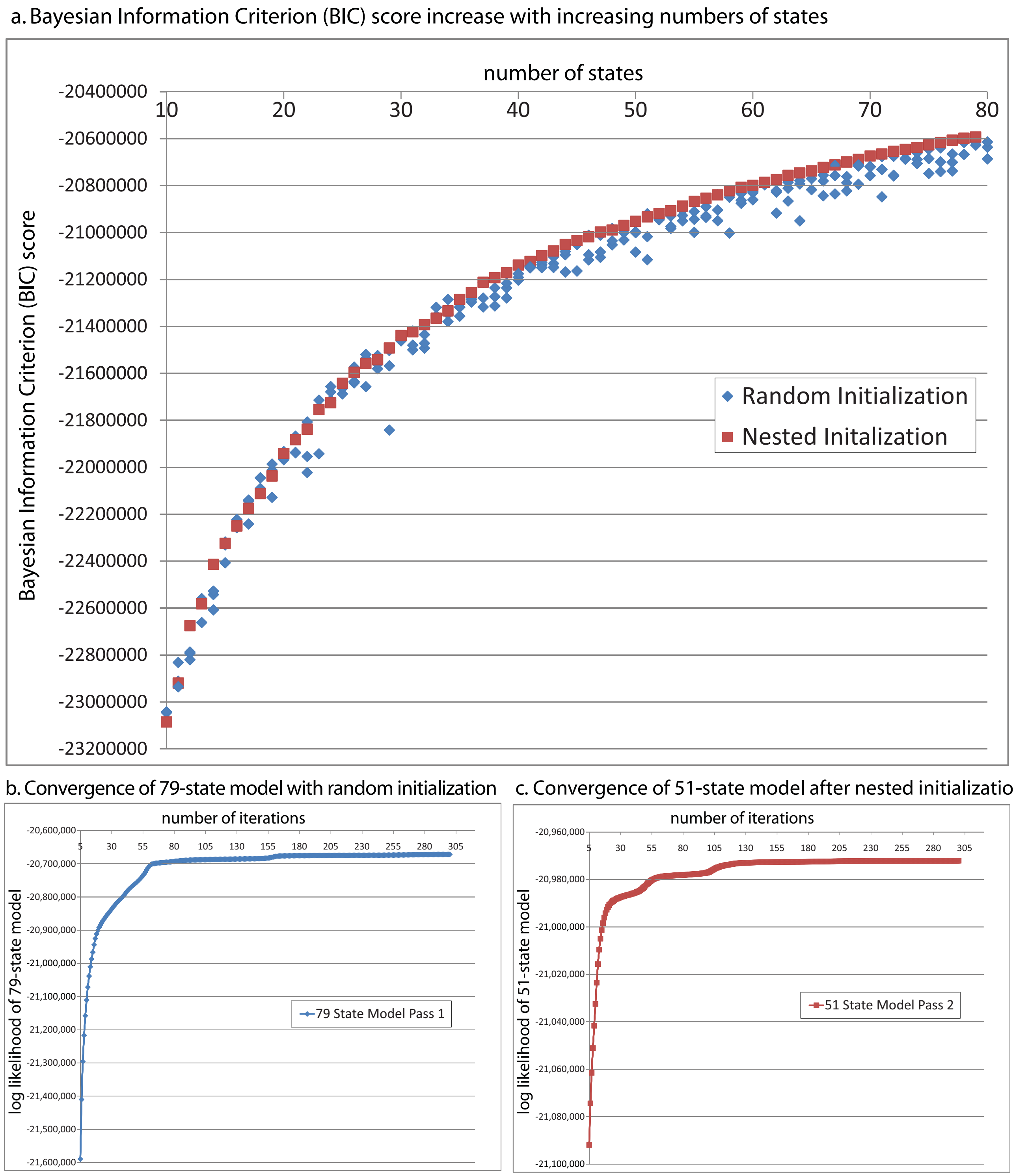
State co-occurrence at distance of 10kb

State co-occurrence at distance of 20kb

Supplementary Figure 6: Chromatin state co-occurrence enrichments at distances of 0kb, 2kb, 10kb, and 20kb. The grid shows the log base 2 fold enrichment for the frequency with which each pair of states co-occur at a fixed distance based on genomic coordinates relative to how often they would be expected to co-occur at the distance based on their size and if state occurrences were independent. Four distances are shown at a fixed gap of 0bp (top left) 2kb (top right), 10kb (bottom left), and 20kb (bottom right). These tables show several noteworthy longer-distance spatial relationships between states. For instance, large scale repressed states are depleted at each of the shown distances from most of the promoter, transcribed, and active intergenic states. Also, transcribed states enrich at all the distances shown relative to all promoter states except for the repressed promoter state (state 4). In computing the log fold enrichments, a pseudocountof 10^{-8} was added to the ratio before taking the log to smooth values close to 0.

state	ChromaSig clusters																Total coverage of each state (over all clusters)
	C7	C8	C1	C4	C2	C3	C11	C12	C13	C5	C10	C6	C9	C16	C14	C15	
1	53.6	3.7	41.6	1.2	0.0	0.0	0.0	0.0	0.0	14.8	33.1	1.6	0.3	0.0	0.0	0.0	74.5
2	**53.7**	12.0	27.7	1.0	0.0	0.0	0.0	0.0	0.0	35.2	34.3	7.7	0.5	0.1	0.0	0.0	69.0
3	43.8	47.3	24.6	0.6	0.4	0.1	0.0	0.0	0.0	29.0	7.2	6.2	2.7	0.2	0.1	0.0	58.7
4	8.8	**113.8**	3.2	0.1	0.7	0.0	0.0	0.0	0.0	6.6	0.3	1.4	10.7	17.8	1.2	0.0	34.7
5	28.5	7.0	46.3	1.9	1.2	0.1	0.0	0.0	0.0	18.9	2.2	0.4	1.3	0.0	0.0	0.0	56.7
6	38.6	4.0	58.3	0.8	0.4	0.1	0.0	0.0	0.0	21.2	2.5	0.4	0.2	0.0	0.0	0.0	70.7
7	29.4	0.5	**77.1**	0.3	0.0	0.0	0.0	0.0	0.0	14.7	0.6	0.1	0.1	0.0	0.0	0.0	77.6
8	8.7	0.0	48.8	2.4	0.0	0.1	0.0	0.0	0.0	5.7	1.5	0.0	0.0	0.0	0.0	0.0	44.3
9	7.7	0.0	28.6	6.0	0.4	0.6	0.0	0.0	0.0	5.3	1.2	0.1	0.0	0.0	0.0	0.0	30.2
10	4.6	0.0	55.9	22.1	0.2	0.2	0.0	0.0	0.0	0.6	9.8	0.0	0.0	0.0	0.0	0.0	58.9
11	2.9	0.0	53.9	18.6	0.7	0.4	0.0	0.0	0.0	0.4	2.2	0.0	0.2	0.0	0.0	0.0	52.7
12	0.6	0.0	9.5	**65.3**	2.6	3.3	0.2	0.0	0.0	0.0	5.2	0.0	0.5	0.0	0.0	0.0	42.7
13	0.1	0.0	7.7	58.2	13.9	4.2	0.0	0.0	0.0	0.0	0.4	0.0	0.2	0.0	0.0	0.0	37.3
14	3.9	0.0	8.4	14.9	1.5	4.0	1.7	0.1	0.3	0.8	9.7	0.4	1.7	0.0	0.1	0.0	20.9
15	0.5	0.0	4.6	19.8	9.7	4.5	0.3	0.0	0.0	0.1	0.5	0.0	0.2	0.0	0.0	0.0	15.7
16	1.2	0.0	3.6	8.1	3.7	2.6	0.4	0.1	0.2	0.5	1.0	0.2	0.6	0.0	0.0	0.0	9.1
17	0.6	0.0	7.9	42.0	5.5	22.1	2.3	0.0	0.0	0.0	1.9	0.0	1.4	0.0	0.0	0.0	34.0
18	0.7	0.0	2.4	8.1	2.1	11.9	2.4	0.1	0.0	0.1	0.8	0.0	2.1	0.0	0.0	0.0	10.5
19	1.1	0.1	1.0	1.1	1.1	3.1	1.6	0.4	0.1	0.3	0.4	0.0	1.8	0.0	0.0	0.0	3.9
20	6.9	0.3	20.2	27.6	3.7	4.1	0.7	0.1	0.0	0.3	45.7	0.1	6.2	0.0	0.0	0.0	47.2
21	1.3	0.2	22.1	21.6	41.7	17.7	0.5	0.0	0.0	0.0	2.6	0.0	14.2	0.0	0.0	0.0	37.6
22	0.1	0.0	2.4	8.9	**96.9**	**62.3**	3.0	0.0	0.0	0.0	0.1	0.0	7.7	0.0	0.0	0.0	29.5
23	0.9	0.5	4.2	2.6	33.8	17.2	2.2	0.1	0.0	0.1	0.6	0.0	25.2	0.1	0.0	0.0	14.7
24	0.3	0.0	0.3	0.7	0.2	15.3	**27.8**	7.0	0.1	0.1	0.6	0.0	4.1	0.0	0.1	0.0	18.5
25	0.0	0.0	0.0	0.0	0.0	1.9	24.9	**38.9**	2.1	0.0	0.1	0.0	0.4	0.0	0.2	0.0	22.7
26	0.3	0.0	0.1	0.0	0.1	0.8	5.0	6.7	0.4	0.1	0.1	0.0	1.1	0.0	0.1	0.0	5.0
27	0.2	0.0	0.9	0.9	5.4	20.3	17.3	1.0	0.0	0.0	0.1	0.0	2.4	0.0	0.0	0.0	13.8
28	0.0	0.0	0.0	0.0	0.0	0.0	0.1	4.6	**87.4**	0.0	0.0	0.0	0.8	0.0	0.4	0.0	14.9
29	16.9	1.2	8.5	4.9	0.1	0.7	0.2	0.0	0.0	5.0	**89.5**	3.7	3.1	0.0	0.1	0.0	44.3
30	18.9	1.3	9.3	1.8	0.2	0.6	1.0	0.3	0.2	12.6	38.1	4.0	1.5	0.0	0.3	0.0	32.6
31	23.7	9.0	7.5	0.1	0.0	0.0	0.1	0.1	0.0	**47.9**	24.1	**48.4**	0.6	0.1	0.3	0.0	42.8
32	9.0	1.1	5.1	2.0	1.4	1.2	1.0	0.2	0.1	3.1	47.3	3.2	11.1	0.0	0.6	0.0	26.5
33	10.7	2.4	8.6	3.6	0.7	0.9	0.7	0.1	0.0	2.8	28.4	2.4	8.8	0.0	0.2	0.0	24.9
34	11.7	1.5	6.8	1.3	0.6	0.9	1.2	0.4	0.2	7.5	15.2	3.2	3.2	0.0	0.3	0.1	20.2
35	3.9	1.2	1.2	0.1	0.2	0.2	0.3	0.3	0.1	2.8	9.2	3.3	5.6	0.1	1.4	0.0	10.7
36	2.3	1.2	0.8	0.0	0.1	0.0	0.1	0.1	0.1	2.0	2.2	2.2	3.6	0.2	1.1	0.1	6.3
37	1.0	1.1	0.3	0.0	0.0	0.0	0.0	0.1	0.3	1.9	0.2	2.3	1.0	0.3	0.5	0.3	3.3
38	8.9	7.4	2.5	0.0	0.0	0.0	0.0	0.0	0.0	23.3	2.0	33.9	0.4	0.3	0.6	0.1	19.5
39	2.3	2.4	0.8	0.0	0.0	0.0	0.2	0.2	0.0	4.8	0.7	9.1	2.2	0.5	0.5	0.0	6.4
40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
41	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	1.1	2.2	3.7
42	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.1	0.0	0.1	3.5	2.6	9.3
43	0.1	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.6	0.2	0.9	1.2	0.5	3.5
44	0.1	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.7	0.8	1.0	2.9	0.2	7.1
45	0.4	16.8	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.5	0.0	0.3	1.2	20.3	5.1	0.2	18.9
46	0.7	3.3	0.3	0.1	6.1	2.7	0.8	0.2	0.3	0.2	2.7	0.5	**113.4**	1.1	7.0	0.1	27.8
47	0.1	1.1	0.0	0.0	0.0	0.0	0.0	2.3	16.8	0.1	0.0	0.2	0.2	**32.8**	**7.4**	**16.0**	34.7
48	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	66.3	0.2	0.0	0.1	0.5	0.9	7.0	14.0	32.7
49	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	18.7	0.3	0.0	0.1	0.0	0.5	3.2	1.5	10.8
50	0.1	0.2	0.0	0.0	0.0	0.0	0.1	0.0	3.9	0.0	0.0	0.0	1.1	0.0	1.7	0.4	4.8
51	0.6	0.0	0.2	0.0	0.0	0.0	0.0	0.0	1.6	2.3	0.0	0.0	4.1	0.0	1.7	0.2	5.1
	C7	C8	C1	C4	C2	C3	C11	C12	C13	C5	C10	C6	C9	C16	C14	C15	All clusters
% of genome covered by cluster	0.58	0.17	0.76	0.50	0.09	0.20	0.43	0.27	0.14	0.10	0.26	0.20	0.07	0.20	2.22	0.51	# 6.73

Supplementary Figure 7: Comparison with published ChromaSig clusters illustrates increased coverage. We compared our state assignments to the published ChromaSig[4] annotation based on a subset of 21 datasets (20 methylation marks and H2AZ). As ChromaSig clusters were learned using 4kb intervals but only cluster centers were reported, we extended each of 49,340 reported genomic loci by 2kb in either side, assigning the full 4kb interval to one of the 16 ChromaSig clusters. We computed the fold enrichment of each state for each location assigned to one of the 16 ChromaSig clusters. We then ordered the ChromaSig clusters to match the 51 states based on the states of maximum enrichment for each cluster, resulting in a general mapping of the correspondence between the chromatin states and ChromaSig clusters. We do however find some overlap between promoter and enhancer assignments, likely due to the difference in resolution between the two methods (4kb vs. 200bp intervals). In addition to the difference in resolution, we find a significant difference in coverage. On average, only 7% of the genome is assigned to a ChromaSig cluster, and although it is higher for promoter (30-78%) and intergenic candidate enhancer (20-44%) states, a large majority of most states remain unassigned by ChromaSig. Each entry denotes the fold enrichment for different ChromaSig clusters, the bottom row indicates the total percentage of all 200bp intervals of the genome that each ChromaSig cluster represents, the last column indicates the percentage of the state that was assigned by ChromaSig to any cluster.

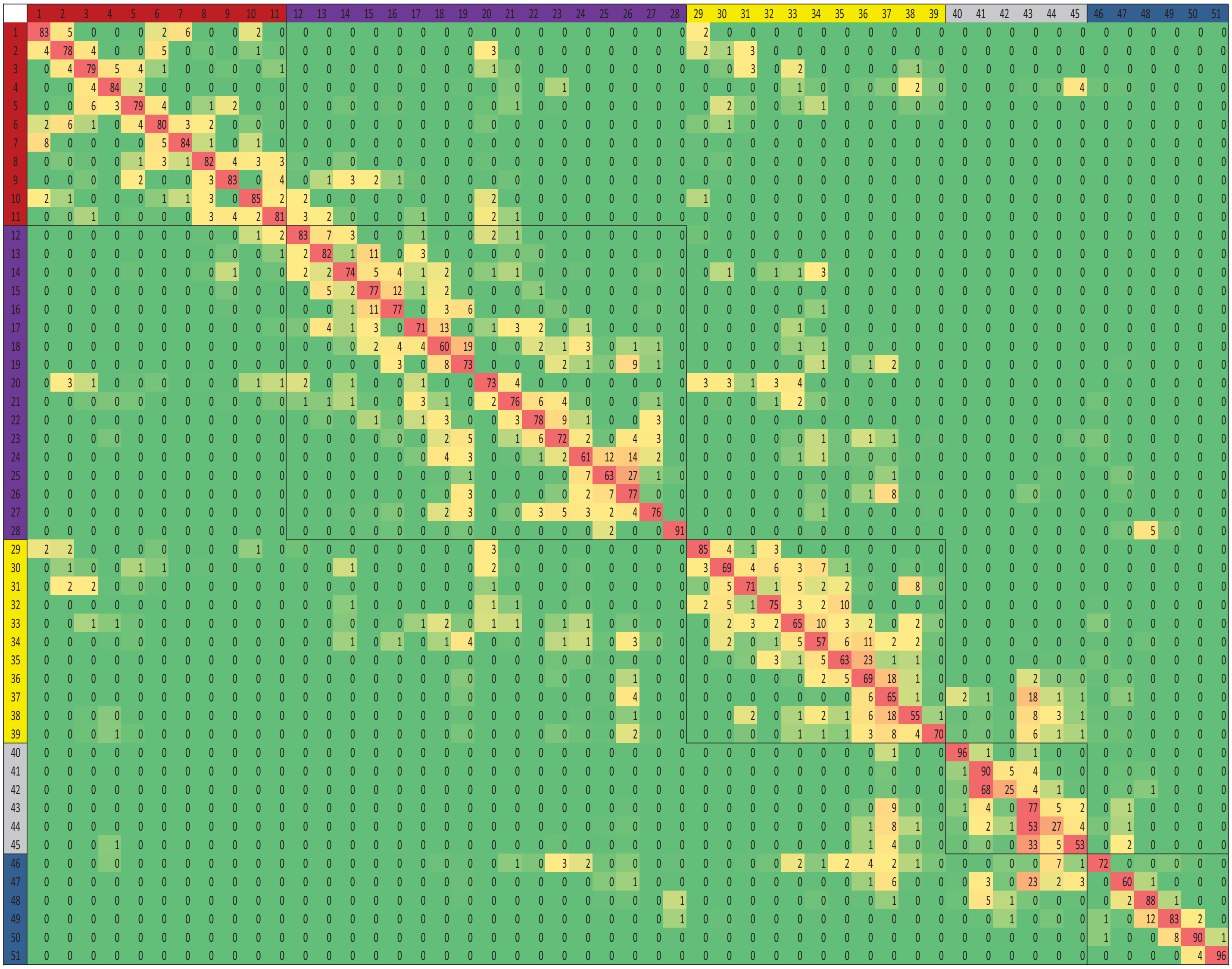


a. Bayesian Information Criterion (BIC) score increase with increasing numbers of states

b. Convergence of 79-state model with random initialization

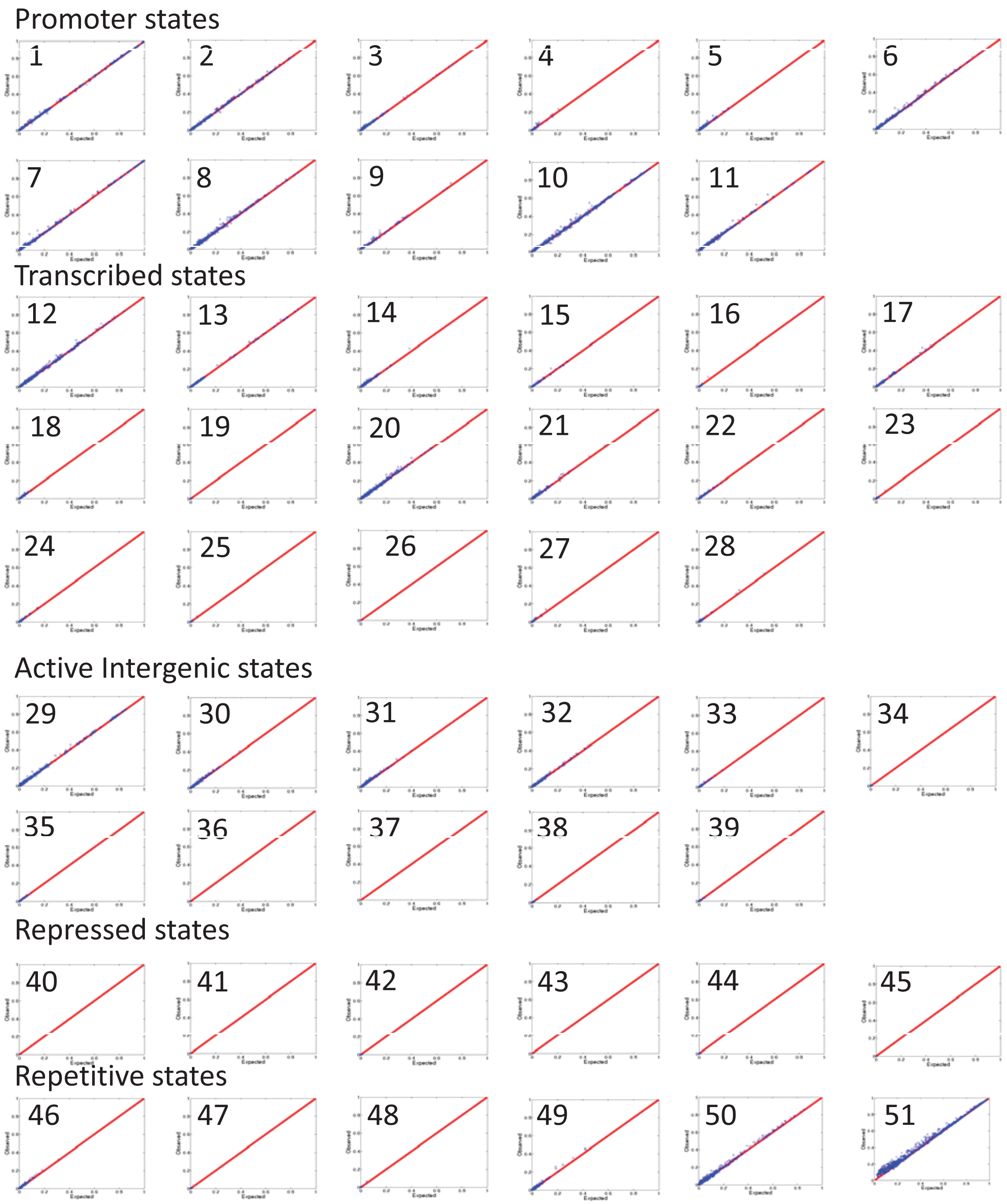
c. Convergence of 51-state model after nested initialization

Supplementary Figure 8: Bayesian Information Criterion (BIC) score with increasing numbers of states and convergence of model training. (a) BIC score both for models with randomly initialized parameters (blue) and for models based on the nested initialization scheme (red). BIC score for each model is its log likelihood score minus a penalty term, computed as the number of parameters in the model divided by two times the natural log of the number of data points (in our case defined as the number of 200bp intervals). The figure shows the BIC scores of the model based on the nested initialization strategy are greater or comparable to the BIC scores obtained based on random initialization. The figure also shows the BIC score alone is not a sufficient criterion to enable selection of a model with a relatively small number of states for this data as it continued to increase past 70 states (see **Supplementary Notes**). (**b and c**) The log likelihood of the model versus the number of full iterations of the expectation-maximization algorithm used for parameter learning. Plots shown are for (b) the 79 state model with highest likelihood from the first pass and (c) the 51 state model from the second pass. The plot shows that the 300 iterations used were sufficient for the model training procedure to have essentially converged on a local maximum.

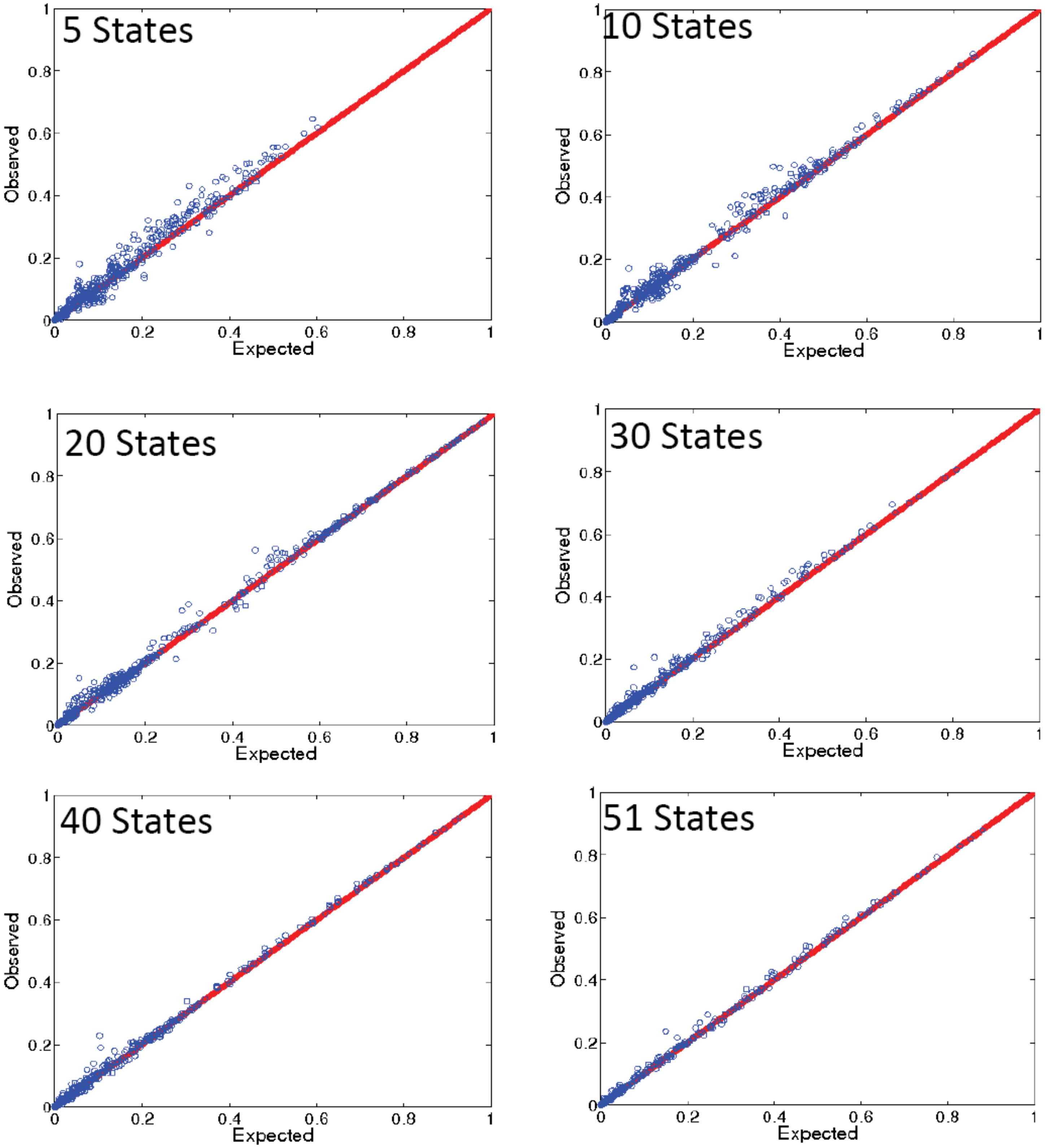


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1	83	5	0	0	0	2	6	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	4	78	4	0	0	5	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	2	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	4	79	5	4	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	4	84	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	4	0	0	0	0	0	0
5	0	0	6	3	79	4	0	1	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	2	6	1	0	4	80	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	8	0	0	0	0	5	84	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	1	3	1	82	4	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	2	0	0	3	83	0	4	0	1	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	2	1	0	0	0	1	1	3	0	85	2	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	0	3	4	2	81	3	2	0	0	0	1	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	1	2	83	7	3	0	0	1	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	1	2	82	1	11	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	1	0	0	2	2	74	5	4	1	2	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	5	2	77	12	1	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	11	77	0	3	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	4	1	3	0	71	13	0	1	3	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4	4	60	19	0	0	2	1	3	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	8	73	0	0	0	2	1	0	9	1	0	0	0	0	0	0	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	3	1	0	0	0	0	0	0	1	1	2	0	1	0	0	1	0	0	73	4	0	0	0	0	0	0	0	3	3	1	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	3	1	0	2	76	6	4	0	0	0	1	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	0	0	3	78	9	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	0	1	6	72	2	0	4	3	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	0	0	1	2	61	12	14	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	7	63	27	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	2	7	77	0	0	0	0	0	0	0	0	0	1	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	3	5	3	2	4	76	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	91	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0
29	2	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	85	4	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	3	69	4	6	3	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5	71	1	5	2	2	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2	5	1	75	3	2	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	1	1	0	1	1	0	0	0	0	0	2	3	2	65	10	3	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	4	0	0	0	1	1	0	3	0	0	0	2	0	1	5	57	6	11	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	5	63	23	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	5	69	18	1	0	0	0	0	2	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	6	65	1	0	2	1	0	18	1	1	0	1	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	1	2	1	6	18	55	1	0	0	0	8	3	1	0	0	0	0	0	0
39	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1	1	3	8	4	70	0	0	0	6	1	1	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	96	1	0	1	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	90	5	4	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	68	25	4	1	0	0	0	1	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	1	4	0	77	5	2	0	1	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	1	0	0	2	1	53	27	4	0	1	0	0	0	0
45	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	0	33	5	53	0	2	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	2	0	0	0	0	0	0	0	0	2	1	2	4	2	1	0	0	0	0	0	7	1	72	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	6	0	0	0	3	0	23	2	3	0	60	1	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	5	1	0	0	0	0	2	88	1	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	12	83	2	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	8	90	1
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	96

Supplementary Figure 9: State discrimination with all 41 marks: overlap in posterior probabilities in genome-wide probabilistic assignments. Given the probabilistic assignment of each genomic location into states, we can evaluate the frequency with which two states show non-zero probability (overlap) in the same genomic interval. Each row (summing to 100%), shows for each state the distribution of overlap in posterior probability for all states, each entry (state1,state2) denoting the average posterior probability of being in state2 for locations assigned to state1. High off diagonal values would denote uncertainty in distinguishing between a given pair of states at specific genomic locations, and specifically the directionality of mis-assignments. Nearly all states (except 42 and 44) contain the majority of the posterior probability at their locations on average, and in most cases substantially more (on average 74%) indicating the states are well-separable according to the model. For states 42 and 44 there is difficulty for the model to confidently distinguish them at any specific location from states 41 and 43 respectively. However for the vast majority of state 41 and 43 the model can be confident they are not 42 and 44 respectively, which is possible due to the high self-transition probabilities for those states.



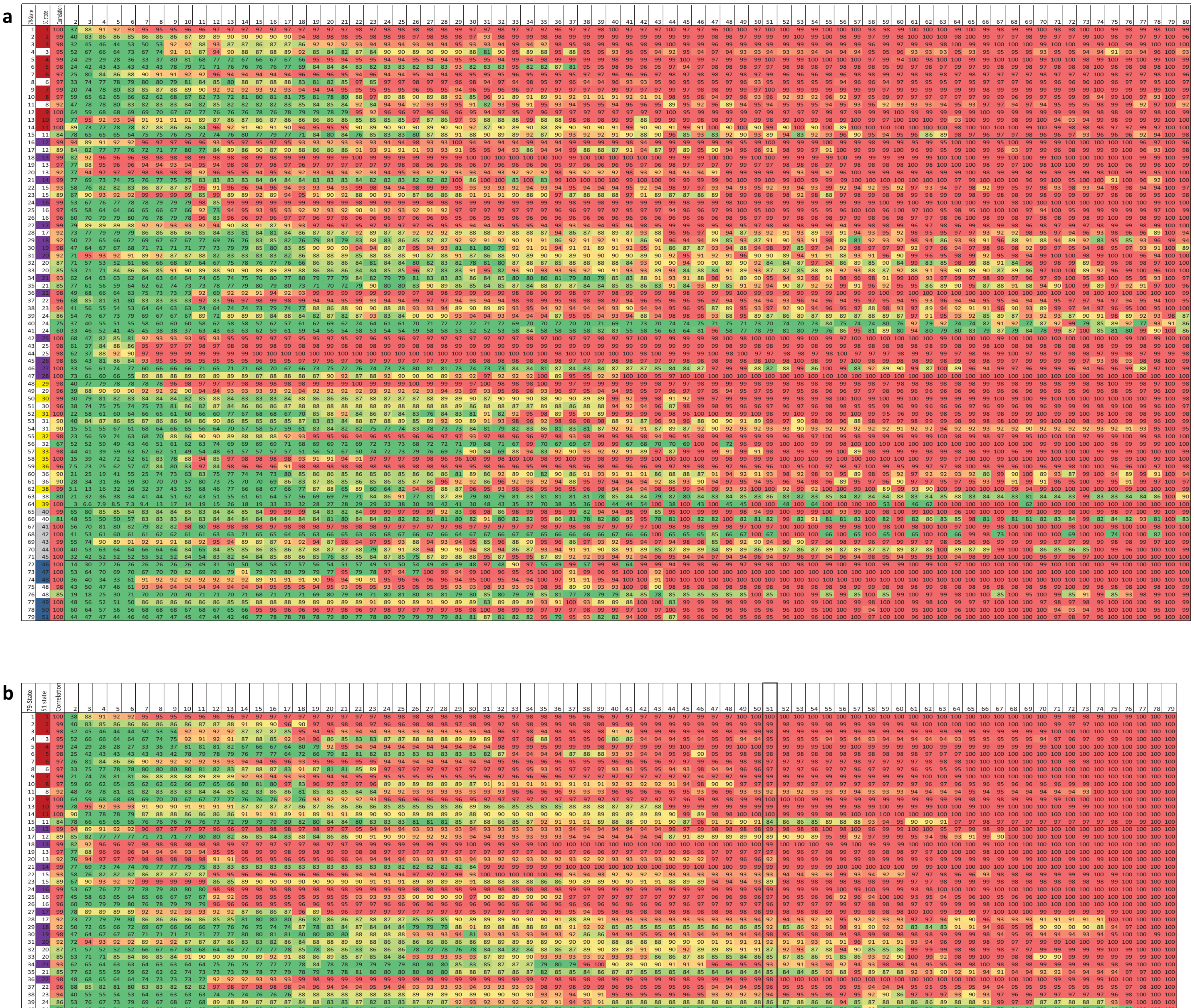
Supplementary Figure 10: Pairwise expected vs. observed mark co-occurrence for each chromatin state in a 51-state model reveals conditional mark independence. We evaluated the assumption of conditional independence of each pair of marks in each state of our 51-state model. Each plot corresponds to one state and each point (blue) corresponds to a pair of marks, and compares the expected frequency of a pair of marks being observed together under the model (x-axis, computed by multiplying the emission probabilities of the two marks), compared to how often a pair of marks in a state are actually observed together (y-axis). When the expected count agrees with the observed counts, points will be on the x=y line (red). The plot validates our model assumption, that conditioned on a state the pairs of marks are independent.



Supplementary Figure 11: Chromatin marks become conditionally independent with increasing numbers of states. A pairwise expected vs. observed mark co-occurrence plot as in **Supplementary Figure 10** is shown from models with increasing numbers of states, showing that larger numbers of states better capture the observed dependencies between chromatin marks. Expected vs. observed pairwise counts are shown for models with 5, 10, 20, 30, 40, and all 51 states, as obtained based on our nested initialization procedure. For each plot, we show the state most correlated to state 6 in the 51 state model in terms of the emission parameters. The comparison of the six plots shows that as more states are added, the points become increasingly closer to the y=x line, meaning that as the number of states increases, pairs of marks become conditionally independent.

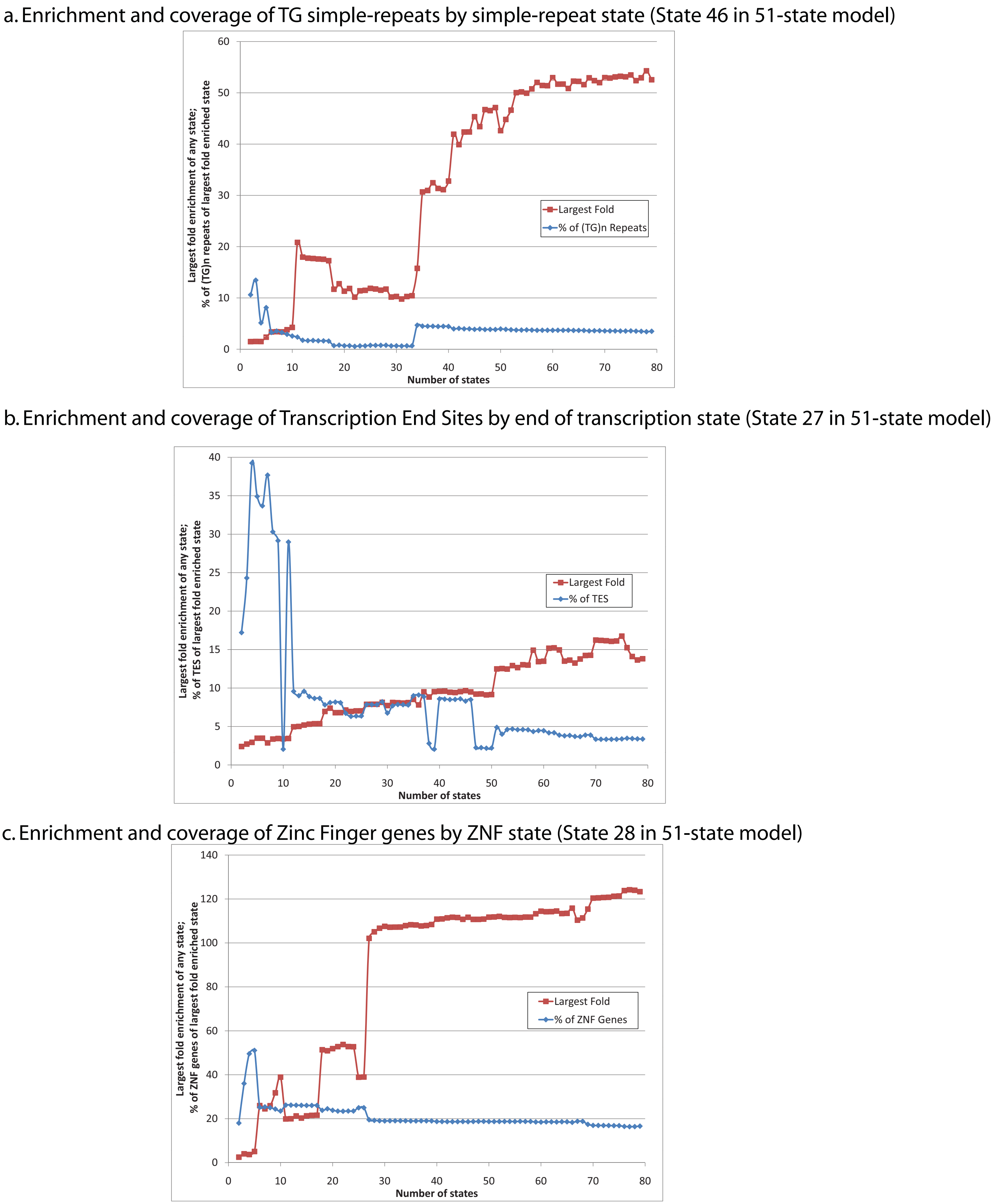
State	H3K14ac	H3K23ac	H4K12ac	H2AK9ac	H4K16ac	H2AK5ac	H4K91ac	H3K4ac	H2BK20ac	H3K18ac	H2BK120ac	H3K27ac	H2BK5ac	H2BK12ac	H3K36ac	H4K5ac	H4K8ac	H3K9ac	PolII	CTCF	H2AZ	H3K4me3	H3K4me2	H3K4me1	H3K9me1	H3K79me3	H3K79me2	H3K79me1	H3K27me1	H2BK5me1	H4K20me1	H3K36me3	H3K36me1	H3R2me1	H3R2me2	H3K27me2	H3K27me3	H4R3me2	H3K9me2	H3K9me3	H4K20me3	51 state	Correlation	Best 51-only	Correlation
1	4.1	26.3	24.1	14.5	35.4	15.4	92.8	92.3	91.8	98.9	99.3	99.7	98.3	72.2	85.0	93.6	90.6	86.3	44.5	14.2	98.7	99.1	97.6	65.4	92.9	4.7	4.0	14.2	24.0	11.9	3.4	0.4	3.1	1.8	2.0	0.2	0.2	0.2	0.5	0.2	1.9	1	1.00		
2	1.5	15.1	7.5	2.6	4.3	2.5	38.9	38.2	40.5	69.7	72.5	83.8	66.3	19.8	32.4	53.1	58.5	44.8	20.6	13.1	95.9	99.5	91.9	45.2	80.8	2.5	1.9	5.6	10.8	3.6	0.7	0.2	1.3	0.7	1.0	0.2	0.7	0.1	0.2	0.1	1.2	2	0.99		
3	0.3	3.5	0.8	0.7	0.4	0.3	6.7	5.1	5.8	16.2	13.7	8.9	2.9	1.0	2.1	5.5	9.7	4.3	5.7	18.1	78.8	96.8	72.2	28.0	67.9	0.3	0.1	0.4	0.9	2.0	3.8	0.1	0.7	2.0	1.4	0.2	5.6	0.1	0.1	0.2	1.3	3	0.98		
4	1.2	11.7	3.7	1.4	1.2	5.1	13.5	12.8	10.6	22.5	24.1	39.5	25.6	3.9	6.4	15.5	17.0	4.7	12.9	9.6	65.1	71.2	83.1	73.9	70.5	4.1	4.5	10.6	12.4	10.1	3.1	0.7	2.6	1.8	1.5	0.3	0.2	0.2	0.2	0.1	0.7	3	0.95		
5	0.1	0.6	0.1	0.4	0.3	0.2	0.9	0.5	0.4	1.2	1.3	0.3	0.2	0.0	0.0	0.1	0.1	0.2	1.1	4.6	12.2	75.3	14.6	17.8	20.4	0.3	0.0	0.0	0.0	0.9	11.7	0.2	0.2	1.3	0.8	0.0	11.3	0.1	0.1	0.5	1.3	4	0.99		
6	0.0	0.2	0.6	0.8	1.6	0.4	16.2	6.9	3.1	8.3	13.7	15.2	13.3	0.8	2.0	0.5	0.8	3.0	43.2	17.3	15.4	81.0	10.5	2.5	14.0	5.7	3.4	1.8	0.0	4.0	12.7	0.1	0.1	0.1	0.4	0.0	0.0	0.0	0.1	0.0	1.1	5	0.98		
7	0.1	2.1	6.1	14.1	15.4	2.7	91.0	87.0	80.8	93.2	97.7	98.3	97.7	44.3	71.2	48.8	45.7	87.1	76.7	27.2	69.0	98.6	36.8	4.0	38.4	3.5	1.6	4.9	0.7	4.5	7.6	0.0	0.7	0.4	0.9	0.0	0.0	0.0	0.1	0.0	1.9	6	0.97		
8	0.0	0.8	1.7	4.0	3.6	0.8	59.7	39.6	26.9	52.5	61.4	68.8	64.4	6.3	18.5	5.1	6.8	35.8	66.2	25.7	49.2	98.5	25.8	3.1	32.8	6.3	3.4	4.8	0.2	7.1	14.6	0.0	0.5	0.4	0.9	0.0	0.0	0.0	0.1	0.1	1.6	6	0.97		
9	2.6	15.9	34.1	55.0	75.2	18.2	99.8	100.0	99.1	100.0	100.0	100.0	100.0	93.2	99.4	97.2	93.4	99.9	91.5	28.3	85.7	99.9	52.2	6.3	48.7	26.1	18.8	13.8	3.8	5.7	10.2	0.2	1.2	0.5	1.5	0.0	0.0	0.0	0.2	0.0	1.9	7	0.99		
10	1.3	13.6	4.6	6.3	2.5	5.5	33.7	34.9	19.6	29.0	49.4	81.8	74.9	14.5	14.6	10.0	5.7	46.3	55.8	9.8	23.5	97.1	55.5	13.2	44.9	85.9	80.8	18.9	1.3	4.2	37.1	6.5	0.3	1.0	0.9	0.0	0.0	0.0	0.0	0.0	0.4	8	0.97		
11	0.6	5.0	6.0	28.8	21.8	4.8	94.2	90.1	73.3	91.5	95.2	97.1	96.7	36.5	55.1	36.3	29.9	90.5	78.1	17.7	49.8	99.8	31.9	3.3	32.6	82.7	71.8	26.6	1.8	11.2	47.6	1.0	0.4	1.2	1.0	0.0	0.0	0.0	0.0	0.1	1.2	8	0.92		
12	0.4	6.5	2.6	0.9	0.4	1.7	2.7	5.5	1.6	1.5	8.7	27.4	20.7	1.8	1.9	1.4	1.4	4.8	32.3	3.8	3.0	77.9	41.4	28.3	37.7	87.3	84.0	12.8	2.5	9.0	48.4	7.4	0.3	0.5	0.8	0.0	0.1	0.0	0.0	0.2	0.5	9	1.00		
13	4.1	28.9	13.0	22.7	26.8	33.4	90.4	89.6	82.3	91.3	96.6	98.4	96.4	57.9	61.8	70.7	50.1	74.6	51.6	10.5	51.8	98.9	97.4	83.6	95.8	79.4	79.4	62.5	27.8	43.4	60.8	3.7	6.9	4.6	2.5	0.1	0.0	0.2	0.0	0.0	0.8	10	0.99		
14	2.0	24.2	4.8	4.4	4.0	10.5	33.0	31.7	18.3	26.1	43.4	63.6	55.0	7.4	7.7	17.0	10.2	20.1	33.6	6.6	16.6	93.0	93.9	95.6	94.9	88.4	89.1	57.9	25.9	58.3	82.2	10.6	6.9	4.6	2.7	0.1	0.0	0.3	0.1	0.1	0.6	11	1.00		
15	0.4	3.1	0.9	1.8	2.2	4.6	23.5	15.4	11.5	18.6	22.6	19.8	16.6	2.2	2.5	3.4	3.2	4.1	21.8	13.0	16.8	53.9	66.3	91.8	82.3	7.5	6.7	28.1	9.7	58.1	86.7	2.4	3.8	6.0	3.5	0.1	0.6	0.3	0.0	0.0	1.1	11	0.84		
16	3.0	16.1	7.6	1.8	11.5	27.0	50.2	39.2	25.9	25.0	43.5	43.8	44.1	4.8	6.1	23.3	12.2	1.4	17.1	2.6	0.3	3.0	64.8	97.4	86.1	90.5	89.8	77.1	44.8	87.1	92.1	12.6	3.5	5.6	3.2	0.3	0.0	0.5	0.2	0.0	0.3	12	0.99		
17	6.1	20.6	14.7	4.6	20.5	53.9	82.2	80.4	74.2	71.3	88.1	82.8	85.2	40.9	36.0	55.4	29.1	6.1	23.6	3.8	4.3	16.2	55.7	96.5	68.1	76.5	77.6	71.0	33.2	60.9	57.2	12.1	2.5	6.7	3.8	0.2	0.0	0.6	0.4	0.1	1.0	12	0.89		
18	1.9	17.2	4.2	0.9	2.2	9.0	8.3	7.1	2.6	3.6	6.6	10.2	5.6	0.5	1.2	7.8	4.5	0.5	11.2	1.7	0.1	5.3	77.0	96.7	96.5	96.9	97.5	70.3	54.6	89.4	93.8	13.3	4.1	3.8	2.2	0.3	0.0	0.3	0.2	0.1	0.3	13	0.99		
19	0.5	5.3	2.8	0.4	0.2	2.1	1.3	2.1	0.6	0.6	2.3	3.5	2.1	0.3	0.6	1.9	1.7	0.1	5.3	0.9	0.0	0.7	32.5	51.7	64.4	87.9	87.3	34.8	16.9	43.2	69.1	4.0	0.5	0.5	0.6	0.1	0.0	0.1	0.2	0.0	0.1	13	0.97		
20	0.2	0.8	2.5	0.6	5.2	6.6	14.7	6.2	2.7	0.9	6.7	2.5	4.9	0.5	0.4	2.0	1.4	0.1	12.9	1.2	0.0	0.2	15.0	46.2	37.9	88.6	83.1	76.0	14.8	87.1	98.5	16.7	0.9	4.7	4.1	0.1	0.0	0.8	0.1	0.2	0.9	13	0.92		
21	0.9	7.6	9.5	1.2	2.4	18.7	21.3	24.8	17.6	10.4	30.3	28.0	29.6	8.5	8.4	14.7	10.6	0.5	9.4	1.9	2.3	2.5	12.6	45.4	23.1	70.8	71.4	38.7	11.7	16.8	22.2	9.2	0.3	1.8	1.1	0.2	0.0	0.1	0.3	0.2	0.6	14	0.99		
22	0.0	0.2	0.6	0.2	0.1	0.3	0.1	0.2	0.1	0.0	0.1	0.2	0.1	0.0	0.1	0.1	0.3	0.0	0.6	0.3	0.0	0.0	0.7	0.4	3.1	11.1	9.4	9.8	1.2	1.4	6.0	1.7	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.1	0.0	15	0.93		
23	0.2	0.9	2.9	0.4	0.8	1.9	1.2	0.9	0.2	0.1	0.7	0.8	0.7	0.2	0.2	0.7	1.2	0.1	3.9	0.7	0.0	0.0	6.0	5.5	19.2	49.5	40.6	47.3	10.5	39.4	66.5	16.3	0.3	1.5	1.3	0.2	0.0	0.2	0.2	0.5	0.4	15	0.89		
24	0.2	0.9	2.7	0.2	0.2	1.1	0.3	1.0	0.3	0.1	0.9	1.1	0.9	0.2	0.4	0.3	1.0	0.0	4.2	0.7	0.0	0.1	3.5	4.1	13.0	88.8	89.3	24.0	3.8	9.9	39.9	2.1	0.1	0.4	0.6	0.0	0.0	0.0	0.1	0.8	0.1	16	0.99	15	0.97
25	0.1	1.0	2.4	0.4	0.3	3.9	2.0	2.9	1.4	0.7	2.7	1.5	1.4	0.7	1.1	1.2	1.9	0.0	1.7	0.9	0.8	0.3	1.0	2.6	2.9	40.8	38.9	19.8	2.8	1.9	2.3	6.6	0.1	0.8	0.7	0.2	0.1	0.0	0.2	2.3	1.0	16	0.97		
26	0.0	0.2	0.1	0.1	0.1	0.1	0.5	0.5	0.1	0.0	0.6	0.3	0.4	0.0	0.1	0.2	0.2	0.0	0.3	0.2	0.0	0.0	1.7	6.0	5.3	23.8	15.1	3.7	0.5	1.8	9.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16	0.96		
27	1.4	10.0	2.9	1.0	2.3	9.3	7.7	6.6	2.5	3.9	3.8	6.5	2.8	0.4	1.0	8.0	5.0	0.5	3.6	1.1	0.2	0.9	58.2	76.9	88.0	37.0	39.3	71.7	52.1	77.9	74.4	19.5	3.2	5.3	2.2	1.0	0.1	0.5	0.5	0.1	0.5	17	0.99		
28	0.5	5.4	1.2	0.4	0.3	2.8	1.8	2.5	0.8	1.5	1.9	3.2	1.0	0.3	0.8	3.2	2.2	0.2	1.4	0.8	0.1	0.3	21.7	29.2	49.5	16.5	18.7	33.0	15.7	12.9	16.1	3.5	0.5	0.4	0.5	0.4	0.0	0.1	0.3	0.0	0.1	17	0.92		
29	0.1	0.4	1.1	0.2	0.2	0.3	0.1	0.2	0.0	0.0	0.1	0.3	0.2	0.0	0.1	0.2	0.6	0.0	0.6	0.3	0.0	0.0	1.4	0.3	5.0	1.6	0.6	8.1	4.2	4.3	2.8	3.7	0.1	0.2	0.4	0.1	0.0	0.0	0.1	0.1	0.1	18	0.92		
30	0.1	0.4	0.3	0.2	0.0	0.9	0.2	0.4	0.1	0.1	0.3	0.1	0.1	0.1	0.2	0.2	0.2	0.0	0.2	0.3	0.1	0.0	0.4	0.5	1.2	4.2	2.8	6.2	0.5	0.4	1.1	0.9	0.1	0.4	0.4	0.2	0.1	0.0	0.2	0.6	0.1	19	0.98		
31	1.5	11.3	5.7	6.3	10.0	17.9	72.9	65.0	59.1	73.0	78.8	82.5	77.4	24.0	28.9	38.4	29.0	26.1	34.5	13.0	46.5	77.8	89.3	96.2	91.5	7.5	7.5	33.4	22.7	48.8	47.0	1.5	4.9	5.6	3.5	0.1	0.1	0.4	0.1	0.0	1.1	20	0.92		
32	1.9	15.2	6.6	1.5	4.8	14.3	20.9	19.1	11.0	18.2	16.9	43.4	27.7	2.0	4.4	30.4	25.0	1.6	8.4	3.1	4.9	3.0	43.5	81.6	67.1	5.9	4.8	33.0	41.8	40.3	23.4	24.0	2.1	3.1	1.7	0.7	0.1	0.3	0.4	0.0	0.3	20	0.87		
33	2.2	4.7	1.7	2.0	4.8	28.4	48.4	38.6	33.9	39.4	41.6	24.5	23.2	8.7	9.2	17.1	9.5	0.5	3.1	3.8	11.2	1.4	13.9	92.2	21.4	1.0	0.8	8.8	15.5	19.0	2.3	1.2	1.3	5.8	2.7	0.8	0.3	0.4	0.5	0.1	1.2	20	0.85		
34	0.1	0.3	0.4	0.7	1.2	4.6	13.5	6.2	5.0	3.1	8.0	1.8	2.9	0.8	0.5	0.7	0.6	0.2	7.5	7.2	0.4	4.4	5.3	49.1	12.8	1.3	0.8	7.6	1.5	27.6	73.7	0.8	0.4	4.0	2.8	0.0	1.3	0.3	0.0	0.1	0.6	21	0.93		
35	0.9	1.9	3.4	2.6	18.6	34.8	74.9	50.5	43.9	28.6	57.4	26.8	36.9	14.9	9.0	12.0	5.5	0.9	11.5	3.1	0.2	0.6	18.5	78.2	35.6	14.0	13.4	52.3	19.4	73.8	71.4	10.5	1.6	10.1	6.9	0.5	0.1	2.2	0.1	0.0	1.3	21	0.85		
36	0.0	0.0	0.4	0.5	5.4	1.3	6.1	0.9	0.3	0.0	0.8	0.3	0.6	0.0	0.0	0.2	0.2	0.0	6.0	0.9	0.0	0.0	1.9	1.7	5.5	3.1	1.0	27.3	4.2	65.9	86.5	14.7	0.2	3.7	3.9	0.1	0.0	1.0	0.0	0.0	0.5	22	0.98	23	0.97
37	0.2	0.5	2.1	0.9	9.2	6.1	18.7	6.3	2.9	1.0	6.3	3.8	6.0	0.3	0.3	2.2	1.9	0.2	12.6	2.4	0.0	0.1	14.1	41.1	33.1	17.5	11.2	67.0	16.5	88.7	98.2	22.3	1.6	9.7	7.1	0.2	0.0	2.6	0.0	0.0	1.0	22	0.96		
38	0.0	0.1	0.1	0.2	0.4	0.5	1.1	0.3	0.1	0.0	0.3	0.1	0.1	0.0	0.0	0.1	0.0	0.0	1.0	0.5	0.0	0.0	0.2	0.4	0.5	0.5	0.0	2.5	0.3	3.6	22.6	2.5	0.0	0.4	0.6	0.0	0.2	0.1	0.0	0.0	0.1	23	0.94		
39	0.4	2.2	2.1	0.7	2.6	2.5	1.8	1.4	0.3	0.4	0.4	1.7	0.9	0.1	0.3	2.7	2.7	0.2	2.1	0.8	0.0	0.1	15.9	11.4	39.1	4.9	2.1	37.9	35.4	53.6	29.7	38.2	1.4	5.4	2.2	1.1	0.0	0.5	0.4	0.2	0.5	24	0.86		
40	0.6	4.3	3.2	0.9	1.0	3.3	2.1	3.5	1.1	1.8	1.8	8.8	4.1	0.5	1.1	7.2	7.9	0.3	2.7	2.2	0.8	0.3	6.8	11.1	18.0	0.5	0.2	4.5	22.7	7.4	0.7	15.3	1.2	1.8	1.1	1.1	0.1	0.2	0.5	0.1	0.2	24	0.75		
41	0.1	0.2	0.1	0.6	1.8	3.9	5.1	1.7	0.6	0.2	0.9	0.4	0.4	0.1	0.2	0.4	0.1	0.0	1.1	0.5	0.0	0.0	0.3	0.5	0.8	0.3	0.1	2.7	3.4	8.5	1.5	2.6	0.2	1.5	1.3	0.1	0.0	0.3	0.0	0.0	0.2	24	0.60		
42	0.1	0.3	0.2	0.4	0.1	0.4	0.1	0.2	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.2	0.1	0.2	0.3	0.0	0.1	0.3	0.1	0.8	0.8	0.0	0.5	5.1	0.8	0.1	52.2	0.5	1.1	0.8	1.1	0.1	0.2	0.6	4.2	0.7	25	1.00		
43	0.0	0.1	1.3	0.2	0.3	0.2	0.1	0.2	0.0	0.0	0.1	0.4	0.3	0.1	0.1	0.1	0.7	0.0	1.5	0.5	0.0	0.0	0.1	0.0	0.4	0.4	0.0	1.1	2.7	1.0	0.2	12.0	0.0	0.1	0.4	0.0	0.0	0.0	0.0	0.0	0.0	25	0.98		
44	0.1	0.8	1.6	0.7	2.1	1.6	1.5	0.9	0.3	0.1	0.5	1.6	0.9	0.1	0.2	1.7	2.1	0.2	1.9	0.6	0.0	0.0	2.2	0.9	6.9	1.4	0.2	8.7	22.7	16.1	3.5	79.6	0.9	3.1	1.8	0.4	0.0	0.5	0.2	0.3	0.2	25	0.98	24	0.90
45	0.1	0.2	0.2	0.2	0.1	0.2	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.2	0.3	0.0	0.0	0.2	0.0	0.5	0.2	0.0	0.3	1.2	0.4	0.1	2.3	0.1	0.3	0.4	0.2	0.0	0.0	0.2	0.2	0.1	26	0.98		
46	0.0	0.7	5.6	0.5	1.9	1.6	2.1	1.0	0.5	0.1	1.2	3.3	3.9	0.2	0.1	2.6	3.8	0.1	29.9	2.0	0.0	0.1	1.8	2.4	5.3	5.5	1.3	12.9	3.6	9.9	43.1	40.2	0.3	1.0	1.5	0.0	0.0	0.2	0.0	0.3	0.2	27	1.00		
47	0.0	0.0	0.2	0.3	0.0	0.2	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	1.3	0.3	0.0	3.8	0.2	0.3	0.5	14.4	3.1	5.9	1.4	3.3	1.6	77.9	0.1	2.8	2.9	0.3	0.1	0.9	0.3	44.1	78.4	28	1.00		
48	4.4	5.0	6.7	4.5	19.8	61.5	90.5	87.8	88.5	81.0	93.7	73.7	78.0	62.9	50.2	51.3	22.4	2.2	7.0	2.3	8.8	1.3	12.8	86.5	19.2	2.7	3.1	17.4	18.5	27.9	4.9	3.5	0.7	6.5	3.7	0.4	0.1	0.4	0.7	0.2	6.5	29	0.98		
49	4.6	14.7	17.3	13.7	28.6	63.3	96.7	96.8	97.0	98.0	99.2	99.2	98.8	89.0	86.7	89.8	67.1	42.5	41.4	10.4	37.8	42.5	64.2	90.2	55.1	3.5	3.4	11.6	15.8	19.6	7.6	1.9	1.8	3.6	2.9	0.2	0.0	0.3	0.6	0.0	2.2	29	0.96		
50	1.2	3.6	10.5	3.3	2.7	15.6	29.9	45.3	47.3	34.3	66.2	74.7	78.8	41.7	38.0	26.8	21.2	7.2	14.7	4.0	13.6	13.0	12.2	30.6	7.0	4.1	4.7	4.4	1.8	2.0	1.3	1.8	0.1	0.7	0.9	0.0	0.1	0.1	0.2	0.4	1.7	30	0.99		
51	0.3	0.9	3.5	0.6	0.6	5.7	8.7	10.7	9.8	3.1	18.5	14.3	17.5	7.7	5.0	3.4	3.3	0.1	2.4	1.4	2.1	0.6	1.4	13.2	1.3	1.7	1.6	3.3	1.5	1.5	0.4	0.5	0.0	0.5	0.5	0.0	0.0	0.0	0.1	0.1	0.2	30	0.96	34	0.91
52	1.3	6.6	4.3	1.6	1.7	5.6	12.9	14.1	16.9	30.7	25.8	30.0	16.1	7.7	10.9	41.4	52.2	2.9	1.1	5.9	66.9	9.4	12.1	32.8	7.6	0.6	0.4	1.1	8.1	1.6	0.0	0.8	1.3	1.8	1.1	1.1	0.5	0.2	0.8	0.2	0.7	31	1.00		
53	4.7	15.2	11.5	4.4	11.4	17.7	58.8	58.0	61.2	80.8	77.2	82.2	70.8	29.6	36.0	82.6	78.6	11.7	4.8	4.9	76.5	22.5	47.3	84.0	39.8	1.2	1.1	5.2	26.1	11.2	0.2	1.8	2.0	3.1	2.0	1.1	0.2	0.3	1.0	0.2	2.6	31	0.90		
54	0.3	1.4	0.9	0.7	0.3	3.3	3.0	3.9	3.6	4.2	4.8	3.2	1.8	1.7	2.3	4.0	5.8	0.2	0.2	0.4	11.5	0.3	0.4	3.2	0.4	0.3	0.1	0.3	1.8	0.3	0.0	0.1	0.3	0.5	0.6	0.2	0.1	0.1	0.4	0.3	0.3	31	0.90		
55	1.1	1.2	0.9	1.8	6.2	56.6	75.0	62.7	74.1	47.9	63.4	10.4	11.9	32.2	26.9	17.8	5.3	0.3	0.9	1.4	3.9	0.1	0.7	21.0	1.7	0.9	0.8	5.6	7.2	7.6	1.1	2.2	0.8	5.3	2.2	1.5	0.4	0.4	0.7	0.2	2.7	32	0.98		
56	0.2	1.0	1.1	1.2	6.0	26.0	38.4	18.5	14.0	4.6	13.9	4.7	5.7	1.9	2.8	5.1	2.0	0.2	3.1	1.1	0.1	0.0	2.1	10.7	6.4	5.3	4.4	30.1	14.2	25.3	13.1	16.8	0.6	4.4	2.6	0.6	0.1	0.5	0.1	0.1	0.2	32	0.67		
57	0.9	2.8	0.3	0.7	0.6	5.6	6.6	5.6	3.6	7.9	5.8	3.4	1.6	0.7	1.1	2.7	1.9	0.3	1.1	4.0	8.6	1.2	9.4	62.4	11.9	0.3	0.2	1.7	3.8	5.2	1.5	0.2	0.8	2.6	1.4	0.6	0.6	0.2	0.3	0.1	0.2	33	0.98		
58	0.6	1.1	0.1	1.1	1.3	35.9	34.4	23.5	23.6	19.3	14.2	1.3	0.6	3.7	6.1	4.6	1.5	0.2	0.3	1.1	2.0	0.1	0.3	3.2	0.5	0.4	0.2	1.5	2.6	1.9	0.3	1.0	0.9	4.3	2.0	2.6	0.7	0.3	0.5	0.1	0.4	35	1.00		
59	0.1	0.2	0.1	0.4	0.2	7.8	4.8	3.1	2.1	1.3	1.6	0.1	0.1	0.5	0.7	0.2	0.1	0.0	0.1	0.4	0.1	0.0	0.0	0.1	0.1	0.2	0.1	0.3	0.3	0.2	0.0	0.2	0.1	0.7	0.6	0.2	0.1	0.1	0.1	0.2	0.2	36	0.96		
60	0.7	1.5	0.1	1.1	0.4	12.8	3.5	4.6	2.0	6.7	1.1	0.2	0.1	0.4	1.7	1.3	0.7	0.3	0.1	0.6	1.6	0.1	0.3	0.7	0.6	0.3	0.0	0.4	0.9	0.9	0.1	2.1	2.1	6.0	2.7	5.7	3.6	0.6	1.1	0.5	0.4	36	0.90		
61	0.0	0.1	0.0	0.3	0.1	1.1	0.3	0.4	0.1	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.2	0.1	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.1	0.4	0.4	0.2	0.3	0.1	0.1	0.2	0.1	36	0.90	37	0.89
62	0.5	3.0	0.5	0.8	0.1	0.7	0.3	0.7	0.4	2.3	0.8	1.4	0.4	0.3	0.8	5.0	11.7	0.4	0.1	2.1	59.1	2.3	2.3	5.0	0.9	0.2	0.1	0.1	1.6	0.3	0.0	0.4	1.2	1.3	1.1	1.6	2.5	0.2	0.7	0.3	0.3	38	0.99		
63	0.1	0.3	0.0	0.3	0.0	0.4	0.1	0.2	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.2	0.0	0.1	0.4	1.6	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.2	0.5	0.5	0.3	0.6	0.1	0.1	0.1	0.1	38	0.80		
64	0.1	0.3	0.4	0.5	0.1	1.1	0.6	0.7	1.0	0.7	0.9	0.3	0.2	0.3	0.4	1.0	1.6	0.1	4.1	88.4	10.9	1.1	1.3	4.0	1.0	0.2	0.0	0.2	0.4	0.7	0.6	0.2	0.7	2.5	1.2	0.2	0.8	0.1	0.0	0.2	0.6	39	1.00		
65	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.1	0.1	0.0	0.1	0.3	0.1	40	0.99		
66	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	40	0.81		
67	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.5	0.3	0.1	0.1	1.1	2.0	0.4	41	1.00		
68	0.1	0.1	0.0	0.9	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.2	0.1	0.6	0.6	0.1	0.0	0.1	0.1	0.1	0.0	0.0	2.0	0.5	0.0	0.5	1.5	2.2	2.5	9.8	1.3	0.9	6.1	2.2	4.3	42	1.00		
69	0.0	0.1	0.0	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.3	0.5	1.1	0.1	0.6	0.6	0.1	43	0.99		
70	0.5	1.0	0.0	1.1	0.2	0.3	0.0	0.3	0.0	0.2	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.3	0.1	0.6	1.5	0.2	0.2	0.4	0.3	0.1	0.0	0.0	0.3	1.3	0.0	1.4	3.4	5.7	3.2	9.0	10.3	1.0	2.6	0.5	1.1	44	1.00		
71	0.1	0.7	0.0	0.5	0.0	0.4	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.6	0.6	0.3	0.1	0.3	0.1	0.1	0.0	0.0	0.0	0.1	0.2	0.3	0.5	1.0	0.9	0.4	16.7	0.2	0.3	1.1	0.4	45	1.00		
72	1.7	0.8	0.5	1.6	4.6	5.1	4.3	3.2	0.6	4.3	2.3	0.9	0.7	0.5	0.5	0.7	0.4	0.6	0.8	2.5	1.2	3.2	2.6	31.3	6.6	0.4	0.1	0.9	10.4	73.8	19.3	7.6	20.8	54.0	28.5	10.5	7.7	14.5	4.5	1.9	19.2	46	1.00		
73	0.0	0.1	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.1	0.7	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	1.4	0.1	0.4	0.7	0.3	3.3	0.1	0.7	34.3	1.8	47	1.00		
74	0.1	0.0	0.0	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.3	0.2	0.0	0.0	0.0	0.6	0.0	0.2	0.3	0.2	0.0	2.2	0.2	0.9	1.5	0.9	0.7	1.3	1.0	18.4	55.1	48	1.00		
75	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.1	0.1	0.1	0.5	0.4	0.1	0.2	0.7	4.2	7.7	48	0.98		
76	1.1	0.9	2.6	1.1	1.4	14.1	13.7	15.7	17.1	11.7	22.3	8.6	11.6	9.4	6.0	5.8	4.0	0.2	1.0	0.8	7.8	6.4	1.6	16.7	1.5	6.4	4.0	7.0	4.2	6.1	0.7	10.2	0.4	4.5	3.0	1.0	0.9	0.7	1.7	12.3	70.4	48	0.85		
77	2.8	0.6	0.6	4.9	11.6	0.1	0.2	0.3	0.2	0.1	0.1	0.1	0.2	0.2	1.5	0.4	0.8	2.9	1.0	3.2	4.0	2.6	0.3	0.4	0.8	1.3	0.1	0.0	12.1	5.6	0.3	19.2	11.5	14.5	43.8	17.9	5.3	57.1	28.7	51.4	86.0	49	1.00		
78	24.3	5.1	9.5	18.6	46.3	1.8	1.4	2.4	1.1	1.1	1.2	0.7	0.8	1.9	7.9	3.0	3.5	15.2	16.5	28.3	19.3	41.2	9.6	19.2	16.1	9.2	1.6	1.5	64.5	41.5	9.7	82.4	72.1	73.8	86.0	68.2	31.9	87.9	79.2	81.5	97.4	50	1.00		
79	80.5	40.8	69.9	70.3	94.4	22.4	23.7	33.5	30.8	21.7	23.2	13.6	20.3	27.7	57.6	37.9	47.4	68.3	72.5	82.0	75.8	95.9	64.5	77.3	79.5	52.3	27.4	28.0	95.5	88.4	63.1	96.3	97.0	98.8	99.0	91.5	90.2	97.8	96.7	92.9	99.7	51	1.00		

Supplementary Figure 12: Emission probabilities of 79-state model used for nested initialization. Emission parameters (multiplied by 100 for clarity) of the 79-state model showing the best BIC score of all models learned across the three random initializations for models between 2 and 80 states. For each state in the 79-state model is shown the state with the highest correlation of emission parameters (column labeled '51 state'). Best matches for a state from the 51-state model with a state of the 79-state are colored according to our promoter/transcribed/active-intergenic/repressed/repetitive color scheme. The 'correlation' column shows the correlation value with the best state from the 51-state model. Five states from the 51-state model did not have a best bi-directional best match with a state in the 79 state model, and the best match and correlation of these are listed in the last two columns.



Supplementary Figure 13: Advantage of nested-initialization strategy for consistent state recovery using a small number of states. **a. State recovery using random-initialization strategy.** Recovery of each state of 79-state model (rows) with randomly-initialized models at increasing numbers of states (columns). Each entry denotes the correlation (multiplied by 100 for clarity) between the emission matrix of the 79-state model (as shown in **Supplementary Figure 12**), and the emission matrix of the best-correlated state in the randomly-initialized model. The figure shows that some states (e.g. 64, 72) are recovered under some random initializations with few states but not recovered in other random initializations with more states, in contrast to nested-initialization model shown in the bottom panel.

b. State recovery for nested-initialization strategy. Same recovery values shown for nested-initialization models. This figure shows that once a state is recovered by a lower-complexity model, it is then consistently recovered for higher-complexity models in most cases. Black box denotes the 51-state nested-initialization model which was analyzed in detail in this paper, selected as the model with the smallest number of states that recovered the end of transcription state (state 27 in the 51-state model, state 46 in the 79-state model).

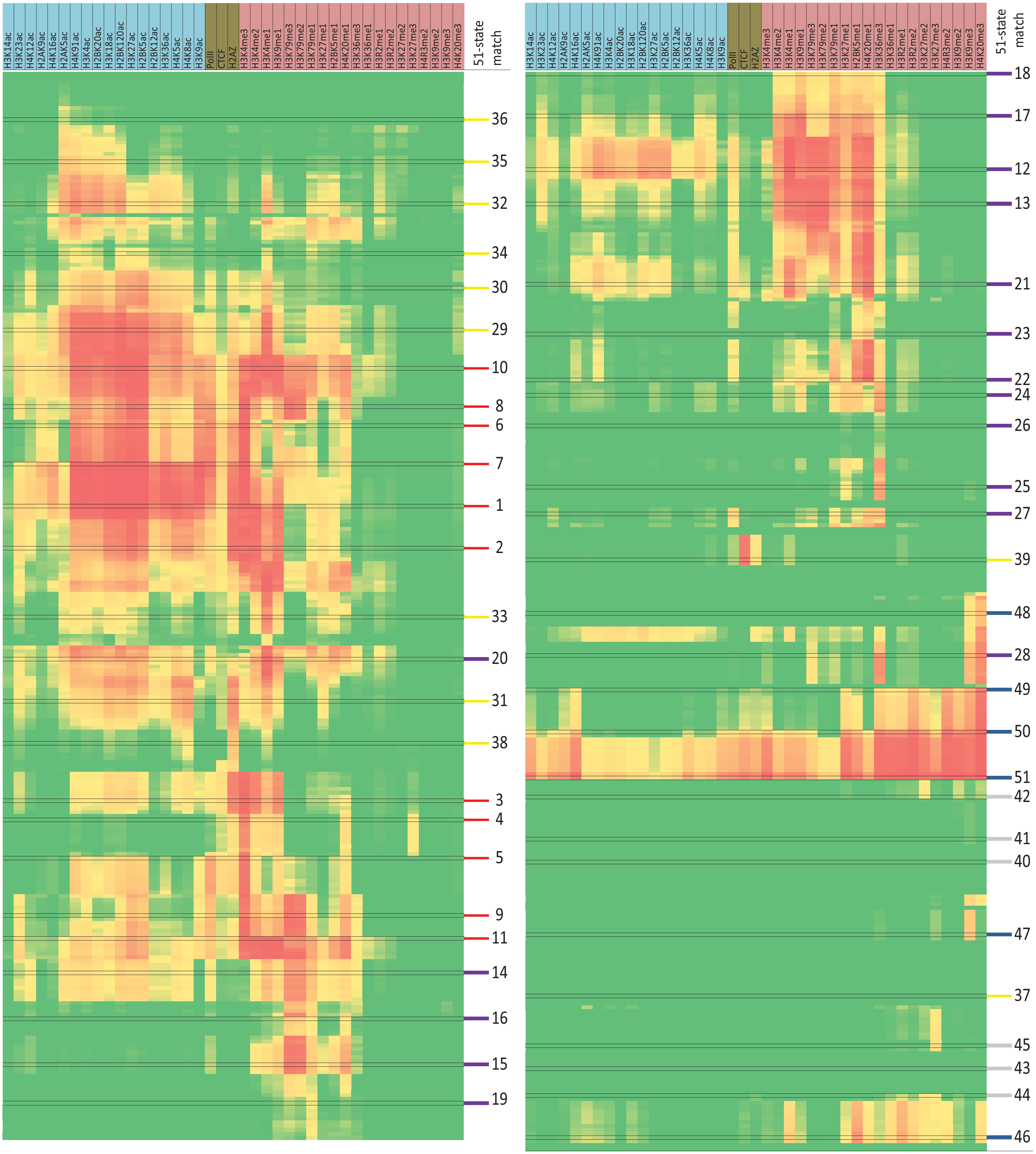


a. Enrichment and coverage of TG simple-repeats by simple-repeat state (State 46 in 51-state model)

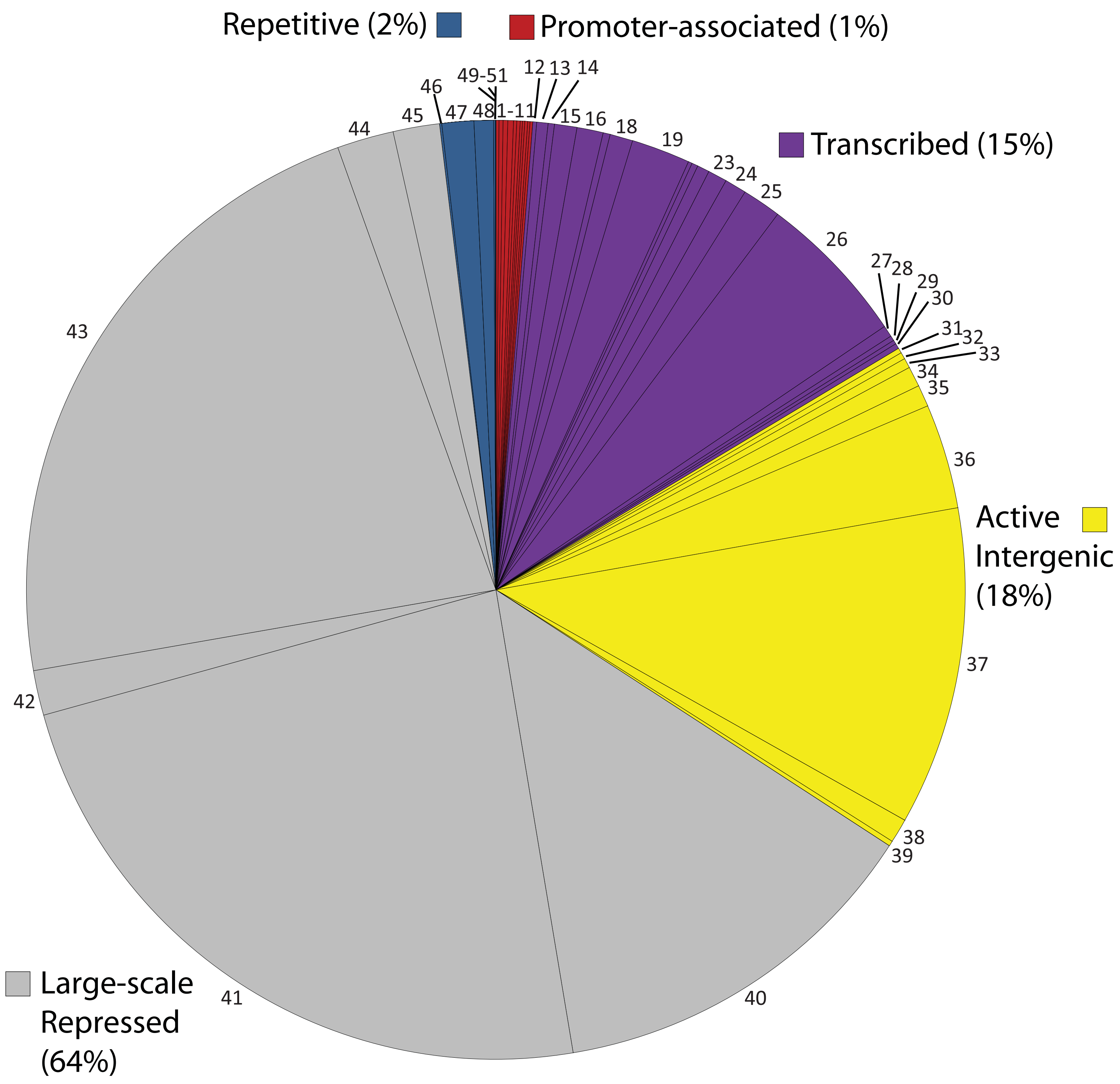
b. Enrichment and coverage of Transcription End Sites by end of transcription state (State 27 in 51-state model)

c. Enrichment and coverage of Zinc Finger genes by ZNF state (State 28 in 51-state model)

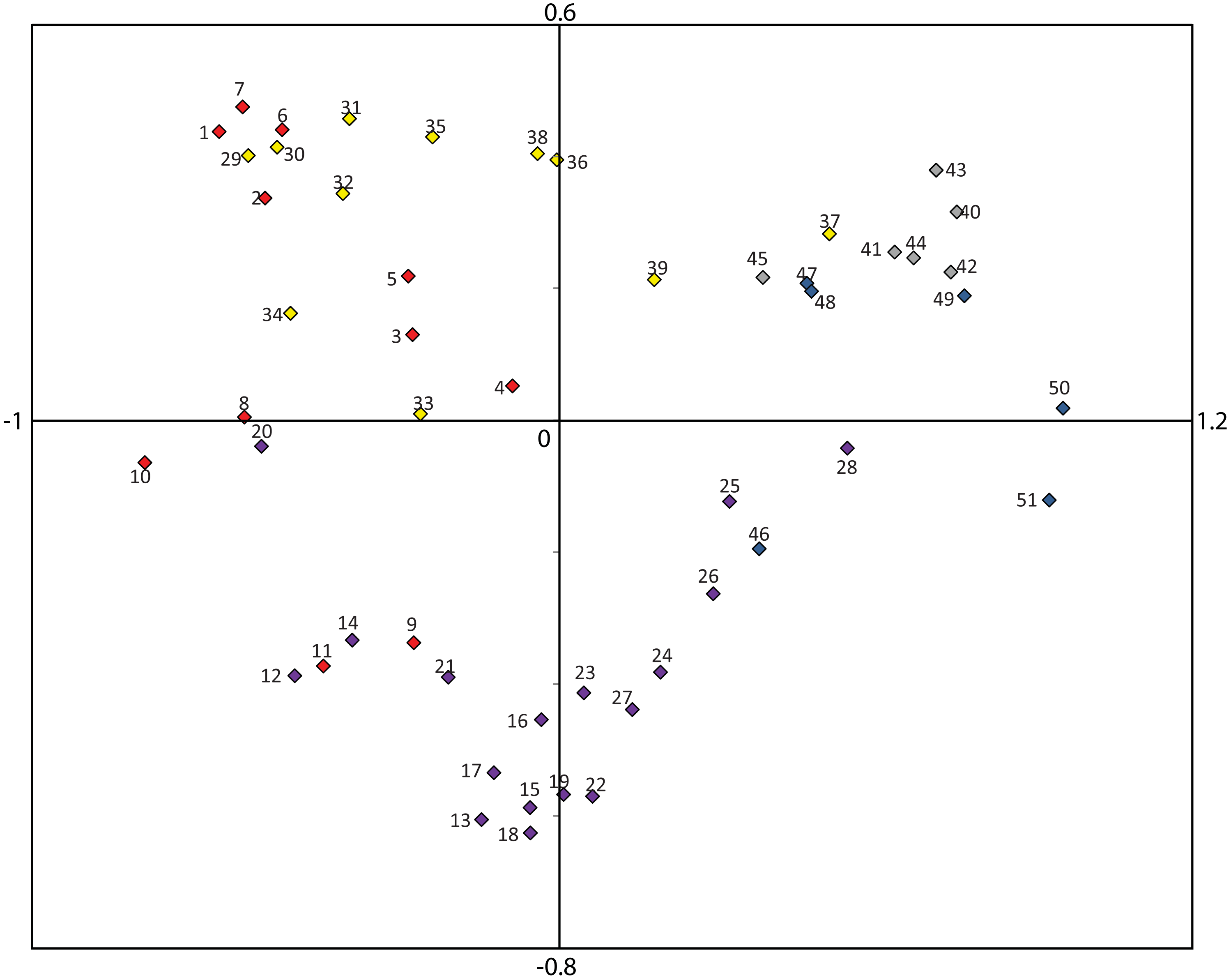
Supplementary Figure 14: Maximal state enrichments for three different types of genomic elements by models of increasing numbers of states. Each plot shows the maximal fold enrichment for the genomic element across all available states in the set of nested initialized models (**Supplementary Figure 13b**) (red line), and the corresponding coverage for that state (blue line), for models with increasing numbers of states. **a. TG simple repeats**. For models with at least 35 states, a state of the model has greater than 30 fold enrichment for TG simple repeats recovering about 4% of all TG simple repeats, which was not observed in models with fewer numbers of states **b. Transcription End Sites**. For models with at least 51 states, the largest enrichment becomes consistently greater than 12-fold, while capturing approximately 5% of all transcription end sites. **c. ZNF genes**. For models with at least 27 states, the enrichment for ZNF genes is consistently greater than 100-fold, while a single state captures approximately 20% of all ZNF genes.



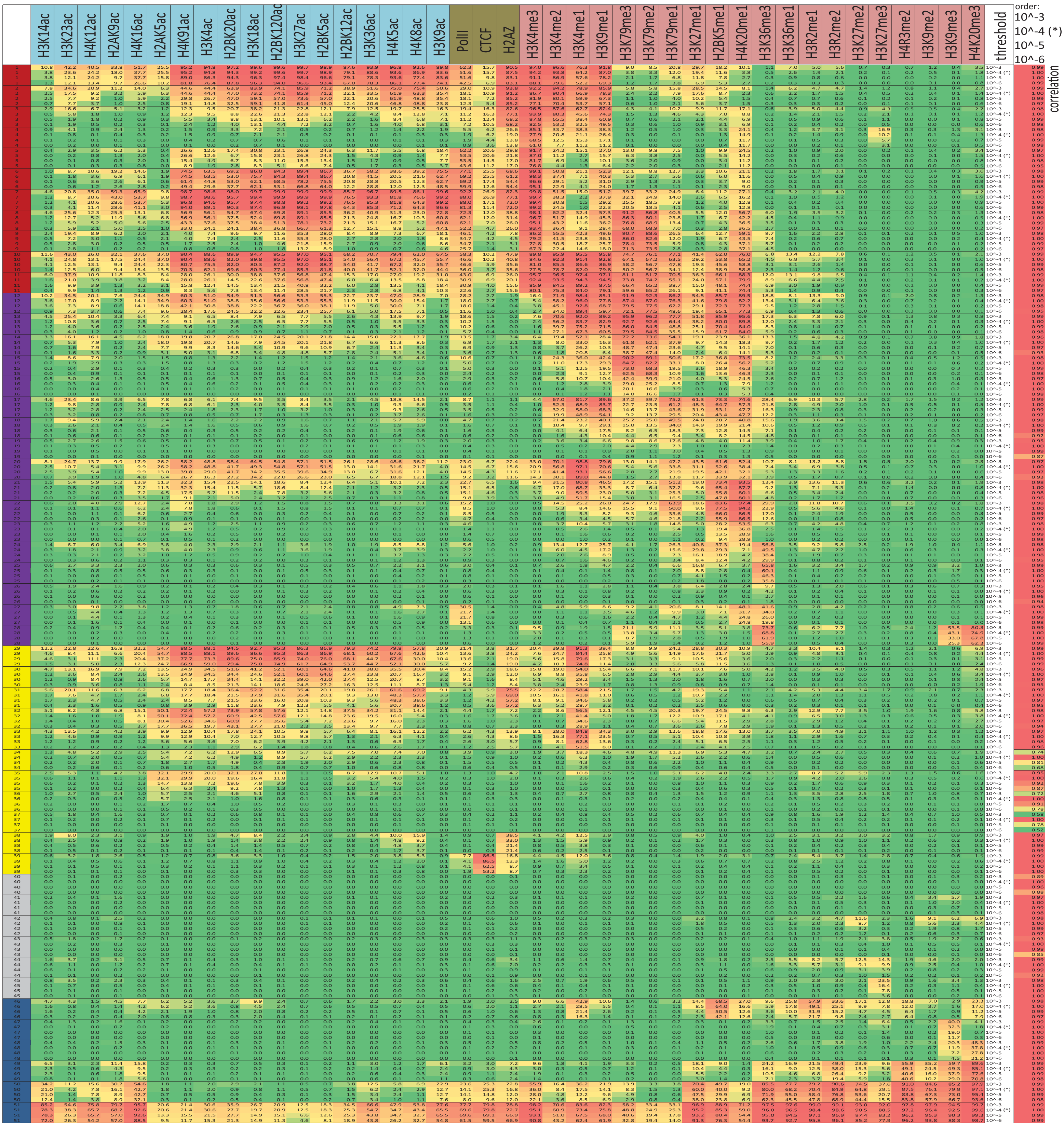
Supplementary Figure 15: Recovery of states from 10 random-initialization 51-state models using the nested-initialization 51-state model. Each row shows the emission parameters for all 51 states of each of 10 randomly-initialized models, clustered together with the 51 states of our nested-initialization model (boxed and labeled with their state ID). Rows are ordered using the optimal leaf ordering clustering method[5], with the matrix split in two halves for display. The figure shows that the states of the model analyzed here are recovered in multiple other random initializations, and that the 51-state model has good coverage of the states found in other random initializations.



Supplementary Figure 16: **Percent genome coverage across chromatin states**. Pie chart showing the portion of the genome assigned to each chromatin state, colored by group.

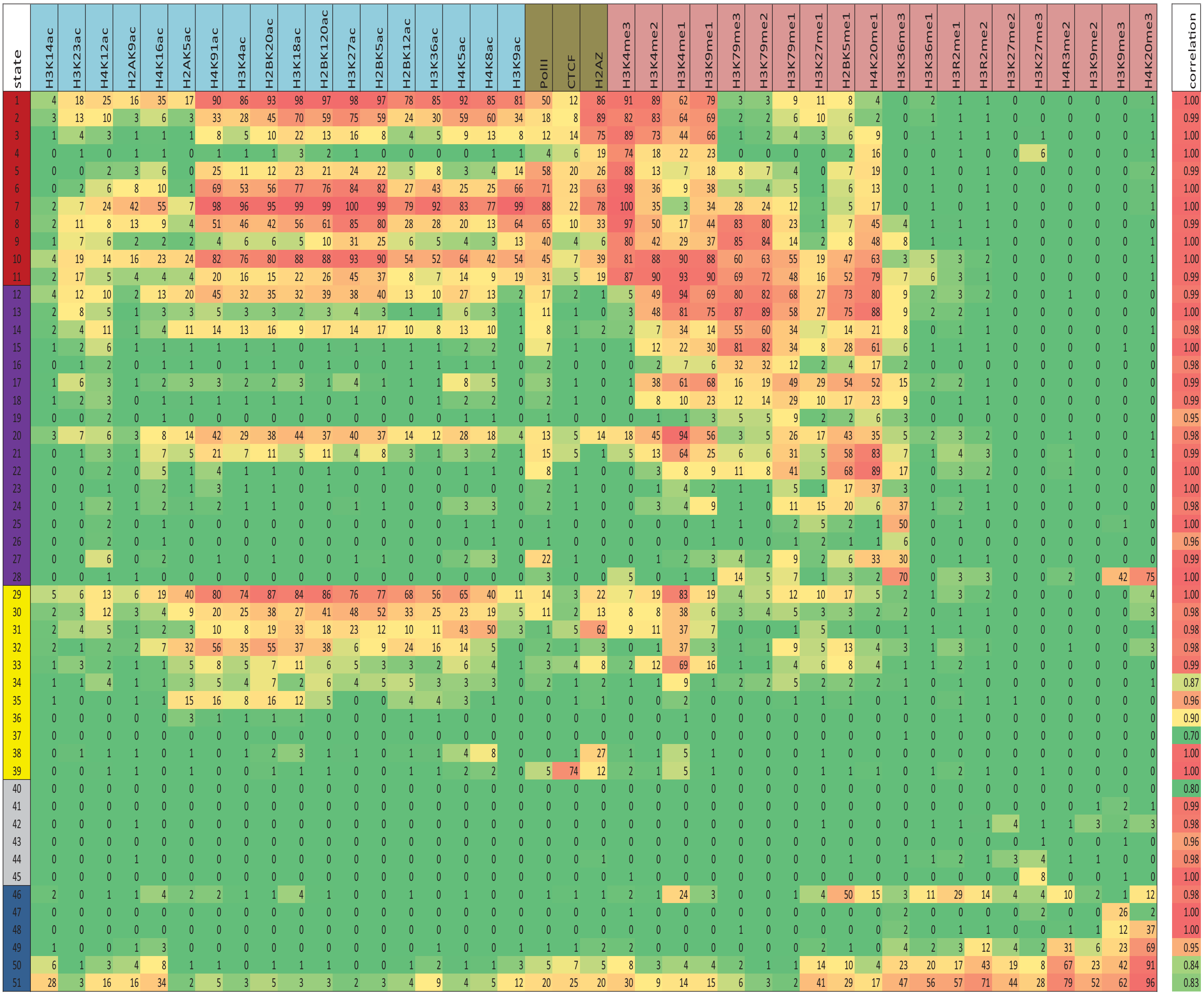


Supplementary Figure 17: Chromatin state emission vector distances visualized using Multi-Dimensional Scaling (MDS). Relative distances between chromatin states projected into a 2-dimensional space based on a multi-dimensional scaling (MDS) approach (implemented in the Matlab cmdscale function). Distances are measured as 1 minus the standard pairwise correlation coefficient between the vectors of emission parameters for each pair of chromatin states. Figure shows that the states capture largely distinct areas of the emission space, and reveal groupings that are largely consistent with the biological interpretation of the functional associations of each state.



(*) = model was learned at 10^-4 cutoff, hence correlations with emission vector at 10^-4 are by definition 1.0

Supplementary Figure 18: Robustness of chromatin states to mark detection thresholds. Each row shows the resulting frequency of each mark in each state, at varying thresholds for the Poisson distribution cut-off. For the model and state assignments inferred at the 10^{-4} cutoff, we evaluate the percentage of the 200bp intervals assigned to each state that would be called ‘present’ at 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} cutoffs, and computed the correlation to the emission vector for each cutoff (since the model was learned at the 10^{-4} cutoff, the emission matrix is by definition the frequency with which marks are observed above the cutoff, and thus the correlation is always 1.0). The high correlation with all other cutoffs indicates that the chromatin mark combinations learned are robust at several different thresholds across three orders of magnitude in the probability cutoff. Even for states with very low emission frequencies (e.g. states 40-45), the correlation remains surprisingly high. The only exceptions are the three Alu-associated intergenic states (states 34, 36, 37), perhaps because acetylation marks that were most associated with these states were sequenced less deeply, coupled with the overall low frequency of most marks in these states making them more sensitive to such fluctuations in acetylation sequencing depth.



state	H3K14ac	H3K23ac	H4K12ac	H2AK9ac	H4K16ac	H2AK5ac	H4K91ac	H3K4ac	H2BK20ac	H3K18ac	H2BK120ac	H3K27ac	H2BK5ac	H2BK12ac	H3K36ac	H4K5ac	H4K8ac	H3K9ac	PolII	CTCF	H2AZ	H3K4me3	H3K4me2	H3K4me1	H3K9me1	H3K79me3	H3K79me2	H3K79me1	H3K27me1	H2BK5me1	H4K20me1	H3K36me3	H3K36me1	H3R2me1	H3R2me2	H3K27me2	H3K27me3	H4R3me2	H3K9me2	H3K9me3	H4K20me3	correlation
1	4	18	25	16	35	17	90	86	93	98	97	98	97	78	85	92	85	81	50	12	86	91	89	62	79	3	3	9	11	8	4	0	2	1	1	0	0	0	0	0	1	1.00
2	3	13	10	3	6	3	33	28	45	70	59	75	59	24	30	59	60	34	18	8	89	82	83	64	69	2	2	6	10	6	2	0	1	1	1	0	0	0	0	0	1	0.99
3	1	4	3	1	1	1	8	5	10	22	13	16	8	4	5	9	13	8	12	14	75	89	73	44	66	1	2	4	3	6	9	0	1	1	1	0	1	0	0	0	1	1.00
4	0	1	0	1	1	0	1	1	1	3	2	1	0	0	0	0	1	1	4	6	19	74	18	22	23	0	0	0	0	2	16	0	0	1	0	0	6	0	0	0	1	1.00
5	0	0	2	3	6	0	25	11	12	23	21	24	22	5	8	3	4	14	58	20	26	88	13	7	18	8	7	4	0	7	19	0	1	0	1	0	0	0	0	0	2	0.99
6	0	2	6	8	10	1	69	53	56	77	76	84	82	27	43	25	25	66	71	23	63	98	36	9	38	5	4	5	1	6	13	0	1	0	1	0	0	0	0	0	1	1.00
7	2	7	24	42	55	7	98	96	95	99	99	100	99	79	92	83	77	99	88	22	78	100	35	3	34	28	24	12	1	5	17	0	1	0	1	0	0	0	0	0	1	1.00
8	2	11	8	13	9	4	51	46	42	56	61	85	80	28	28	20	13	64	65	10	33	97	50	17	44	83	80	23	1	7	45	4	1	1	1	0	0	0	0	0	0	0.99
9	1	7	6	2	2	2	4	6	6	5	10	31	25	6	5	4	3	13	40	4	6	80	42	29	37	85	84	14	2	8	48	8	1	1	1	0	0	0	0	0	1	1.00
10	4	19	14	16	23	24	82	76	80	88	88	93	90	54	52	64	42	54	45	7	39	81	88	90	88	60	63	55	19	47	63	3	5	3	2	0	0	0	0	0	1	1.00
11	2	17	5	4	4	4	20	16	15	22	26	45	37	8	7	14	9	19	31	5	19	87	90	93	90	69	72	48	16	52	79	7	6	3	1	0	0	0	0	0	1	0.99
12	4	12	10	2	13	20	45	32	35	32	39	38	40	13	10	27	13	2	17	2	1	5	49	94	69	80	82	68	27	73	80	9	2	3	2	0	0	1	0	0	0	0.99
13	2	8	5	1	3	3	5	3	3	2	3	4	3	1	1	6	3	1	11	1	0	3	48	81	75	87	89	58	27	75	88	9	2	2	1	0	0	0	0	0	0	1.00
14	2	4	11	1	4	11	14	13	16	9	17	14	17	10	8	13	10	1	8	1	2	2	7	34	14	55	60	34	7	14	21	8	0	1	1	0	0	0	0	0	1	0.98
15	1	2	6	1	1	1	1	1	1	0	1	1	1	1	1	2	2	0	7	1	0	1	12	22	30	81	82	34	8	28	61	6	1	1	1	0	0	0	0	1	0	1.00
16	0	1	2	0	1	1	0	1	1	0	1	0	0	1	1	1	1	0	2	0	0	0	2	7	6	32	32	12	2	4	17	2	0	0	0	0	0	0	0	0	0	0.98
17	1	6	3	1	2	3	3	2	2	3	1	4	1	1	1	8	5	0	3	1	0	1	38	61	68	16	19	49	29	54	52	15	2	2	1	0	0	0	0	0	0	0.99
18	1	2	3	0	1	1	1	1	1	1	0	1	0	0	1	2	2	0	2	1	0	0	8	10	23	12	14	29	10	17	23	9	0	1	1	0	0	0	0	0	0	0.99
19	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	1	3	5	5	9	2	2	6	3	0	0	0	0	0	0	0	0	0	0.95
20	3	7	6	3	8	14	42	29	38	44	37	40	37	14	12	28	18	4	13	5	14	18	45	94	56	3	5	26	17	43	35	5	2	3	2	0	0	1	0	0	1	0.98
21	0	1	3	1	7	5	21	7	11	5	11	4	8	3	1	3	2	1	15	5	1	5	13	64	25	6	6	31	5	58	83	7	1	4	3	0	0	1	0	0	1	0.99
22	0	0	2	0	5	1	4	1	1	0	1	0	1	0	0	1	1	0	8	1	0	0	3	8	9	11	8	41	5	68	89	17	0	3	2	0	0	1	0	0	0	1.00
23	0	0	1	0	2	1	3	1	1	0	1	0	0	0	0	0	0	0	2	1	0	0	1	4	2	1	1	5	1	17	37	3	0	1	1	0	0	1	0	0	0	1.00
24	0	1	2	1	2	1	2	1	1	0	0	1	1	0	0	3	3	0	2	1	0	0	3	4	9	1	0	11	15	20	6	37	1	2	1	0	0	0	0	0	0	0.98
25	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	1	0	2	5	2	1	50	0	1	0	0	0	0	0	1	0	1.00
26	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	2	1	1	6	0	0	0	0	0	0	0	0	0	0.96
27	0	0	6	0	2	0	1	0	1	0	0	1	1	0	0	2	3	0	22	1	0	0	1	2	3	4	2	9	2	6	33	30	0	1	1	0	0	0	0	0	0	0.99
28	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	5	0	1	1	14	5	7	1	3	2	70	0	3	3	0	0	2	0	42	75	1.00
29	5	6	13	6	19	40	80	74	87	84	86	76	77	68	56	65	40	11	14	3	22	7	19	83	19	4	5	12	10	17	5	2	1	3	2	0	0	0	0	0	4	1.00
30	2	3	12	3	4	9	20	25	38	27	41	48	52	33	25	23	19	5	11	2	13	8	8	38	6	3	4	5	3	3	2	2	0	1	1	0	0	0	0	0	3	0.98
31	2	4	5	1	2	3	10	8	19	33	18	23	12	10	11	43	50	3	1	5	62	9	11	37	7	0	0	1	5	1	0	1	1	1	0	0	0	0	0	0	1	0.98
32	2	1	2	2	7	32	56	35	55	37	38	6	9	24	16	14	5	0	2	1	3	0	1	37	3	1	1	9	5	13	4	3	1	3	1	0	0	1	0	0	3	0.98
33	1	3	2	1	1	5	8	5	7	11	6	5	3	3	2	6	4	1	3	4	8	2	12	69	16	1	1	4	6	8	4	1	1	2	1	0	0	0	0	0	0	0.99
34	1	1	4	1	1	3	5	4	7	2	6	4	5	5	3	3	3	0	2	1	2	1	1	9	1	2	2	5	2	2	2	1	0	1	0	0	0	0	0	0	1	0.87
35	1	0	0	1	1	15	16	8	16	12	5	0	0	4	4	3	1	0	0	1	1	0	0	2	0	0	0	1	1	1	0	1	0	1	1	1	0	0	0	0	0	0.96
36	0	0	0	0	0	3	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.90
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0.70
38	0	1	1	1	0	1	0	1	2	3	1	1	0	1	1	4	8	0	0	1	27	1	1	5	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1.00
39	0	0	1	1	0	1	0	0	1	1	1	0	0	1	1	2	2	0	5	74	12	2	1	5	1	0	0	0	1	1	1	0	1	2	1	0	1	0	0	0	0	1.00
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.80
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0.99
42	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	4	1	1	3	2	3	0.98
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0.96
44	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	2	1	3	4	1	1	0	0	0.98
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	1	0	1.00
46	2	0	1	1	4	2	2	1	1	4	1	0	0	1	0	1	0	0	1	1	1	2	1	24	3	0	0	1	4	50	15	3	11	29	14	4	4	10	2	1	12	0.98
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	0	0	26	2	1.00
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	1	1	0	0	1	1	12	37	1.00
49	1	0	0	1	3	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	2	2	0	0	0	0	0	0	2	1	0	4	2	3	12	4	2	31	6	23	69	0.95
50	6	1	3	4	8	1	1	0	1	1	1	0	0	1	2	1	1	3	5	7	5	8	3	4	4	2	1	1	14	10	4	23	20	17	43	19	8	67	23	42	91	0.84
51	28	3	16	16	34	2	5	3	5	3	3	2	3	4	9	4	5	12	20	25	20	30	9	14	15	6	3	2	41	29	17	47	56	57	71	44	28	79	52	62	96	0.83

Supplementary Figure 19: Correlation of mark presence calls with background model based on nucleosome density. This figure shows for each mark the percentage of each state in which it would be considered detected if the background Poisson model was computed based on the local nucleosome tag density[1] in a 1kb window centered at each window (**Supplementary Notes**). The resulting mark-presence frequency calls are highly correlated with the emission matrix computed without adjusting for nucleosome read counts (correlation values for each state shown in the right-most column). All correlations are above 0.95, except for state 40 (unmappable), states 50 and 51 (repetitive with strong mapping bias), and the Alu-associated intergenic states (states 34, 36, 37), showing that the states are largely robust with or without input signal corrections.

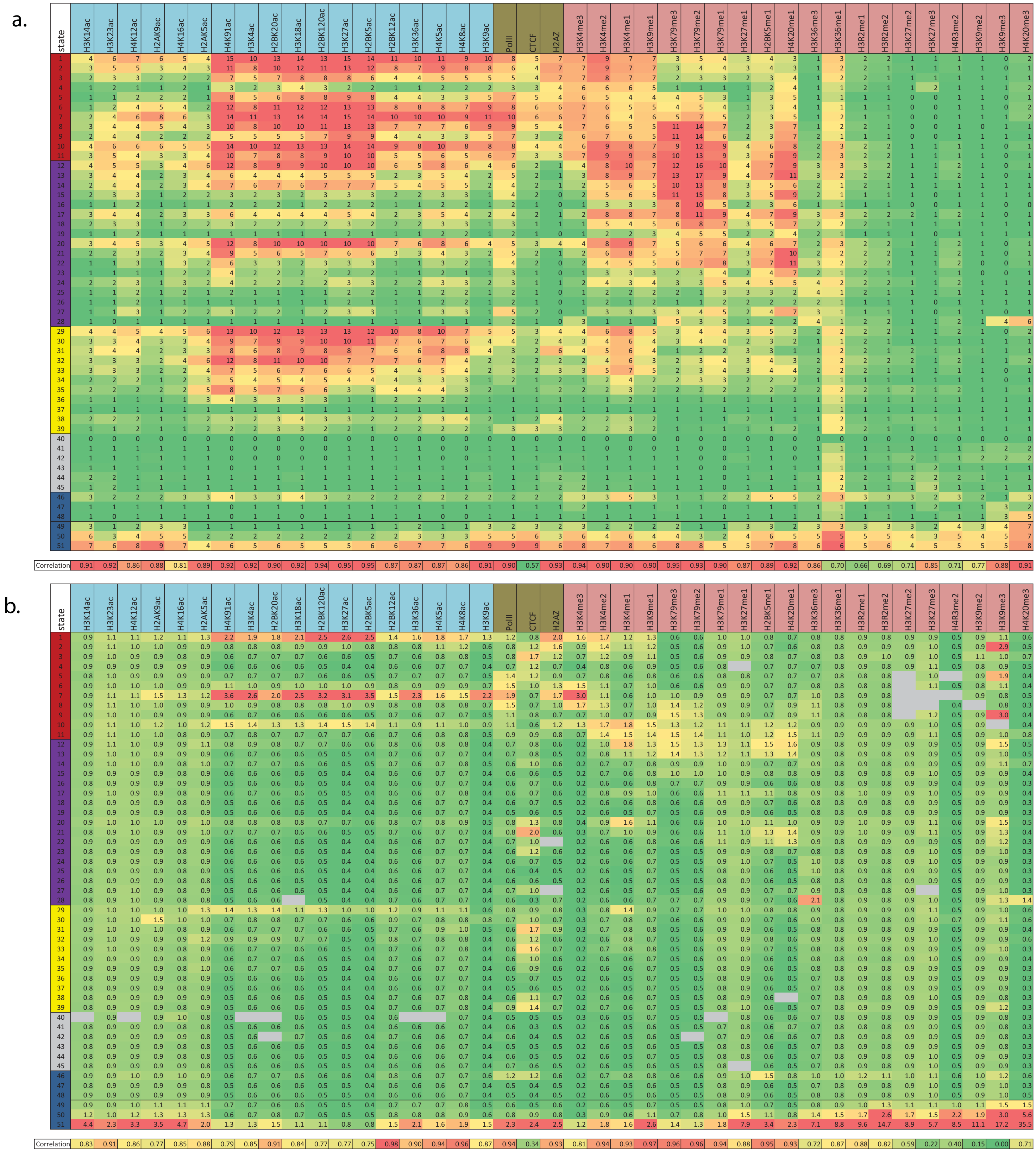
a.

state	H3K14ac	H3K23ac	H4K12ac	H2AK9ac	H4K16ac	H2AK5ac	H4K91ac	H3K4ac	H2BK20ac	H3K18ac	H2BK120ac	H3K27ac	H2BK5ac	H2BK12ac	H3K36ac	H4K5ac	H4K8ac	H3K9ac	PolII	CTCF	H2AZ	H3K4me3	H3K4me2	H3K4me1	H3K9me1	H3K79me3	H3K79me2	H3K79me1	H3K27me1	H2BK5me1	H4K20me1	H3K36me3	H3K36me1	H3R2me1	H3R2me2	H3K27me2	H3K27me3	H4R3me2	H3K9me2	H3K9me3	H4K20me3	IgG
1	4	10	10	10	9	8	62	40	45	57	70	88	93	33	38	41	33	44	22	9	41	56	33	16	16	2	2	4	4	4	2	1	3	2	2	1	1	1	1	0	1	3.0
2	4	8	6	4	4	3	13	11	14	21	20	30	23	10	11	19	19	12	8	9	33	29	27	15	14	1	2	3	4	3	2	1	2	2	2	1	1	1	1	0	1	2.3
3	2	4	3	2	2	2	5	4	6	8	7	7	5	4	4	5	6	5	6	17	22	26	20	9	12	1	1	3	2	3	3	1	2	2	1	1	1	1	1	0	1	2.4
4	1	2	1	1	1	1	2	2	2	3	2	1	1	1	1	1	1	2	2	5	6	13	6	5	5	1	1	1	0	1	4	1	1	2	1	1	3	1	0	1	1	2.3
5	1	1	2	2	3	1	8	5	5	7	7	8	8	3	4	3	3	5	25	16	7	24	5	2	4	2	3	2	1	2	4	0	1	1	1	0	0	1	0	0	1	3.8
6	1	3	4	6	4	2	26	16	16	22	24	32	32	10	13	9	9	20	34	16	19	56	11	3	8	2	3	3	1	2	3	1	1	1	1	0	0	1	1	0	1	3.9
7	3	5	9	21	13	4	107	57	51	69	91	107	134	34	56	34	27	85	52	13	31	112	12	2	7	9	8	5	2	3	4	1	2	1	1	0	0	1	1	1	2	4.6
8	3	6	5	7	4	3	19	14	12	15	18	32	30	10	9	7	6	20	31	7	9	60	16	4	8	29	27	7	2	3	10	2	2	2	1	0	0	1	0	0	1	3.2
9	2	5	4	2	2	2	3	4	4	3	5	10	8	4	3	3	3	5	14	4	2	20	11	5	7	31	31	5	2	3	10	2	1	1	1	0	0	1	0	0	1	2.0
10	4	10	7	10	7	10	41	29	29	33	39	50	51	20	17	20	14	22	18	6	13	41	32	32	19	22	23	14	5	13	17	2	3	3	2	1	1	1	1	0	1	2.7
11	3	9	4	4	3	4	9	8	7	8	10	16	13	5	5	7	5	8	12	5	6	24	28	28	18	29	29	12	5	14	25	2	3	2	2	1	1	1	1	0	1	2.2
12	4	8	6	3	5	9	17	12	12	11	14	14	15	7	6	10	7	3	7	3	1	2	14	32	15	32	32	17	7	21	27	3	3	3	2	1	1	2	1	0	1	1.7
13	3	6	4	2	3	4	4	4	3	3	4	4	3	2	2	4	4	2	5	2	0	2	12	18	14	33	34	14	6	19	26	3	3	2	2	1	1	2	1	0	1	1.4
14	2	4	5	2	3	6	6	6	7	5	7	6	7	5	4	6	5	2	4	2	1	1	4	7	4	15	17	9	3	4	5	2	1	2	1	1	1	1	1	0	1	1.4
15	2	3	4	1	2	2	1	2	2	1	2	2	2	2	2	2	2	1	4	1	0	1	5	4	6	22	23	9	3	6	11	2	1	1	1	1	0	1	1	1	0	1.2
16	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	0	0	2	2	2	7	8	4	2	2	3	1	1	1	1	0	0	1	0	0	0	1.0
17	3	6	4	2	3	4	4	3	3	3	3	4	3	2	2	5	4	2	3	1	1	1	11	13	13	6	8	13	7	13	11	4	3	3	2	1	1	2	1	0	1	1.3
18	2	3	3	1	2	2	2	2	2	1	1	2	1	1	2	3	3	1	2	1	0	1	4	3	5	4	5	8	4	5	5	3	2	2	1	1	1	1	1	0	0	1.2
19	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	0	0	2	1	2	2	3	4	2	1	2	1	1	1	1	1	0	1	1	0	0	1.0
20	4	6	5	3	5	8	16	12	13	14	14	15	14	8	7	10	8	4	7	6	5	5	13	30	11	3	4	8	5	11	9	2	3	3	2	1	1	2	1	0	1	1.9
21	2	2	3	2	4	4	9	5	6	4	6	4	5	3	2	4	3	2	7	8	1	2	6	14	6	3	4	9	3	16	26	2	2	3	2	1	1	2	0	0	1	1.7
22	1	1	3	1	4	2	4	2	2	1	2	1	2	1	1	2	2	1	4	2	0	1	3	3	4	4	4	11	3	16	26	4	2	3	2	1	0	2	0	0	1	1.5
23	1	1	1	1	2	2	3	2	2	1	1	1	1	1	1	1	1	1	2	1	0	1	2	2	2	1	1	3	2	5	7	2	1	2	1	1	1	1	0	0	1	1.3
24	2	2	3	1	3	2	2	2	2	1	1	2	2	1	2	3	3	1	2	1	0	1	3	2	4	1	1	5	5	5	3	7	2	2	2	1	1	2	1	0	1	1.3
25	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	2	2	1	2	1	0	1	1	1	2	1	1	2	3	2	1	8	1	2	1	1	1	1	1	1	1	1.2
26	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	2	2	1	1	2	1	1	1	1	0	1	1	0	0	1.0
27	1	1	4	1	2	2	2	1	1	1	1	2	2	1	1	2	3	1	8	2	0	1	2	1	2	2	2	4	2	3	7	5	1	1	1	1	0	1	0	0	0	1.3
28	1	0	1	1	1	1	0	1	1	0	1	0	0	1	1	1	1	1	2	1	0	2	1	1	1	4	2	3	1	1	1	17	1	2	2	1	1	2	1	10	51	1.2
29	4	5	7	5	6	14	37	26	33	28	36	31	35	25	17	20	13	6	6	3	7	3	7	23	6	2	3	5	4	5	2	1	2	3	2	1	1	2	1	0	2	1.9
30	3	4	5	3	3	5	8	9	11	9	13	15	16	11	8	8	7	4	5	3	4	2	4	7	3	2	2	3	2	2	1	1	1	2	1	1	1	1	1	1	2	1.5
31	3	5	4	2	3	4	6	6	8	11	9	10	7	6	6	13	15	4	2	7	16	3	5	8	4	1	1	2	3	2	1	1	2	2	1	2	1	1	1	0	1	1.5
32	3	3	3	3	4	12	20	13	17	12	14	5	6	10	7	7	4	2	2	2	2	1	3	8	3	1	2	4	3	4	2	2	2	3	2	2	1	2	1	0	2	1.5
33	3	4	3	2	2	4	5	4	5	6	5	4	3	3	3	5	4	2	2	5	3	2	5	13	5	1	1	3	3	3	2	1	2	2	1	1	1	1	1	0	1	1.5
34	2	2	3	1	2	3	4	3	4	3	4	3	3	3	3	3	3	1	2	1	1	1	2	3	2	1	2	3	2	2	1	1	1	1	1	1	1	1	1	0	1	1.2
35	2	2	2	2	2	8	8	6	7	6	5	2	1	4	4	4	2	1	1	1	2	1	1	2	1	1	1	2	2	2	1	1	2	2	2	2	1	2	1	0	1	1.3
36	1	1	1	1	1	3	2	2	2	2	2	1	1	2	2	2	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.1
37	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1.0
38	2	3	2	1	1	2	1	2	2	3	2	2	1	2	2	3	4	1	1	1	7	1	2	2	1	0	0	1	1	1	0	1	2	1	1	1	1	1	1	1	1	1.2
39	1	1	2	1	1	2	1	1	2	2	1	1	1	2	2	2	2	1	3	66	4	1	2	2	1	0	1	1	1	1	1	1	2	2	1	1	1	1	1	0	1	1.9
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2
41	1	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	2	1	1.1
42	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	1	1	1	2	1	1	2	2	2	1.1
43	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	1	0	0	1	1	1	1	1	2	1	1	1	1	1.0
44	2	2	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1	2	2	1	2	3	2	2	1	1	1.1
45	1	2	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	2	1	1	1	4	1	1	1	1	1.2
46	3	2	2	2	3	3	3	2	2	3	2	1	1	1	2	2	2	2	1	2	1	2	2	6	2	1	1	1	3	17	5	2	5	7	5	3	3	4	2	1	6	1.7
47	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0	0	0	1	0	0	1	0	0	1	1	1	1	1	2	1	1	6	1	1.1
48	1	0	1	1	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0	0	0	1	0	0	1	0	0	1	1	1	1	1	1	1	1	4	11	1.3
49	3	1	2	4	4	1	0	1	1	1	1	0	1	1	2	1	1	3	2	3	2	2	1	1	1	1	0	0	3	2	1	4	4	3	8	4	2	14	5	12	62	9.2
50	9	3	5	10	11	2	1	2	2	1	1	1	1	2	4	2	3	7	7	15	5	6	4	5	4	3	2	2	10	9	3	13	15	13	29	13	6	47	15	35	262	23.7
51	70	21	45	67	64	9	10	11	14	9	8	6	8	14	33	17	21	41	52	103	42	44	25	24	28	18	11	10	72	50	29	76	107	96	187	83	52	198	105	227	1696	67.0
Correlation	0.93	0.98	0.94	0.99	0.96	0.99	0.85	0.92	0.96	0.92	0.85	0.84	0.81	0.99	0.95	0.97	0.98	0.91	0.98	0.99	0.97	0.83	0.98	0.96	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.94	0.84	0.93	0.94	0.85	0.93	0.85	0.84	0.98	0.96	

b.

state	**H3K14ac**	H3K23ac	**H4K12ac**	H2AK9ac	**H4K16ac**	H2AK5ac	H4K91ac	H3K4ac	H2BK20ac	H3K18ac	H2BK120ac	H3K27ac	H2BK5ac	H2BK12ac	H3K36ac	H4K5ac	H4K8ac	H3K9ac	PolII	CTCF	H2AZ	H3K4me3	H3K4me2	H3K4me1	H3K9me1	H3K79me3	H3K79me2	H3K79me1	H3K27me1	H2BK5me1	H4K20me1	H3K36me3	**H3K36me1**	H3R2me1	**H3R2me2**	**H3K27me2**	H3K27me3	**H4R3me2**	**H3K9me2**	H3K9me3	H4K20me3	IgG
1	1	3	3	3	3	3	20	13	15	19	23	29	31	11	12	13	11	14	7	3	13	18	11	5	5	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	3.0
2	2	3	3	2	2	1	6	5	6	9	9	13	10	4	5	8	8	5	3	4	14	12	12	7	6	1	1	1	2	1	1	0	1	1	1	0	0	1	0	0	1	2.3
3	1	2	1	1	1	1	2	2	2	4	3	3	2	2	2	2	3	2	2	7	9	11	8	4	5	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	1	2.4
4	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	2	2	6	3	2	2	0	0	0	0	1	2	0	1	1	0	0	1	0	0	0	1	2.3
5	0	0	1	1	1	0	2	1	1	2	2	2	2	1	1	1	1	1	7	4	2	6	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	3.8
6	0	1	1	1	1	0	7	4	4	6	6	8	8	2	3	2	2	5	9	4	5	14	3	1	2	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	3.9
7	1	1	2	5	3	1	23	13	11	15	20	23	29	7	12	7	6	19	11	3	7	24	3	1	2	2	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0	4.6
8	1	2	1	2	1	1	6	4	4	5	6	10	9	3	3	2	2	6	10	2	3	19	5	1	3	9	8	2	1	1	3	0	1	1	0	0	0	0	0	0	0	3.2
9	1	2	2	1	1	1	1	2	2	2	3	5	4	2	2	1	1	3	7	2	1	10	5	3	3	15	15	3	1	1	5	1	1	1	1	0	0	0	0	0	0	2.0
10	2	4	3	4	3	4	15	11	11	12	15	19	19	8	6	8	5	8	7	2	5	16	12	12	7	8	9	5	2	5	7	1	1	1	1	0	0	1	0	0	0	2.7
11	1	4	2	2	1	2	4	4	3	4	5	7	6	2	2	3	2	4	6	2	3	11	13	13	8	13	13	6	2	6	11	1	2	1	1	0	0	1	0	0	0	2.2
12	2	5	4	2	3	5	10	7	7	6	8	8	9	4	3	6	4	2	4	2	1	1	8	19	9	19	19	10	4	12	16	2	2	2	1	1	0	1	0	0	1	1.7
13	2	4	3	1	2	3	3	2	2	2	2	3	2	1	2	3	2	1	4	1	0	1	8	13	10	23	24	10	4	13	18	2	2	2	1	1	0	1	1	0	0	1.4
14	2	3	4	1	2	4	5	5	5	3	5	4	5	4	3	4	3	1	3	1	1	1	3	5	3	11	12	7	2	3	4	2	1	1	1	1	0	1	1	0	1	1.4
15	1	2	3	1	1	2	1	1	1	1	1	1	1	1	1	2	2	1	3	1	0	1	4	4	5	18	19	7	2	5	9	1	1	1	1	1	0	1	1	0	0	1.2
16	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	0	0	2	2	2	7	8	4	2	2	3	1	1	1	1	0	0	1	1	0	0	1.0
17	2	4	3	1	2	3	3	2	2	2	2	3	2	1	2	4	3	1	2	1	1	1	8	10	9	4	6	10	5	10	8	3	2	2	1	1	0	1	1	0	1	1.3
18	2	2	3	1	2	2	1	1	1	1	1	1	1	1	1	2	2	1	2	1	0	1	3	3	4	4	4	7	3	4	4	2	1	1	1	1	0	1	1	0	0	1.2
19	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	0	0	2	1	2	2	3	4	2	1	2	1	1	1	1	1	0	1	1	0	0	1.0
20	2	3	3	2	2	4	9	6	7	7	7	8	8	4	3	6	4	2	4	3	3	2	7	16	6	1	2	4	3	6	5	1	1	2	1	1	0	1	0	0	1	1.9
21	1	1	2	1	2	3	5	3	3	2	4	2	3	2	1	2	2	1	4	5	0	1	3	8	4	2	2	5	2	9	15	1	1	2	1	0	0	1	0	0	1	1.7
22	1	1	2	1	2	1	3	1	1	1	1	1	1	1	1	1	1	1	3	1	0	1	2	2	3	3	3	7	2	10	17	3	1	2	2	1	0	1	0	0	1	1.5
23	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	0	0	1	2	1	1	1	3	1	4	6	1	1	1	1	1	0	1	0	0	0	1.3
24	1	2	2	1	2	2	2	1	1	1	1	2	1	1	1	2	2	1	2	1	0	1	2	2	3	1	1	4	4	4	2	5	1	2	1	1	0	1	1	0	1	1.3
25	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	0	1	1	1	1	1	1	2	3	2	1	7	1	1	1	1	1	1	1	1	1	1.2
26	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	2	2	1	1	2	1	1	1	1	0	1	1	0	0	1.0
27	1	1	3	1	2	1	1	1	1	1	1	1	2	1	1	2	2	1	6	1	0	0	1	1	2	2	1	3	2	2	5	4	1	1	1	0	0	1	0	0	0	1.3
28	1	0	1	1	1	1	0	0	1	0	0	0	0	1	1	0	1	1	1	1	0	2	1	1	1	3	2	2	1	1	1	15	1	2	2	1	0	1	1	8	44	1.2
29	2	3	4	3	3	8	20	14	18	15	19	16	19	14	9	11	7	3	3	2	4	1	4	12	3	1	1	3	2	3	1	1	1	1	1	1	0	1	1	0	1	1.9
30	2	2	4	2	2	3	5	6	7	6	8	10	11	7	5	5	5	2	3	2	3	1	3	5	2	1	1	2	1	1	1	1	1	1	1	0	0	1	1	0	1	1.5
31	2	3	3	2	2	2	4	4	5	7	6	6	4	4	4	9	10	2	1	4	10	2	4	5	2	0	1	1	2	1	0	1	1	1	1	1	1	1	1	0	1	1.5
32	2	2	2	2	3	8	13	9	12	8	9	3	4	7	5	5	3	1	1	1	1	1	2	6	2	1	1	3	2	3	1	1	1	2	1	1	1	1	1	0	1	1.5
33	2	2	2	1	2	3	4	3	3	4	3	3	2	2	2	3	3	1	2	3	2	1	4	9	3	1	1	2	2	2	1	1	1	1	1	1	1	1	1	0	1	1.5
34	1	2	2	1	2	3	3	3	3	2	3	2	3	3	2	3	2	1	2	1	1	1	1	2	1	1	1	2	2	1	1	1	1	1	1	1	0	1	1	0	1	1.2
35	2	2	1	1	2	6	6	5	6	5	4	1	1	3	3	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	0	1	1.3
36	1	1	1	1	1	3	2	2	2	2	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1.1
37	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1.0
38	2	2	2	1	1	2	1	2	2	2	2	1	1	2	2	3	4	1	1	1	6	1	1	2	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1.2
39	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	2	34	2	1	1	1	1	0	0	0	1	0	0	0	1	1	1	1	1	1	0	0	0	1.9
40	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	0.2
41	1	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	2	1	1.1
42	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0	0	1	1	1	1	2	1	1	2	2	2	1.1
43	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	1	0	0	1	1	1	1	1	2	1	1	1	1	1.0
44	1	1	1	1	1	1	0	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1	2	2	1	2	2	1	1	1	1	1.1
45	1	1	1	1	1	1	0	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	1	1	1	3	1	1	1	1	1.2
46	2	1	1	1	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	4	1	0	0	1	2	10	3	1	3	4	3	2	2	3	1	1	4	1.7
47	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0	0	0	1	0	0	1	0	0	1	1	1	1	1	2	1	1	6	1	1.1
48	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	3	9	1.3
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	1	1	7	9.2
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	2	1	1	11	23.7
51	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1	1	0	0	0	0	0	0	1	1	0	1	2	1	3	1	1	3	2	3	25	67.0
Correlation	0.67	0.81	0.62	0.89	0.67	0.93	0.94	0.95	0.94	0.95	0.92	0.92	0.90	0.91	0.95	0.93	0.91	0.95	0.96	0.94	0.96	0.92	0.96	0.96	0.91	0.95	0.95	0.92	0.87	0.97	0.96	0.92	0.71	0.86	0.80	0.79	0.85	0.68	0.72	0.98	0.95	

Supplementary Figure 20: Sequence tag enrichments relative to genome average, and relative to input control. a. Sequence tag enrichments relative to genome average. Table shows the fold enrichment for the sequence tag counts of each mark (top) in each state (left) relative to the genome average for that mark (column). Rightmost column shows the same tag enrichments for IgG control[6], showing that repetitive states 49-51 are likely due to sequencing biases (discussed below and in the text). The bottom row shows the correlation of the mark tag enrichment with the emission parameter columns for the remaining states 1-48. **b. Sequence tag enrichments relative to IgG input control**. Table shows the fold enrichment for the number of sequence tags for each mark (top) in each state (left) relative to the IgG control sequence tag enrichments in that state from panel **a** (also repeated in the rightmost column). All marks showed correlations above 0.85 except a set of nine marks (H4K12ac, H4K16ac, H3K14ac, H4R3me2, H3K36me1, H3K9me2, H3K27me2, H3R2me2, H3K23ac). These nine marks generally appear to have been less informative in our state definitions, and may correspond to either lower quality antibodies, or to less biologically informative chromatin modifications. This table shows that even though we did not seek to explicitly model tag enrichments associated with different chromatin states, such differences are captured by the model. The table also shows that repeat-associated states 49-51 had the highest enrichment for the IgG control (9.2-fold, 23.7-fold, and 67-fold), while only a subset of the high-emission marks were actually enriched relative to the IgG control for these states, suggesting that many of the sequence tags mapping to location in these states are likely from highly-repetitive genomic sequences underrepresented in the reference genome. Lastly, this table suggests that the joint modeling of marks in chromatin states may enable detection of lower informative marks in the context of all other marks and their genomic locations.



a.

state	H3K14ac	H3K23ac	H4K12ac	H2AK9ac	H4K16ac	H2AK5ac	H4K91ac	H3K4ac	H2BK20ac	H3K18ac	H2BK120ac	H3K27ac	H2BK5ac	H2BK12ac	H3K36ac	H4K5ac	H4K8ac	H3K9ac	PolII	CTCF	H2AZ	H3K4me3	H3K4me2	H3K4me1	H3K9me1	H3K79me3	H3K79me2	H3K79me1	H3K27me1	H2BK5me1	H4K20me1	H3K36me3	H3K36me1	H3R2me1	H3R2me2	H3K27me2	H3K27me3	H4R3me2	H3K9me2	H3K9me3	H4K20me3
1	4	6	7	6	5	4	15	10	13	14	13	15	14	11	10	11	9	10	8	5	7	7	9	7	7	3	5	4	3	4	3	1	3	2	2	1	1	1	1	0	2
2	3	5	5	3	4	3	11	8	10	12	11	13	12	8	7	9	8	8	6	4	7	7	9	7	7	3	4	4	3	4	3	1	2	2	1	1	1	1	1	0	1
3	2	3	3	2	2	2	7	5	7	8	8	8	6	4	4	5	5	5	5	4	7	7	8	7	7	2	3	3	2	3	4	1	2	2	1	1	1	1	1	1	1
4	1	2	1	1	2	1	3	2	3	4	3	2	2	1	1	2	2	2	3	3	4	6	6	6	5	1	1	1	1	2	4	1	1	2	1	1	2	1	1	1	2
5	1	1	2	2	2	1	8	5	6	8	8	9	8	4	4	3	3	5	7	5	4	6	5	4	4	4	5	3	1	3	5	1	1	1	1	0	0	1	1	0	1
6	1	2	4	5	4	2	12	8	11	12	12	13	13	8	8	8	7	9	8	6	6	7	6	5	5	4	5	4	1	3	4	1	1	1	1	0	0	1	1	0	2
7	2	4	6	8	6	3	14	11	13	14	14	15	14	10	10	10	9	11	10	6	6	7	6	4	6	5	7	5	2	3	5	1	2	1	1	0	0	1	1	1	2
8	3	4	4	5	4	3	10	8	10	10	11	13	13	7	7	7	6	9	9	5	4	7	6	5	5	11	14	7	2	3	7	1	2	2	1	0	0	1	0	1	1
9	2	4	4	2	2	2	5	5	5	5	7	9	9	4	4	3	3	5	7	3	2	6	7	6	5	11	14	6	2	3	7	2	1	1	1	0	0	1	1	1	1
10	4	6	6	6	5	5	14	10	12	13	13	14	14	9	8	10	8	8	8	4	4	6	9	8	7	9	12	9	4	6	8	2	3	2	2	1	1	1	1	0	1
11	3	5	4	3	3	4	10	7	8	8	9	10	10	5	5	6	5	6	7	3	3	7	9	9	8	10	13	9	3	6	9	2	3	2	2	1	1	1	1	0	1
12	4	5	5	3	4	6	12	8	9	9	10	10	10	6	5	8	6	4	6	2	1	4	8	10	7	12	16	10	4	7	9	3	3	3	2	1	1	2	1	0	1
13	3	4	4	2	3	4	6	4	4	4	5	5	5	2	3	5	4	2	5	2	1	3	8	9	7	13	17	9	4	7	11	3	2	2	2	1	1	2	1	0	1
14	2	4	4	2	3	4	7	6	7	6	7	7	7	5	4	5	5	2	4	2	1	2	5	6	5	10	13	8	3	5	6	2	1	2	1	1	1	1	1	1	1
15	2	3	3	1	2	2	3	2	3	2	3	3	3	2	2	3	3	1	4	2	0	2	5	6	5	11	15	8	3	5	9	2	1	1	1	1	0	1	1	1	1
16	1	1	2	1	1	1	2	2	2	1	2	2	2	1	1	2	2	1	2	1	0	1	3	3	3	8	10	5	2	3	6	1	1	1	1	0	0	1	1	0	0
17	3	4	4	2	3	3	6	4	4	4	4	5	4	2	3	5	4	2	4	2	1	2	8	8	7	8	11	9	4	7	9	3	3	2	2	2	1	2	1	0	1
18	2	3	3	1	2	2	3	2	2	2	2	3	2	2	2	3	3	1	3	1	1	1	5	4	5	6	8	7	3	5	7	2	2	2	1	1	1	1	1	1	1
19	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	2	1	0	1	2	2	3	4	5	5	2	2	4	2	1	1	1	1	0	1	1	1	1
20	3	4	5	3	4	5	12	8	10	10	10	10	10	7	6	8	6	4	5	3	4	4	8	9	7	5	6	6	4	6	7	2	3	2	2	1	1	2	1	0	1
21	2	2	3	2	3	4	9	5	6	5	7	6	6	3	3	4	3	2	5	2	1	2	6	8	5	5	7	7	3	7	10	2	2	2	2	1	1	2	1	0	1
22	1	1	3	1	3	2	6	3	3	2	3	3	3	1	1	3	2	1	4	2	0	2	4	5	5	6	7	8	3	7	11	3	2	2	2	1	1	2	0	0	1
23	1	1	1	1	2	2	4	2	2	2	2	2	2	1	1	2	1	1	3	1	0	1	3	3	3	2	3	4	2	4	7	2	1	2	1	1	1	1	0	0	1
24	2	2	3	1	3	2	4	2	2	2	2	3	3	2	2	3	3	2	3	1	1	2	4	3	4	3	3	5	4	5	5	4	2	2	2	1	1	2	1	1	1
25	1	1	2	1	2	1	2	1	1	1	1	2	2	1	1	2	2	1	2	1	0	1	2	2	2	2	1	3	3	3	2	4	1	2	1	1	1	1	1	1	1
26	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	2	1	0	1	2	1	2	1	1	2	2	2	2	2	1	1	1	1	0	1	1	1	1
27	1	1	3	1	2	2	3	2	2	1	2	3	3	1	1	3	3	1	5	2	0	1	3	3	3	3	4	5	2	4	7	3	1	1	1	1	0	1	1	0	1
28	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	0	3	1	1	1	5	3	3	1	2	2	4	1	2	2	1	1	2	1	4	6
29	4	4	5	4	5	6	13	10	12	13	13	13	12	10	8	10	7	5	5	3	4	4	6	8	5	3	4	4	3	5	3	2	2	2	2	1	1	2	1	0	2
30	3	3	4	3	3	4	9	7	9	9	10	10	11	7	6	7	6	4	4	2	4	3	4	6	4	3	4	4	2	3	2	1	1	2	1	1	1	1	1	1	1
31	3	4	4	2	3	3	8	6	8	9	8	8	7	6	6	8	8	4	3	2	6	4	5	6	4	1	2	2	3	3	1	1	2	2	1	2	1	1	1	1	1
32	3	3	3	2	4	6	12	8	11	10	10	7	7	7	6	7	4	2	2	2	2	2	4	6	3	2	3	4	3	4	3	2	2	2	2	2	1	2	1	1	2
33	3	3	3	2	2	4	7	5	6	7	6	6	5	4	4	5	4	2	3	2	3	3	5	7	5	2	3	4	3	4	4	2	2	2	1	1	1	1	1	0	1
34	2	2	2	1	2	3	5	4	5	4	5	4	4	4	3	3	3	1	2	1	2	1	2	4	2	2	3	3	2	2	2	1	1	1	1	1	1	1	1	0	1
35	2	2	2	2	2	5	8	5	7	6	6	3	3	4	4	4	3	2	1	1	2	2	2	3	2	2	2	2	2	2	2	2	2	2	1	2	1	2	1	1	1
36	1	1	1	1	1	3	4	3	3	3	3	1	1	2	2	2	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
38	2	2	2	1	1	2	3	2	3	4	3	3	2	2	2	3	4	2	1	2	4	2	2	3	2	1	1	1	2	1	1	1	2	1	1	1	1	1	1	1	1
39	1	1	2	1	1	2	2	2	3	3	2	2	1	2	2	2	2	1	3	3	3	2	2	3	2	1	1	1	1	1	2	1	2	2	1	1	1	1	1	1	1
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	0	0	0	1	0	0	1	1	0	1	1	1	1	1	1	1	1	2	2
42	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	2	1	1	2	2	2
43	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	2	1	1	1	1
44	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	0	1	1	1	1	2	2	1	2	2	2	1	1	1
45	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	0	1	1	1	1	2	1	1	1	3	1	1	1	1
46	3	2	2	2	3	3	4	3	3	4	3	2	2	2	2	2	2	2	2	2	2	3	3	5	3	1	1	2	2	5	5	2	3	3	3	3	2	3	2	1	3
47	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	0	1	1	1	1	1	1	1	1	2	1	1	3	2
48	1	0	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	5
49	3	1	2	3	3	1	1	1	1	1	1	1	1	1	2	1	1	3	2	3	2	3	2	2	2	2	1	1	3	3	2	3	3	3	3	3	2	4	3	4	7
50	6	3	4	5	5	2	2	2	2	2	2	2	2	2	3	3	3	5	5	6	3	6	4	6	4	4	3	2	4	6	5	5	5	5	5	4	3	5	4	5	7
51	7	6	8	9	7	4	6	5	6	5	5	5	6	6	7	7	6	9	9	9	6	8	7	8	6	8	8	5	5	7	8	6	6	5	6	4	5	5	5	5	8

Correlation	0.91	0.92	0.86	0.88	0.81	0.89	0.92	0.92	0.90	0.92	0.94	0.95	0.95	0.87	0.87	0.87	0.86	0.91	0.90	0.57	0.93	0.94	0.90	0.90	0.90	0.95	0.93	0.90	0.87	0.89	0.92	0.86	0.70	0.66	0.69	0.71	0.85	0.71	0.77	0.88	0.91

b.

state	H3K14ac	H3K23ac	H4K12ac	H2AK9ac	H4K16ac	H2AK5ac	H4K91ac	H3K4ac	H2BK20ac	H3K18ac	H2BK120ac	H3K27ac	H2BK5ac	H2BK12ac	H3K36ac	H4K5ac	H4K8ac	H3K9ac	PolII	CTCF	H2AZ	H3K4me3	H3K4me2	H3K4me1	H3K9me1	H3K79me3	H3K79me2	H3K79me1	H3K27me1	H2BK5me1	H4K20me1	H3K36me3	H3K36me1	H3R2me1	H3R2me2	H3K27me2	H3K27me3	H4R3me2	H3K9me2	H3K9me3	H4K20me3
1	0.9	1.1	1.1	1.2	1.1	1.3	2.2	1.9	1.8	2.1	2.5	2.6	2.5	1.4	1.6	1.8	1.7	1.3	1.2	0.8	2.0	1.6	1.7	1.2	1.3	0.6	0.6	1.0	1.0	0.8	0.7	0.8	0.8	0.9	0.8	0.9	0.9	0.5	0.9	1.1	0.6
2	0.9	1.1	1.0	1.0	0.9	0.9	0.8	0.8	0.8	0.9	0.9	1.0	0.8	0.8	0.8	1.1	1.2	0.6	0.8	1.2	1.6	0.9	1.4	1.1	1.2	0.5	0.6	0.9	1.0	0.7	0.6	0.8	0.8	0.9	0.9	0.9	1.0	0.5	0.9	2.9	0.5
3	0.9	1.0	0.9	0.9	0.8	0.9	0.6	0.7	0.7	0.7	0.6	0.6	0.5	0.7	0.6	0.8	0.8	0.5	0.8	1.7	1.2	0.7	1.2	0.9	1.1	0.5	0.6	0.9	0.9	0.8	0.7	0.8	0.8	0.9	1.0	0.9	1.0	0.5	1.1	1.0	0.7
4	0.9	0.9	0.9	0.9	0.9	0.9	0.6	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.7	1.2	0.7	0.4	0.8	0.6	0.9	0.5	0.6	0.8		0.7	0.7	0.7	0.8	0.9	0.8	0.9	1.1	0.5	0.8	0.9	0.5
5	0.8	1.0	0.9	0.9	0.9	0.9	0.7	0.7	0.7	0.7	0.6	0.6	0.5	0.7	0.7	0.7	0.7	0.5	1.4	1.2	0.9	0.7	0.8	0.6	0.8	0.6	0.6	0.8	0.8	0.7	0.7	0.7	0.8	0.8	0.8		1.0		0.9	1.9	0.4
6	0.9	1.0	1.0	1.0	0.9	0.9	1.1	1.0	0.9	1.0	1.0	1.0	0.9	0.8	0.9	0.9	0.9	0.7	1.5	1.0	1.3	1.5	1.1	0.7	1.0	0.6	0.6	0.9	0.9	0.7	0.7	0.8	0.8	0.8	0.8		1.1	0.5	0.8	1.1	0.4
7	0.9	1.1	1.1	1.5	1.3	1.2	3.6	2.6	2.0	2.5	3.2	3.1	3.5	1.5	2.3	1.6	1.5	2.2	1.9	0.7	1.7	3.0	1.1	0.6	1.0	1.0	0.9	0.9	0.9	0.7	0.7	1.0	0.8	0.9	0.8				0.9	0.8	0.5
8	0.9	1.1	1.0	1.0	0.9	0.9	1.0	0.9	0.8	0.8	0.8	1.0	0.9	0.8	0.8	0.8	0.8	0.7	1.5	0.7	1.0	1.7	1.3	0.7	1.0	1.4	1.2	0.9	0.9	0.7	0.9	1.1	0.8	0.9	0.8			0.4		0.8	0.3
9	0.9	1.0	1.0	0.9	0.9	0.9	0.6	0.7	0.6	0.6	0.6	0.6	0.5	0.7	0.6	0.7	0.7	0.5	1.1	0.8	0.7	0.7	1.0	0.7	0.9	1.5	1.3	0.9	0.9	0.7	0.9	1.1	0.8	0.8	0.8		1.2	0.5	0.9	3.0	0.4
10	0.9	1.1	1.0	1.2	1.0	1.2	1.5	1.4	1.3	1.3	1.4	1.5	1.4	1.1	0.9	1.1	1.0	0.9	1.1	0.6	1.2	1.3	1.7	1.8	1.5	1.3	1.2	1.1	1.1	1.2	1.2	0.9	0.9	0.9	0.9	0.9	1.0	0.5	0.9		0.4
11	0.9	1.1	1.0	1.0	0.8	0.9	0.7	0.8	0.7	0.7	0.7	0.7	0.6	0.7	0.6	0.8	0.8	0.5	0.9	0.9	0.8	0.7	1.4	1.5	1.4	1.5	1.4	1.1	1.0	1.2	1.5	1.1	0.9	0.9	0.9	0.9	1.1	0.5	0.9	1.0	0.8
12	0.9	1.1	1.0	0.9	0.9	1.1	0.8	0.9	0.8	0.7	0.7	0.6	0.6	0.8	0.6	0.8	0.8	0.4	0.7	0.8	0.6	0.2	1.0	1.8	1.3	1.5	1.3	1.3	1.1	1.5	1.6	0.9	0.8	0.9	0.9	0.9	0.9	0.5	0.9	1.5	0.5
13	0.9	1.0	1.0	0.9	0.9	0.9	0.6	0.7	0.6	0.6	0.5	0.5	0.4	0.7	0.6	0.7	0.7	0.4	0.7	0.8	0.5	0.2	0.8	1.1	1.2	1.4	1.3	1.2	1.1	1.3	1.4	0.9	0.8	0.9	0.8	0.9	0.9	0.5	0.9	1.0	0.5
14	0.9	1.0	1.0	0.9	0.8	1.0	0.7	0.7	0.7	0.6	0.6	0.5	0.5	0.7	0.6	0.7	0.8	0.5	0.6	1.0	0.6	0.2	0.7	0.7	0.8	0.9	0.9	1.0	0.9	0.7	0.7	0.9	0.8	0.9	0.8	0.9	0.9	0.5	0.9	1.1	0.7
15	0.9	0.9	0.9	0.9	0.8	0.9	0.6	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.6	0.6	0.5	0.2	0.7	0.6	0.8	1.0	1.0	1.0	0.9	0.8	0.8	0.9	0.8	0.9	0.8	0.9	1.0	0.5	0.9	0.9	0.4
16	0.8	0.9	0.9	0.9	0.9	0.9	0.5	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.6	0.7	0.6	0.2	0.6	0.6	0.8	0.7	0.7	0.9	0.9	0.6	0.6	0.8	0.8	0.9	0.8	0.9	0.9	0.4	0.9	0.9	0.3
17	0.9	1.0	0.9	0.9	0.8	0.9	0.6	0.7	0.6	0.6	0.5	0.5	0.4	0.6	0.6	0.7	0.8	0.4	0.6	0.7	0.6	0.2	0.8	0.9	1.0	0.6	0.6	1.1	1.1	1.1	0.8	0.9	0.8	1.0	0.9	0.9	1.0	0.5	0.9	0.9	0.4
18	0.8	0.9	0.9	0.9	0.8	0.9	0.5	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.6	0.7	0.5	0.2	0.6	0.6	0.8	0.6	0.6	0.9	0.9	0.7	0.6	0.8	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.3
19	0.8	0.9	0.9	0.9	0.8	0.9	0.5	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.5	0.6	0.5	0.2	0.6	0.5	0.7	0.5	0.6	0.9	0.9	0.6	0.5	0.8	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.3
20	0.9	1.0	1.0	1.0	0.9	1.0	0.8	0.8	0.8	0.8	0.7	0.7	0.6	0.8	0.7	0.8	0.9	0.5	0.8	1.3	0.8	0.4	0.9	1.6	1.1	0.6	0.6	1.0	1.0	1.1	1.0	0.9	0.9	1.0	0.9	0.9	1.1	0.6	0.9	1.5	0.5
21	0.8	0.9	1.0	0.9	0.9	1.0	0.7	0.7	0.7	0.6	0.6	0.5	0.5	0.7	0.6	0.7	0.7	0.4	0.8	2.0	0.6	0.3	0.7	1.0	0.9	0.6	0.6	1.1	1.0	1.3	1.4	0.9	0.9	1.0	0.9	0.9	1.1	0.5	0.9	1.3	0.4
22	0.9	0.9	0.9	0.9	0.9	0.9	0.6	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.7	1.0		0.2	0.6	0.6	0.8	0.6	0.6	1.1	0.9	1.1	1.3	0.9	0.8	0.9	0.9	0.9	0.9	0.5	0.9	1.2	0.3
23	0.8	0.9	0.9	0.9	0.8	0.9	0.6	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.6	1.2	0.6	0.2	0.6	0.6	0.7	0.5	0.5	0.9	0.9	0.8	0.7	0.8	0.8	0.9	0.8	0.9	1.0	0.5	0.9	1.0	0.3
24	0.8	0.9	0.9	0.9	0.8	0.9	0.6	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.6	0.7	0.5	0.2	0.6	0.6	0.7	0.5	0.5	0.9	1.0	0.7	0.5	1.0	0.8	0.9	0.8	0.9	1.0	0.5	0.9	0.9	0.3
25	0.8	0.9	0.9	0.9	0.8	0.9	0.5	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.6	0.4	0.5	0.2	0.6	0.5	0.7	0.5	0.5	0.8	0.9	0.6	0.5	1.0	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.3
26	0.8	0.9	0.9	0.9	0.8	0.8	0.5	0.6	0.6	0.5	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.6	0.6	0.5	0.2	0.6	0.5	0.7	0.5	0.5	0.8	0.9	0.6	0.5	0.8	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.3
27	0.8	0.9	1.0	0.9	0.8	0.9	0.5	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.7	1.0		0.2	0.6	0.5	0.7	0.5	0.6	0.9	0.9	0.7	0.7	0.9	0.8	0.9	0.8	0.9		0.5	0.9	1.0	0.3
28	0.8	0.9	1.0	0.9	0.8	0.8	0.5	0.6	0.6		0.5	0.4	0.4	0.7	0.6	0.7	0.7	0.4	0.6	0.3	0.7	0.2	0.6	0.6	0.7	0.7	0.7	0.9	0.9	0.7	0.6	2.1	0.8	0.9	0.8	0.9	1.0	0.5	0.9	1.3	1.4
29	0.9	1.0	1.0	1.0	1.0	1.3	1.4	1.3	1.4	1.1	1.3	1.0	1.0	1.2	0.9	1.1	1.1	0.6	0.8	0.9	0.8	0.3	0.8	1.4	0.9	0.7	0.7	1.0	1.0	0.8	0.8	0.9	0.8	0.9	0.9	0.9	1.1	0.5	0.9	1.0	0.6
30	0.9	1.0	1.0	1.5	1.0	1.0	0.7	0.8	0.8	0.7	0.7	0.6	0.6	0.9	0.7	0.8	0.8	0.8	0.7	1.0	0.7	0.3	0.7	0.8	0.7	0.6	0.6	0.9	0.9	0.7	0.6	0.9	0.8	0.9	0.8	0.9	1.0	0.7	0.9	1.1	0.6
31	0.9	1.0	1.0	1.0	0.8	0.9	0.6	0.7	0.7	0.7	0.6	0.6	0.5	0.7	0.6	0.9	1.0	0.5	0.6	1.7	0.9	0.3	0.7	0.8	0.8	0.5	0.6	0.9	0.9	0.6	0.5	0.8	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.4
32	0.9	1.0	0.9	0.9	0.9	1.2	0.9	0.9	0.9	0.7	0.7	0.5	0.5	0.8	0.7	0.8	0.8	0.4	0.6	1.2	0.6	0.2	0.6	0.8	0.7	0.5	0.6	0.9	0.9	0.8	0.6	0.9	0.8	0.9	0.9	0.9	1.0	0.5	0.9	0.9	0.6
33	0.9	1.0	0.9	0.9	0.8	0.9	0.6	0.7	0.7	0.6	0.6	0.5	0.4	0.7	0.6	0.7	0.8	0.4	0.6	1.6	0.7	0.2	0.7	0.8	0.8	0.5	0.6	0.9	0.9	0.7	0.5	0.8	0.8	0.9	0.8	0.9	0.9	0.5	0.9	1.0	0.3
34	0.9	0.9	0.9	0.9	0.8	0.9	0.6	0.7	0.7	0.6	0.5	0.5	0.4	0.7	0.6	0.7	0.7	0.4	0.6	1.0	0.6	0.2	0.6	0.6	0.7	0.5	0.6	0.9	0.9	0.6	0.5	0.8	0.8	0.9	0.8	0.9	1.0	0.5	0.9	0.9	0.4
35	0.9	0.9	0.9	0.9	0.8	1.0	0.6	0.7	0.7	0.6	0.5	0.4	0.4	0.7	0.6	0.7	0.7	0.4	0.6	0.9	0.6	0.2	0.6	0.6	0.7	0.5	0.5	0.8	0.9	0.6	0.5	0.8	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.3
36	0.8	0.9	0.9	0.9	0.8	0.9	0.5	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.5	0.7	0.5	0.2	0.6	0.5	0.7	0.5	0.5	0.8	0.9	0.6	0.5	0.7	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.3
37	0.8	0.9	0.9	0.9	0.8	0.9	0.5	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.6	0.7	0.7	0.4	0.5	0.6	0.5	0.2	0.6	0.5	0.7	0.5	0.5	0.8	0.9	0.6	0.5	0.7	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.9	0.3
38	0.8	0.9	0.9	0.9	0.9	0.9	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.7	0.6	0.7	0.8	0.4	0.6	1.1	0.7	0.2	0.6	0.6	0.7	0.5	0.5	0.8	0.9	0.6		0.7	0.8	0.9	0.8	0.9	1.0	0.5	0.9	0.8	0.3
39	0.8	0.9	0.9	0.9	0.8	0.9	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.7	0.6	0.7	0.8	0.4	0.9	1.4	0.7	0.2	0.7	0.6	0.7	0.5	0.5	0.8	1.0	0.6	0.5	0.7	0.8	0.9	0.9	0.9	0.9	0.5	0.9	1.2	0.3
40		0.9		0.9	1.0	0.8	0.5			0.6	0.5	0.5	0.4	0.6			0.7	0.4	0.5	0.5	0.5	0.3	0.6	0.5	0.7	0.5	0.5		0.8	0.6	0.6	0.7	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.8	0.3
41	0.8	0.9	0.9	0.9	0.8	0.8	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.6	0.6	0.6	0.7	0.4	0.6	0.3	0.5	0.2	0.6	0.5	0.6	0.5	0.5	0.7	0.8	0.6	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.5	0.9	0.8	0.3
42	0.8	0.9	0.9	0.9	0.9	0.8	0.5	0.6		0.7	0.5	0.6	0.4	0.7	0.6	0.6	0.7	0.4	0.6	0.4	0.5	0.2	0.6	0.5	0.6	0.5		0.7	0.9	0.6	0.6	0.7	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.8	0.3
43	0.8	0.9	0.9	0.9	0.8	0.8	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.6	0.6	0.6	0.7	0.4	0.5	0.5	0.5	0.2	0.6	0.5	0.6	0.5	0.5	0.8	0.8	0.6	0.5	0.7	0.8	0.9	0.8	0.9	0.9	0.5	0.9	0.8	0.3
44	0.8	0.9	0.9	0.9	0.8	0.8	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.6	0.6	0.6	0.7	0.4	0.6	0.5	0.5	0.2	0.6	0.5	0.7	0.5	0.5	0.8	0.8	0.6	0.5	0.7	0.8	0.9	0.8	0.9	1.0	0.5	0.9	0.9	0.3
45	0.8	0.9	0.9	0.9	0.9	0.8	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.6	0.6	0.6	0.7	0.4	0.7	0.6	0.5	0.2	0.6	0.5	0.6	0.5	0.5	0.8		0.6	0.5	0.7	0.8	0.9	0.8	0.9	1.0	0.5	0.9	0.9	0.3
46	0.9	0.9	1.0	0.9	1.0	0.9	0.6	0.7	0.7	0.6	0.6	0.6	0.4	0.8	0.6	0.7	0.7	0.6	1.2	1.2	0.6	0.2	0.7	0.8	0.8	0.6	0.6	0.9	1.0	1.5	0.8	1.0	1.0	1.2	1.1	1.0	1.1	0.6	1.0	1.2	0.6
47	0.8	0.9	0.9	0.9	0.9	0.8	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.6	0.6	0.6	0.7	0.4	0.5	0.4	0.5	0.2	0.6	0.5	0.6	0.5	0.5	0.8	0.9	0.6	0.5	0.7	0.8	0.9	0.8	0.9	1.0	0.5	0.9	1.0	0.3
48	0.9	0.9	0.9	0.9	0.9	0.9	0.5	0.6	0.6	0.6	0.5	0.5	0.4	0.6	0.6	0.6	0.7	0.4	0.5	0.4	0.5	0.2	0.6	0.5	0.6	0.5	0.5	0.8	0.9	0.6	0.5	0.8	0.8	0.9	0.8	0.9	1.0	0.5	0.9	1.0	0.5
49	0.9	0.9	1.0	1.1	1.1	1.1	0.7	0.7	0.6	0.6	0.5	0.5	0.4	0.7	0.7	0.7	0.8	0.5	0.6	0.5	0.6	0.2	0.6	0.5	0.7	0.5	0.5	0.8	1.0	0.7	0.5	0.8	0.9	1.0	1.3	1.1	1.1	1.0	1.1	1.5	1.5
50	1.2	1.0	1.2	1.3	1.3	1.3	0.6	0.7	0.8	0.7	0.5	0.5	0.5	0.8	0.8	0.8	0.9	0.6	0.8	0.9	0.8	0.3	0.9	0.6	1.1	0.7	0.8	1.0	1.5	1.1	0.8	1.4	1.5	1.7	2.6	1.7	1.5	2.2	1.9	3.0	5.6
51	4.4	2.3	3.3	3.5	4.7	2.0	1.3	1.3	1.5	1.1	1.1	0.8	0.8	1.5	2.1	1.6	1.9	1.5	2.3	2.4	2.5	1.2	1.8	1.6	2.6	1.4	1.3	1.8	7.9	3.4	2.3	7.1	8.8	9.6	14.7	8.9	5.7	8.5	11.1	17.2	35.5

Correlation	0.83	0.91	0.86	0.77	0.85	0.88	0.79	0.85	0.91	0.84	0.77	0.77	0.75	0.98	0.90	0.94	0.96	0.87	0.94	0.34	0.93	0.81	0.94	0.93	0.97	0.96	0.96	0.94	0.88	0.95	0.93	0.72	0.87	0.88	0.82	0.59	0.22	0.40	0.15	0.00	0.71

Supplementary Figure 21: Chromatin states capture tag intensities outside binary cutoffs. Sequence tag enrichments for intervals above and below detection threshold demonstrate that chromatin states capture additional information about signal intensity beyond the simple presence/absence information encoded in the binary cutoffs. **a. Below cutoff.** Fold enrichment for the sequence tag counts of each mark (top) in each state (left) relative to the genome average for that mark (column), for all intervals where the mark is 'absent' (i.e. number of sequence tags for that mark is below the 'presence' threshold for that mark). The bottom row shows high correlation between the emission parameters of a given mark across states and the tag frequency levels for that mark in 'absent' regions, even though when learning the model no information on the below threshold signal levels was given. **b. Above cutoff.** Same figure as top panel for intervals where the mark was 'present' (at or above the 'presence' threshold for that mark). Cells corresponding to states with zero posterior probability for observing the mark are colored gray. Again we find positive correlations between these tag enrichments and the emission parameters even though no information on the exact signal level above threshold was given to the model. The high correlations for both above-threshold and below-threshold intervals suggest that chromatin states with high emission probability for a given mark are more likely to have more tags for that mark, both below and above the threshold, compared to states that have lower emission probability for that mark.

Average Expression Level

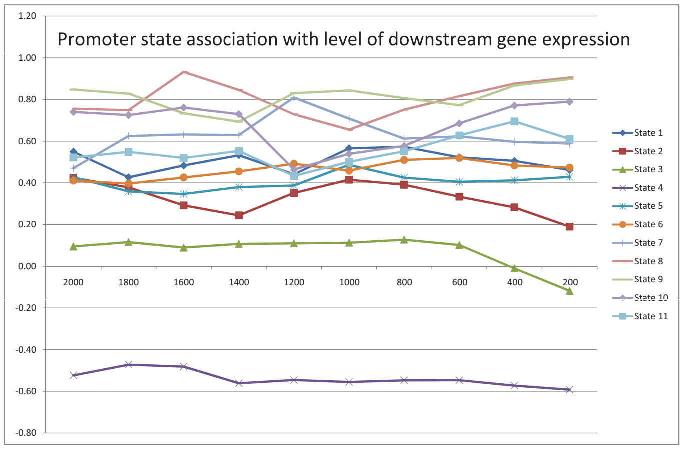
state	2000	1800	1600	1400	1200	1000	800	600	400	200
1	0.55	0.43	0.48	0.53	0.44	0.57	0.57	0.52	0.50	0.46
2	0.42	0.38	0.29	0.24	0.35	0.41	0.39	0.33	0.28	0.19
3	0.09	0.12	0.09	0.11	0.11	0.11	0.13	0.10	-0.01	-0.12
4	-0.52	-0.47	-0.48	-0.56	-0.55	-0.55	-0.55	-0.55	-0.57	-0.59
5	0.43	0.36	0.35	0.38	0.39	0.49	0.42	0.40	0.41	0.43
6	0.41	0.40	0.43	0.45	0.49	0.46	0.51	0.52	0.48	0.47
7	0.47	0.62	0.63	0.63	0.81	0.71	0.61	0.62	0.60	0.59
8	0.76	0.75	0.93	0.85	0.73	0.66	0.75	0.82	0.88	0.91
9	0.85	0.83	0.73	0.69	0.83	0.84	0.81	0.77	0.87	0.90
10	0.74	0.72	0.76	0.73	0.46	0.54	0.58	0.68	0.77	0.79
11	0.52	0.55	0.52	0.55	0.43	0.50	0.55	0.63	0.69	0.61

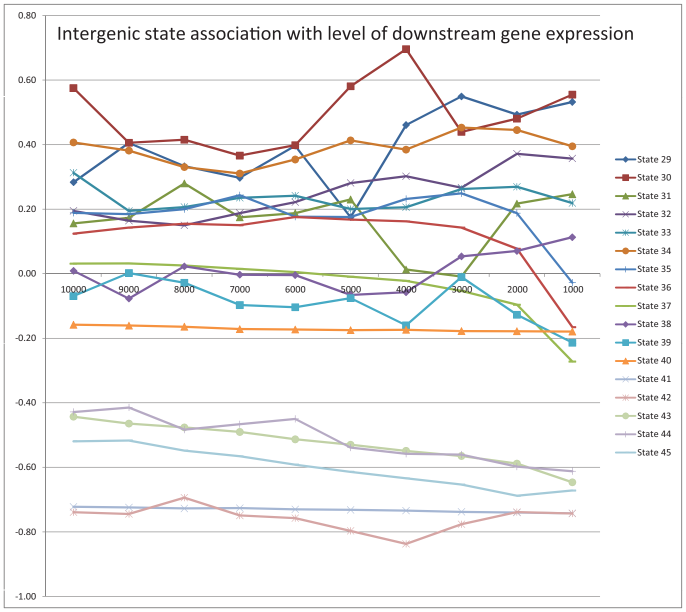
Number of Locations for Average

2000	1800	1600	1400	1200	1000	800	600	400	200
73	82	102	130	182	315	561	904	1110	650
94	126	140	183	263	430	590	773	731	465
165	170	202	267	361	508	659	801	748	693
169	196	222	268	334	420	534	670	827	983
121	153	160	187	243	315	415	563	872	1767
117	131	157	195	261	373	561	833	1228	1513
63	66	77	105	143	209	351	648	1071	1240
53	68	83	82	108	140	165	195	235	327
89	94	98	110	101	94	81	88	100	146
65	69	83	81	119	179	239	250	173	112
94	116	132	172	206	216	172	115	118	97

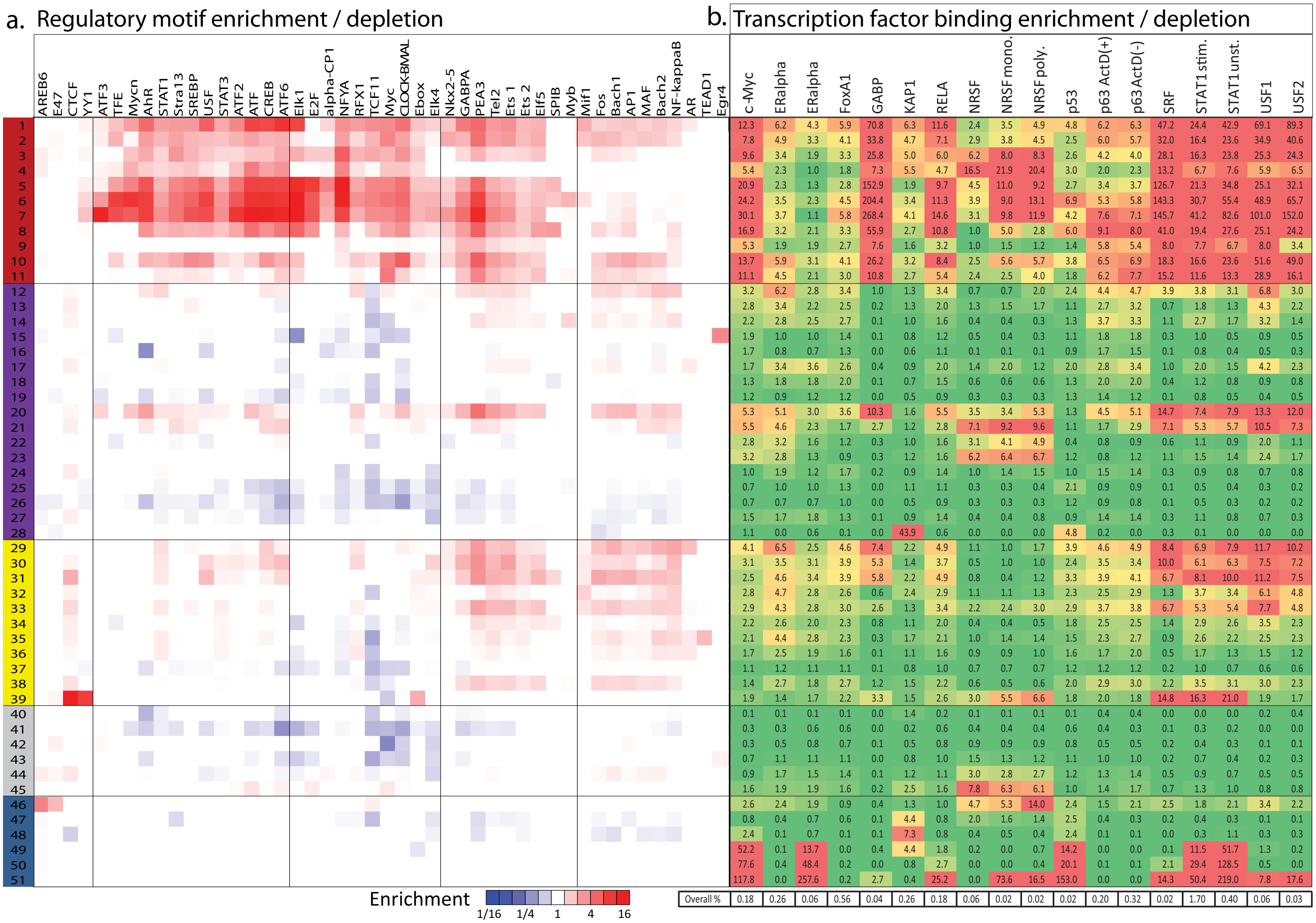
state	10000	9000	8000	7000	6000	5000	4000	3000	2000	1000
1	0.33	0.29	0.19	0.23	0.38	0.44	0.10	0.62	0.55	0.57
2	0.25	0.27	0.22	0.22	0.27	0.13	0.01	0.08	0.42	0.41
3	0.02	0.27	-0.06	0.31	-0.07	0.13	0.18	0.17	0.09	0.11
4	-0.15	-0.43	-0.31	-0.32	-0.28	-0.37	-0.41	-0.55	-0.52	-0.55
5	0.33	0.27	0.28	0.43	0.41	0.42	0.41	0.29	0.43	0.49
6	0.23	0.40	0.23	0.35	0.45	0.41	0.45	0.37	0.41	0.46
7	0.36	0.33	0.34	0.38	0.34	0.63	0.45	0.59	0.47	0.71
8	0.29	0.53	0.43	0.55	0.61	0.54	0.70	0.55	0.76	0.66
9	0.63	0.47	0.74	0.64	0.70	0.60	0.78	0.69	0.85	0.84
10	0.55	0.21	0.31	0.28	0.34	0.40	0.39	0.51	0.74	0.54
11	0.32	0.35	0.62	0.55	0.36	0.35	0.39	0.41	0.52	0.50
12	0.66	0.38	0.41	0.56	0.67	0.75	0.53	0.37	0.43	0.59
13	0.38	0.42	0.39	0.43	0.50	0.39	0.42	0.54	0.39	0.38
14	0.53	0.58	0.72	0.83	0.68	0.75	0.77	0.64	0.80	0.76
15	0.36	0.46	0.36	0.32	0.28	0.34	0.33	0.37	0.36	0.40
16	0.41	0.37	0.42	0.46	0.42	0.44	0.49	0.45	0.45	0.36
17	0.10	0.17	0.20	0.26	0.26	0.11	0.19	0.29	0.31	0.33
18	0.22	0.16	0.19	0.23	0.24	0.25	0.23	0.32	0.22	0.22
19	0.29	0.30	0.32	0.32	0.35	0.38	0.40	0.37	0.37	0.20
20	0.28	0.37	0.12	0.21	0.33	0.30	0.37	0.34	0.35	0.44
21	0.20	0.18	0.31	0.41	0.41	0.34	0.29	0.37	0.39	0.34
22	0.25	0.28	0.31	0.27	0.32	0.34	0.29	0.28	0.31	0.25
23	0.21	0.26	0.21	0.20	0.16	0.18	0.20	0.20	0.12	0.01
24	0.14	0.29	0.35	0.27	0.25	0.20	0.22	0.19	0.16	0.14
25	0.14	0.16	0.14	0.14	0.17	0.13	0.14	0.10	0.15	0.11
26	0.18	0.14	0.12	0.11	0.09	0.09	0.07	0.06	0.01	-0.03
27	0.23	0.26	0.29	0.35	0.40	0.43	0.50	0.43	0.47	0.43
28	-0.05	0.01	0.00	0.11	0.07	0.03	0.01	0.03	-0.01	-0.07
29	0.28	0.41	0.33	0.30	0.40	0.18	0.46	0.55	0.49	0.53
30	0.58	0.41	0.41	0.37	0.40	0.58	0.70	0.44	0.48	0.55
31	0.16	0.17	0.28	0.17	0.19	0.23	0.01	-0.01	0.22	0.25
32	0.19	0.16	0.15	0.19	0.22	0.28	0.30	0.27	0.37	0.36
33	0.31	0.19	0.21	0.24	0.24	0.20	0.21	0.26	0.27	0.22
34	0.41	0.38	0.33	0.31	0.35	0.41	0.38	0.45	0.45	0.39
35	0.19	0.18	0.20	0.24	0.18	0.18	0.23	0.25	0.19	-0.03
36	0.12	0.14	0.15	0.15	0.18	0.17	0.16	0.14	0.08	-0.17
37	0.03	0.03	0.03	0.01	0.00	-0.01	-0.02	-0.05	-0.10	-0.27
38	0.01	-0.08	0.02	0.00	0.00	-0.07	-0.06	0.05	0.07	0.11
39	-0.07	0.00	-0.03	-0.10	-0.10	-0.08	-0.16	-0.01	-0.13	-0.21
40	-0.16	-0.16	-0.16	-0.17	-0.17	-0.17	-0.17	-0.18	-0.18	-0.18
41	-0.72	-0.72	-0.73	-0.73	-0.73	-0.73	-0.73	-0.74	-0.74	-0.74
42	-0.74	-0.74	-0.69	-0.75	-0.76	-0.80	-0.84	-0.78	-0.74	-0.74
43	-0.44	-0.46	-0.48	-0.49	-0.51	-0.53	-0.55	-0.56	-0.59	-0.65
44	-0.43	-0.41	-0.48	-0.47	-0.45	-0.54	-0.56	-0.56	-0.60	-0.61
45	-0.52	-0.52	-0.55	-0.57	-0.59	-0.61	-0.63	-0.65	-0.69	-0.67
46	-0.15	-0.24	0.08	-0.19	-0.27	-0.26	-0.20	-0.35	-0.16	-0.44
47	-0.36	-0.34	-0.32	-0.38	-0.29	-0.32	-0.36	-0.48	-0.60	-0.73
48	-0.32	-0.27	-0.29	-0.34	-0.33	-0.32	-0.26	-0.29	-0.28	-0.47
49	counts too small (see table on the right)									
50	counts too small (see table on the right)									
51	counts too small (see table on the right)									

10000	9000	8000	7000	6000	5000	4000	3000	2000	1000
57	53	45	50	50	63	49	62	73	315
64	52	65	57	68	71	71	80	94	430
68	70	86	90	83	83	90	100	165	508
64	67	85	75	84	94	116	128	169	420
81	78	56	68	56	72	79	82	121	315
72	67	68	61	72	79	83	84	117	373
55	53	49	45	49	47	30	52	63	209
46	42	46	44	46	51	34	53	53	140
41	46	55	54	49	46	53	64	89	94
55	51	44	42	46	47	46	44	65	179
48	69	63	68	60	59	63	61	94	216
53	52	44	45	53	54	52	58	64	65
145	127	154	155	160	153	158	162	196	110
71	73	80	77	85	87	98	112	174	130
254	285	254	265	268	295	292	287	228	149
311	308	324	314	312	319	355	368	312	162
74	78	66	88	81	83	84	102	111	79
257	232	245	251	262	271	269	260	232	151
700	719	737	745	784	771	788	745	618	384
73	61	75	84	74	83	82	86	110	233
174	189	168	185	181	181	200	212	299	290
307	303	289	291	314	310	322	307	267	222
519	514	527	521	516	532	529	552	586	537
297	298	295	293	288	302	277	235	243	209
422	413	405	385	359	355	348	336	313	334
1598	1579	1566	1526	1501	1442	1425	1415	1304	1143
234	234	235	244	237	231	219	214	222	229
54	70	63	62	64	61	61	55	58	61
49	44	58	53	52	68	66	62	90	227
67	80	74	81	74	65	81	98	162	365
60	61	61	59	56	54	63	67	97	321
93	100	101	100	90	66	105	140	181	222
152	156	141	141	165	167	166	186	245	460
300	312	342	338	343	395	413	526	841	900
283	261	269	291	287	336	373	423	415	277
1423	1470	1524	1602	1705	1785	1902	1940	1796	1032
4003	4037	4086	4064	4039	3908	3784	3514	3024	1980
215	232	224	231	223	265	254	318	470	684
72	73	57	60	58	73	64	59	61	53
883	874	865	852	839	827	818	811	800	784
554	546	543	534	521	515	504	490	473	446
34	37	32	32	36	30	31	30	34	42
3032	2925	2832	2755	2664	2542	2425	2237	1969	1636
249	257	247	241	230	225	206	199	183	174
405	425	435	480	518	556	602	711	874	907
25	28	23	25	29	20	24	30	23	23
188	195	191	178	164	152	142	132	119	86
91	83	83	83	86	90	89	93	84	74
7.9	5.7	5.7	6.0	3.5	4.0	2.2	1.6	4.4	4.9
3.0	2.0	1.8	0.3	1.9	1.7	1.2	0.1	0.0	1.0
1.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	2.0	1.0





Supplementary Figure 22: Chromatin state association with expression level of downstream genes. The tables on the left show the average expression level downstream of chromatin states as a function of distance. The top tables give these values at smaller increments specifically for promoter states (every 200bp distance between 200 and 2000). The bottom tables give these values at larger increments (every 1000bp between 1000 and 10000) for all states with a sufficient number of associated genes (excluding states 49-51). The expression level of the downstream location was computed as described in the **Online Methods** using the CD4 T expression data from (Su et al, 2004)[7]. Averages shown in this table were weighted based on the posterior probability of the upstream state assignment. When using the most likely state assignments, the observed difference between states 30 and 31 (mentioned in the main text) were statistically significant at distances of 1,2,3,4,5, and 10kb based on a 2-sided t-test with p-values ranging from $<10^{-6}$ to 0.02. The middle tables indicate the number of locations that each of these averages is based on. The two graphs display the values for promoter regions (top graph, from the top left table), and the values for active intergenic (29-39) and repressed states (41-45) (bottom graph, data from the bottom-left table).

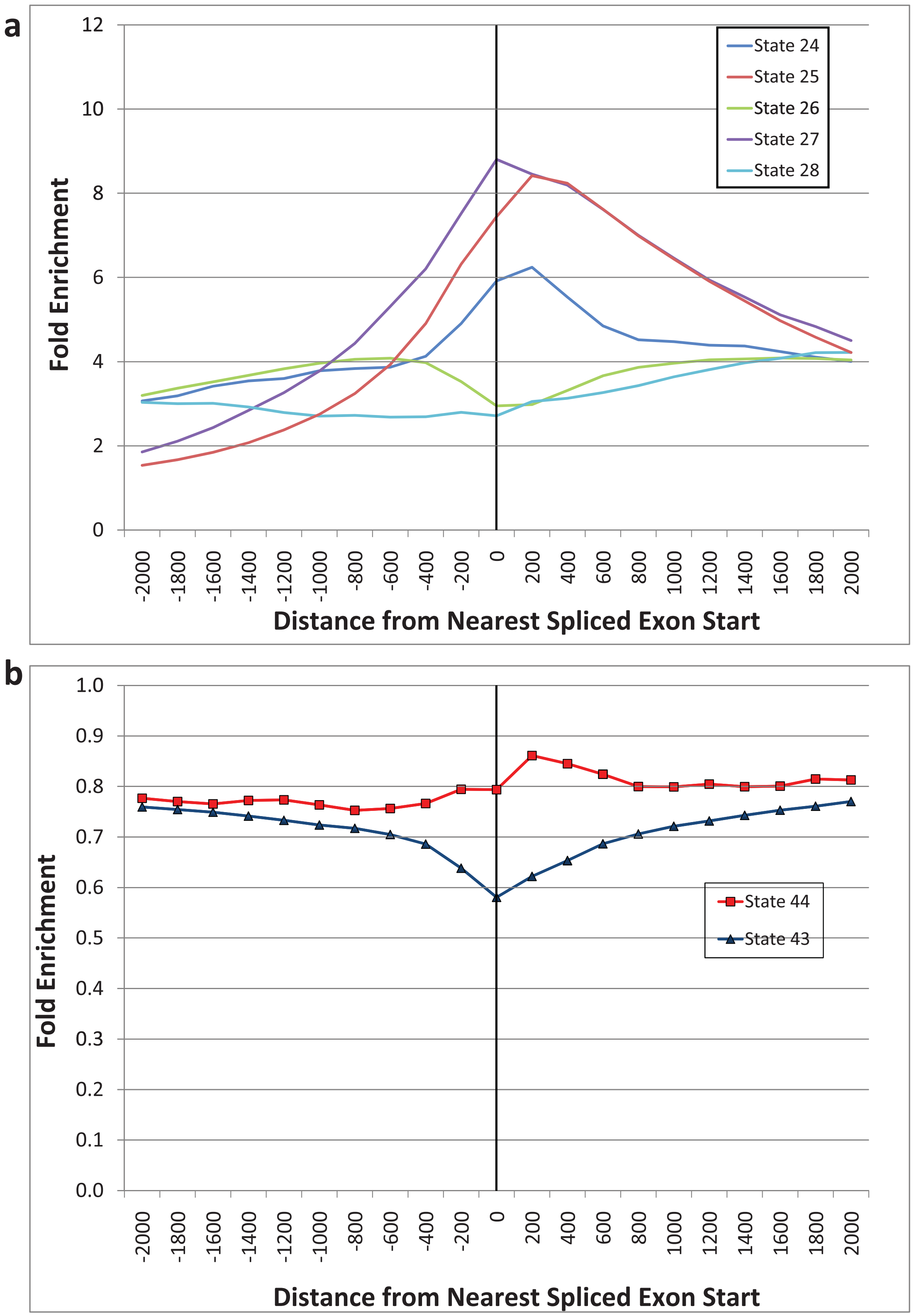


	c-Myc	ERalpha	ERalpha	FoxA1	GABP	KAP1	RELA	NRSF	NRSF mono.	NRSF poly.	p53	p63 ActD(+)	p63 ActD(-)	SRF	STAT1 stim.	STAT1 unst.	USF1	USF2
1	12.3	6.2	4.3	5.9	70.8	6.3	11.6	2.4	3.5	4.9	4.8	6.2	6.3	47.2	24.4	42.9	69.1	89.3
2	7.8	4.9	3.3	4.1	33.8	4.7	7.1	2.9	3.8	4.5	2.5	6.0	5.7	32.0	16.4	23.6	34.9	40.6
3	9.6	3.4	1.9	3.3	25.8	5.0	6.0	6.2	8.0	8.3	2.6	4.2	4.0	28.1	16.3	23.8	25.3	24.3
4	5.4	2.3	1.0	1.8	7.3	5.5	4.7	16.5	21.9	20.4	3.0	2.0	2.3	13.2	6.7	7.6	5.9	6.5
5	20.9	2.3	1.3	2.8	152.9	1.9	9.7	4.5	11.0	9.2	2.7	3.4	3.7	126.7	21.3	34.8	25.1	32.1
6	24.2	3.5	2.3	4.5	204.4	3.4	11.3	3.9	9.0	13.1	6.9	5.3	5.8	143.3	30.7	55.4	48.9	65.7
7	30.1	3.7	1.1	5.8	268.4	4.1	14.6	3.1	9.8	11.9	4.2	7.6	7.1	145.7	41.2	82.6	101.0	152.0
8	16.9	3.2	2.1	3.3	55.9	2.7	10.8	1.0	5.0	2.8	6.0	9.1	8.0	41.0	19.4	27.6	25.1	24.2
9	5.3	1.9	1.9	2.7	7.6	1.6	3.2	1.0	1.5	1.2	1.4	5.8	5.4	8.0	7.7	6.7	8.0	3.4
10	13.7	5.9	3.1	4.1	26.2	3.2	8.4	2.5	5.6	5.7	3.8	6.5	6.9	18.3	16.6	23.6	51.6	49.0
11	11.1	4.5	2.1	3.0	10.8	2.7	5.4	2.4	2.5	4.0	1.8	6.2	7.7	15.2	11.6	13.3	28.9	16.1
12	3.2	6.2	2.8	3.4	1.0	1.3	3.4	0.7	0.7	2.0	2.4	4.4	4.7	3.9	3.8	3.1	6.8	3.0
13	2.8	3.4	2.2	2.5	0.2	1.3	2.0	1.3	1.5	1.7	1.1	2.7	3.2	0.7	1.8	1.3	4.3	2.2
14	2.2	2.8	2.5	2.7	0.1	1.0	1.6	0.4	0.4	0.3	1.3	3.7	3.3	1.1	2.7	1.7	3.2	1.4
15	1.9	1.0	1.0	1.4	0.1	0.8	1.2	0.5	0.4	0.3	1.1	1.8	1.8	0.3	1.0	0.5	0.9	0.5
16	1.7	0.8	0.7	1.3	0.0	0.6	1.1	0.1	0.1	0.1	0.9	1.7	1.5	0.1	0.8	0.4	0.5	0.3
17	1.7	3.4	3.6	2.6	0.4	0.9	2.0	1.4	2.0	1.2	2.0	2.8	3.4	1.0	2.0	1.5	4.2	2.3
18	1.3	1.8	1.8	2.0	0.1	0.7	1.5	0.6	0.6	0.6	1.3	2.0	2.0	0.4	1.2	0.8	0.9	0.8
19	1.2	0.9	1.2	1.2	0.0	0.5	0.9	0.3	0.3	0.3	1.3	1.4	1.2	0.1	0.8	0.5	0.4	0.5
20	5.3	5.1	3.0	3.6	10.3	1.6	5.5	3.5	3.4	5.3	1.3	4.5	5.1	14.7	7.4	7.9	13.3	12.0
21	5.5	4.6	2.3	1.7	2.7	1.2	2.8	7.1	9.2	9.6	1.1	1.7	2.9	7.1	5.3	5.7	10.5	7.3
22	2.8	3.2	1.6	1.2	0.3	1.0	1.6	3.1	4.1	4.9	0.4	0.8	0.9	0.6	1.1	0.9	2.0	1.1
23	3.2	2.8	1.3	0.9	0.3	1.2	1.6	6.2	6.4	6.7	1.2	0.8	1.2	1.1	1.5	1.4	2.4	1.7
24	1.0	1.9	1.2	1.7	0.2	0.9	1.4	1.0	1.4	1.5	1.0	1.5	1.4	0.3	0.9	0.8	0.7	0.8
25	0.7	1.0	1.0	1.3	0.0	1.1	1.1	0.3	0.3	0.4	2.1	0.9	0.9	0.1	0.5	0.4	0.3	0.2
26	0.7	0.7	0.7	1.0	0.0	0.5	0.9	0.3	0.3	0.3	1.2	0.9	0.8	0.1	0.5	0.3	0.2	0.2
27	1.5	1.7	1.8	1.3	0.1	0.9	1.4	0.4	0.4	0.8	0.9	1.4	1.4	0.3	1.1	0.8	0.7	0.3
28	1.1	0.0	0.6	0.1	0.0	43.9	0.6	0.0	0.0	0.0	4.8	0.2	0.0	0.3	0.1	0.0	0.0	0.0
29	4.1	6.5	2.5	4.6	7.4	2.2	4.9	1.1	1.0	1.7	3.9	4.6	4.9	8.4	6.9	7.9	11.7	10.2
30	3.1	3.5	3.1	3.9	5.3	1.4	3.7	0.5	1.0	1.0	2.4	3.5	3.4	10.0	6.1	6.3	7.5	7.2
31	2.5	4.6	3.4	3.9	5.8	2.2	4.9	0.8	0.4	1.2	3.3	3.9	4.1	6.7	8.1	10.0	11.2	7.5
32	2.8	4.7	2.8	2.6	0.6	2.4	2.9	1.1	1.1	1.3	2.3	2.5	2.9	1.3	3.7	3.4	6.1	4.8
33	2.9	4.3	2.8	3.0	2.6	1.3	3.4	2.2	2.4	3.0	2.9	3.7	3.8	6.7	5.3	5.4	7.7	4.8
34	2.2	2.6	2.0	2.3	0.8	1.1	2.0	0.4	0.4	0.5	1.8	2.5	2.5	1.4	2.9	2.6	3.5	2.3
35	2.1	4.4	2.8	2.3	0.3	1.7	2.1	1.0	1.4	1.4	1.5	2.3	2.7	0.9	2.6	2.2	2.5	2.3
36	1.7	2.5	1.9	1.6	0.1	1.1	1.6	0.9	0.9	1.0	1.6	1.7	2.0	0.5	1.7	1.3	1.5	1.2
37	1.1	1.2	1.1	1.1	0.1	0.8	1.0	0.7	0.7	0.7	1.2	1.2	1.2	0.2	1.0	0.7	0.6	0.6
38	1.4	2.7	1.8	2.7	1.2	1.5	2.2	0.6	0.5	0.6	2.0	2.9	3.0	2.2	3.5	3.1	3.0	2.3
39	1.9	1.4	1.7	2.2	3.3	1.5	2.6	3.0	5.5	6.6	1.8	2.0	1.8	14.8	16.3	21.0	1.9	1.7
40	0.1	0.1	0.1	0.1	0.0	1.4	0.2	0.1	0.1	0.1	0.1	0.4	0.4	0.0	0.0	0.0	0.2	0.4
41	0.3	0.3	0.6	0.6	0.0	0.2	0.6	0.4	0.4	0.4	0.6	0.4	0.3	0.1	0.3	0.2	0.0	0.0
42	0.3	0.5	0.8	0.7	0.1	0.5	0.8	0.9	0.9	0.9	0.8	0.5	0.5	0.2	0.4	0.3	0.1	0.1
43	0.7	1.1	1.1	1.1	0.0	0.8	1.0	1.5	1.3	1.2	1.1	1.0	1.0	0.2	0.7	0.4	0.2	0.3
44	0.9	1.7	1.5	1.4	0.1	1.2	1.1	3.0	2.8	2.7	1.2	1.3	1.4	0.5	0.9	0.7	0.5	0.5
45	1.6	1.9	1.9	1.6	0.2	2.5	1.6	7.8	6.3	6.1	1.0	1.4	1.7	0.7	1.3	1.0	0.8	0.8
46	2.6	2.4	1.9	0.9	0.4	1.3	1.0	4.7	5.3	14.0	2.4	1.5	2.1	2.5	1.8	2.1	3.4	2.2
47	0.8	0.4	0.7	0.6	0.1	4.4	0.8	2.0	1.6	1.4	2.5	0.4	0.3	0.2	0.4	0.3	0.1	0.1
48	2.4	0.1	0.7	0.1	0.1	7.3	0.8	0.4	0.5	0.4	2.4	0.1	0.1	0.0	0.3	1.1	0.3	0.3
49	52.2	0.1	13.7	0.0	0.4	4.4	1.8	0.2	0.0	0.7	14.2	0.0	0.0	0.1	11.5	51.7	1.3	0.2
50	77.6	0.4	48.4	0.2	0.0	0.8	2.7	0.0	0.0	0.4	20.1	0.1	0.1	2.1	29.4	128.5	0.5	0.0
51	117.8	0.0	257.6	0.2	2.7	0.4	25.2	0.0	73.6	16.5	153.0	0.0	0.0	14.3	50.4	219.0	7.8	17.6
Overall %	0.18	0.26	0.06	0.56	0.04	0.26	0.18	0.06	0.02	0.02	0.02	0.20	0.32	0.02	1.70	0.40	0.06	0.03

Supplementary Figure 23: Transcription factor binding and motif enrichments.

a. Heatmap showing significant regulatory motif enrichments (red) and depletions (blue) for several transcription factors in both promoter and candidate enhancer states. Vertical black lines distinguish several groups emerging from two-dimensional clustering based on their relative enrichment in promoter states vs. active intergenic states. From left to right, these are: enriched primarily in intergenic, enriched in promoter only, enriched in promoter and depleted in other states, both promoter and intergenic enrichment, enriched in promoter states away from TSS and intergenic states.

b. Heatmap showing fold enrichments (red) and depletions (green) of chromatin states for transcription factor binding (from left to right) c-Myc[8], ERalpha[9], ERalpha[10], FoxA1[9], GABP[11], KAP1[12], RelA[13], NRSF[14], NRSF monoclonal and polyclonal[11], p53[15], p63 (actinomycinD (+) and (-))[16], SRF[11], STAT1 stimulated and unstimulated[17], USF1 and USF2[18]. 'Overall %' row indicates the percentage of 200bp intervals that overlap peak calls for each transcription factor.



Supplementary Figure 24: Spliced exon enrichments/depletion. Fold enrichment (y-axis) relative to distance from the 5'-end of nearest start of a spliced exon (2nd exon or later) (x-axis, shown in the direction of transcription) for subset of transcribed and repressed states. a. **Transcribed states 24-28**. These five states show relative enrichment and depletion patterns with respect to spliced exon boundaries. States 24 and 25 enrichment peaked downstream of the start of the exon while the enrichment of States 21-23 was centered on the start of the exon (**Figure 3c**). These were separated for clarity, due to their different positional biases. **b. Repressed states 43-44**. State 43 shows its greatest depletion nearspliced exon 5' boundaries, while state 44 had less depletion at the interval 200bp downstream of the 5' end of exon as compared to flanking regions. State 44 had a relatively greater frequency for H3K27me2 and H3K27me3 consistent with observations made previously on the association of repressive modifications with exons[19].

state	Resting Unphos. Pol2	Active Unphos. Pol2	Resting Phos. Pol2	Active Phos. Pol2
1	4.3	1.4	3.1	3.1
2	2.2	1.1	2.0	2.3
3	1.4	0.8	1.3	1.6
4	0.6	0.5	0.5	0.6
5	2.2	0.7	1.6	1.6
6	3.6	1.1	2.5	2.5
7	5.1	1.4	3.2	3.4
8	4.1	1.1	3.4	3.2
9	3.0	1.1	3.6	3.5
10	3.2	1.1	3.1	3.0
11	2.3	1.1	2.8	3.3
12	1.5	1.0	2.5	2.6
13	1.2	1.0	2.2	2.4
14	1.5	1.0	2.1	2.2
15	1.3	1.0	2.1	2.1
16	1.2	0.9	1.6	1.5
17	1.1	1.0	1.7	2.2
18	1.1	1.0	1.5	1.8
19	1.1	0.9	1.3	1.4
20	1.7	1.0	2.0	2.3
21	1.2	0.9	1.8	2.4
22	0.8	0.9	1.5	2.0
23	0.7	0.8	1.0	1.4
24	1.0	1.0	1.4	2.0
25	1.0	1.0	1.3	1.7
26	1.0	1.0	1.2	1.5
27	1.5	1.1	3.4	3.7
28	1.1	1.1	1.3	1.3
29	2.2	1.0	2.2	2.0
30	2.3	1.1	2.3	2.0
31	1.4	1.1	1.3	1.7
32	1.0	0.9	1.1	1.3
33	1.2	1.0	1.3	1.6
34	1.2	1.0	1.3	1.5
35	0.9	1.0	0.9	1.0
36	0.8	0.9	0.8	0.9
37	0.9	0.9	0.8	0.8
38	1.0	1.0	0.9	1.1
39	1.4	0.9	1.4	1.5
40	0.8	1.0	0.7	0.9
41	0.9	1.0	0.8	0.7
42	0.9	1.1	0.8	0.7
43	0.9	1.0	0.8	0.7
44	0.8	1.0	0.8	0.7
45	0.8	1.0	0.7	0.7
46	0.7	0.9	0.7	0.8
47	1.0	1.0	0.9	0.8
48	0.7	1.0	0.7	0.6
49	0.6	0.7	0.5	0.4
50	0.7	0.9	0.6	0.5
51	1.5	1.8	1.4	1.4

Supplementary Figure 25: Elongating vs. resting Pol2 enrichments relative to an IgG control. The table shows the fold enrichments for phosphorylated (elongating) and unphosphorylated (stalled) Pol2 for sequence reads in resting and active CD4 T[1] relative to IgG control[6]. While the Pol2 data used in learning our model was not specific to either form, this table shows that the highly-expressed transcribed state 27 is much more enriched for phosphorylated Pol2 in both active and resting cells, while in contrast the active TSS states (States 5-7) were more enriched for the unphosphorylated Pol2 in resting cells, showing that the remaining marks are a good predictor for the form of Pol2. The read counts for the enrichments were determined based on the 5’ end of the read after applying a 100 bp shift in the 5’ to 3’ direction.

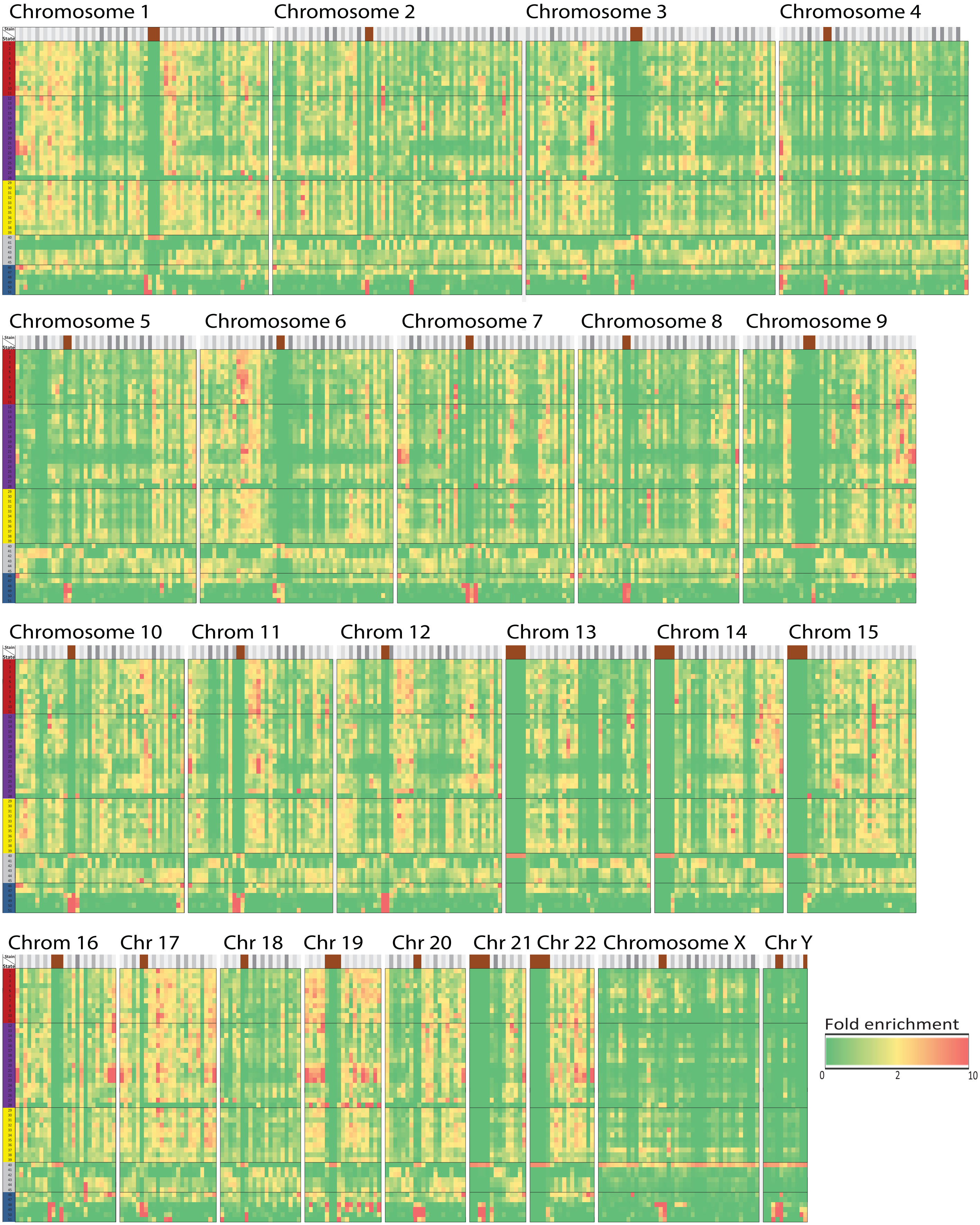
States ordered by GC content

state	aa/tt	at	ta	ga/tc	ag/ct	ac/gt	ca/tg	cg	gc	cc/gg	GC%
5	4.8	2.9	2.5	5.4	6.4	4.0	5.3	8.7	10.9	11.7	64.0
4	4.9	3.0	2.5	5.9	7.0	4.3	6.0	6.9	9.7	10.7	61.4
6	5.1	3.0	2.6	5.9	6.9	4.5	5.8	7.4	9.7	10.6	61.3
7	5.0	2.8	2.6	6.2	7.1	4.7	5.8	7.2	9.2	10.3	60.9
22	5.1	4.0	2.9	6.1	7.9	5.3	8.2	2.5	7.3	9.1	55.5
21	5.2	4.0	3.0	6.1	7.9	5.2	8.0	2.5	7.1	9.3	55.4
23	5.8	4.2	3.2	6.0	7.7	5.1	7.7	2.7	7.1	9.1	54.5
8	6.8	4.4	3.9	6.0	7.1	4.8	6.4	4.7	7.4	8.8	54.0
3	6.8	4.4	3.8	6.1	7.3	4.8	6.7	4.2	7.3	8.5	53.4
10	5.9	4.1	3.4	6.4	8.0	5.3	7.6	2.7	6.6	8.4	53.3
1	6.4	4.1	3.6	6.4	7.8	5.2	7.2	3.4	6.7	8.1	52.9
20	6.2	4.7	3.7	6.2	8.0	5.3	8.1	1.7	6.2	8.0	51.6
46	4.2	5.9	4.0	4.7	5.9	8.5	11.6	1.8	6.2	6.0	50.9
12	6.4	4.8	3.8	6.3	8.1	5.3	8.1	1.3	6.0	7.8	50.7
32	6.5	5.0	3.8	6.4	8.1	5.2	8.0	1.3	5.8	7.8	50.5
2	7.2	4.8	4.1	6.3	7.7	5.2	7.3	2.7	6.2	7.4	50.1
11	7.2	5.1	4.4	6.1	7.7	5.2	7.5	2.2	6.0	7.5	49.6
29	6.9	5.1	4.1	6.4	8.1	5.2	7.9	1.3	5.6	7.4	49.4
35	7.0	5.4	4.2	6.4	8.0	5.1	7.9	1.2	5.6	7.4	49.0
33	7.6	5.7	4.7	6.1	7.7	5.2	7.8	1.4	5.5	6.9	47.6
17	7.6	6.0	5.1	6.1	7.8	5.4	8.0	1.0	5.3	6.4	46.4
31	8.1	5.8	4.9	6.3	7.9	5.2	7.7	1.2	5.2	6.2	46.0
13	8.0	6.1	5.2	6.0	7.6	5.3	7.8	1.2	5.2	6.4	46.0
36	8.3	6.4	5.2	6.1	7.5	5.0	7.6	1.3	5.1	6.5	45.6
24	8.0	6.2	5.3	6.1	7.7	5.3	7.8	1.1	5.2	6.2	45.6
45	8.5	6.5	5.4	6.2	7.4	5.0	7.4	1.4	5.0	6.4	45.2
39	8.8	6.6	5.7	5.9	7.3	5.1	7.4	1.4	5.1	6.1	44.3
27	8.9	6.5	5.7	5.9	7.3	5.2	7.4	1.3	5.0	6.0	44.2
44	8.4	6.9	5.6	6.2	7.5	5.2	7.8	0.9	4.8	5.9	44.1
34	9.2	6.7	5.8	6.0	7.3	5.0	7.3	1.3	4.8	6.0	43.7
38	9.1	6.8	5.8	6.2	7.5	5.1	7.4	1.0	4.7	5.6	43.0
30	9.4	6.8	5.9	6.1	7.3	5.1	7.3	1.1	4.5	5.6	42.6
51	8.9	8.5	4.8	7.3	6.3	5.1	7.7	2.1	3.7	5.1	42.5
37	9.4	7.2	6.1	6.0	7.1	5.0	7.3	1.1	4.6	5.7	42.5
14	9.4	7.0	6.1	6.0	7.3	5.1	7.3	1.0	4.4	5.6	42.3
9	10.1	7.2	6.6	5.6	6.7	5.0	6.6	2.1	4.8	5.6	42.1
18	9.5	7.2	6.4	5.9	7.3	5.2	7.4	0.9	4.4	5.3	41.7
25	9.5	7.3	6.4	6.0	7.2	5.2	7.4	1.0	4.4	5.2	41.6
43	9.7	7.7	6.4	6.0	7.1	5.0	7.3	0.9	4.2	5.3	41.1
48	10.1	7.4	5.8	6.2	7.0	5.0	7.5	1.0	4.2	5.0	41.0
16	10.2	7.4	6.6	5.7	6.9	5.1	7.0	1.3	4.5	5.2	40.9
50	9.4	10.7	4.2	8.0	5.2	4.2	8.0	1.8	2.8	5.4	40.9
28	9.7	7.6	6.5	6.0	7.0	5.3	7.5	0.9	4.1	5.1	40.9
19	10.3	7.6	6.8	5.7	6.9	5.1	7.1	1.1	4.3	5.1	40.3
26	10.2	7.7	6.8	5.8	6.9	5.1	7.1	1.0	4.2	5.0	40.2
49	10.0	9.6	4.8	7.5	5.9	4.5	7.8	1.5	3.2	4.8	39.9
40	10.1	8.1	6.9	5.9	6.8	5.1	7.3	0.9	4.0	4.9	39.8
15	10.4	7.8	7.0	5.7	6.9	5.2	7.1	1.0	4.0	4.8	39.5
42	10.0	8.3	7.1	5.9	7.0	5.1	7.4	0.7	4.0	4.7	39.3
47	10.8	8.6	7.4	5.8	6.7	5.0	7.1	0.6	3.6	4.4	37.6
41	11.1	9.1	7.8	5.8	6.6	5.0	7.0	0.6	3.4	4.1	36.6

States ordered by state id

state	aa/tt	at	ta	ga/tc	ag/ct	ac/gt	ca/tg	cg	gc	cc/gg	GC%
1	6.4	4.1	3.6	6.4	7.8	5.2	7.2	3.4	6.7	8.1	52.9
2	7.2	4.8	4.1	6.3	7.7	5.2	7.3	2.7	6.2	7.4	50.1
3	6.8	4.4	3.8	6.1	7.3	4.8	6.7	4.2	7.3	8.5	53.4
4	4.9	3.0	2.5	5.9	7.0	4.3	6.0	6.9	9.7	10.7	61.4
5	4.8	2.9	2.5	5.4	6.4	4.0	5.3	8.7	10.9	11.7	64.0
6	5.1	3.0	2.6	5.9	6.9	4.5	5.8	7.4	9.7	10.6	61.3
7	5.0	2.8	2.6	6.2	7.1	4.7	5.8	7.2	9.2	10.3	60.9
8	6.8	4.4	3.9	6.0	7.1	4.8	6.4	4.7	7.4	8.8	54.0
9	10.1	7.2	6.6	5.6	6.7	5.0	6.6	2.1	4.8	5.6	42.1
10	5.9	4.1	3.4	6.4	8.0	5.3	7.6	2.7	6.6	8.4	53.3
11	7.2	5.1	4.4	6.1	7.7	5.2	7.5	2.2	6.0	7.5	49.6
12	6.4	4.8	3.8	6.3	8.1	5.3	8.1	1.3	6.0	7.8	50.7
13	8.0	6.1	5.2	6.0	7.6	5.3	7.8	1.2	5.2	6.4	46.0
14	9.4	7.0	6.1	6.0	7.3	5.1	7.3	1.0	4.4	5.6	42.3
15	10.4	7.8	7.0	5.7	6.9	5.2	7.1	1.0	4.0	4.8	39.5
16	10.2	7.4	6.6	5.7	6.9	5.1	7.0	1.3	4.5	5.2	40.9
17	7.6	6.0	5.1	6.1	7.8	5.4	8.0	1.0	5.3	6.4	46.4
18	9.5	7.2	6.4	5.9	7.3	5.2	7.4	0.9	4.4	5.3	41.7
19	10.3	7.6	6.8	5.7	6.9	5.1	7.1	1.1	4.3	5.1	40.3
20	6.2	4.7	3.7	6.2	8.0	5.3	8.1	1.7	6.2	8.0	51.6
21	5.2	4.0	3.0	6.1	7.9	5.2	8.0	2.5	7.1	9.3	55.4
22	5.1	4.0	2.9	6.1	7.9	5.3	8.2	2.5	7.3	9.1	55.5
23	5.8	4.2	3.2	6.0	7.7	5.1	7.7	2.7	7.1	9.1	54.5
24	8.0	6.2	5.3	6.1	7.7	5.3	7.8	1.1	5.2	6.2	45.6
25	9.5	7.3	6.4	6.0	7.2	5.2	7.4	1.0	4.4	5.2	41.6
26	10.2	7.7	6.8	5.8	6.9	5.1	7.1	1.0	4.2	5.0	40.2
27	8.9	6.5	5.7	5.9	7.3	5.2	7.4	1.3	5.0	6.0	44.2
28	9.7	7.6	6.5	6.0	7.0	5.3	7.5	0.9	4.1	5.1	40.9
29	6.9	5.1	4.1	6.4	8.1	5.2	7.9	1.3	5.6	7.4	49.4
30	9.4	6.8	5.9	6.1	7.3	5.1	7.3	1.1	4.5	5.6	42.6
31	8.1	5.8	4.9	6.3	7.9	5.2	7.7	1.2	5.2	6.2	46.0
32	6.5	5.0	3.8	6.4	8.1	5.2	8.0	1.3	5.8	7.8	50.5
33	7.6	5.7	4.7	6.1	7.7	5.2	7.8	1.4	5.5	6.9	47.6
34	9.2	6.7	5.8	6.0	7.3	5.0	7.3	1.3	4.8	6.0	43.7
35	7.0	5.4	4.2	6.4	8.0	5.1	7.9	1.2	5.6	7.4	49.0
36	8.3	6.4	5.2	6.1	7.5	5.0	7.6	1.3	5.1	6.5	45.6
37	9.4	7.2	6.1	6.0	7.1	5.0	7.3	1.1	4.6	5.7	42.5
38	9.1	6.8	5.8	6.2	7.5	5.1	7.4	1.0	4.7	5.6	43.0
39	8.8	6.6	5.7	5.9	7.3	5.1	7.4	1.4	5.1	6.1	44.3
40	10.1	8.1	6.9	5.9	6.8	5.1	7.3	0.9	4.0	4.9	39.8
41	11.1	9.1	7.8	5.8	6.6	5.0	7.0	0.6	3.4	4.1	36.6
42	10.0	8.3	7.1	5.9	7.0	5.1	7.4	0.7	4.0	4.7	39.3
43	9.7	7.7	6.4	6.0	7.1	5.0	7.3	0.9	4.2	5.3	41.1
44	8.4	6.9	5.6	6.2	7.5	5.2	7.8	0.9	4.8	5.9	44.1
45	8.5	6.5	5.4	6.2	7.4	5.0	7.4	1.4	5.0	6.4	45.2
46	4.2	5.9	4.0	4.7	5.9	8.5	11.6	1.8	6.2	6.0	50.9
47	10.8	8.6	7.4	5.8	6.7	5.0	7.1	0.6	3.6	4.4	37.6
48	10.1	7.4	5.8	6.2	7.0	5.0	7.5	1.0	4.2	5.0	41.0
49	10.0	9.6	4.8	7.5	5.9	4.5	7.8	1.5	3.2	4.8	39.9
50	9.4	10.7	4.2	8.0	5.2	4.2	8.0	1.8	2.8	5.4	40.9
51	8.9	8.5	4.8	7.3	6.3	5.1	7.7	2.1	3.7	5.1	42.5

Supplementary Figure 26: State di-nucleotide composition. Left table show the percentage of di-nucleotide pairs in each of the states, grouping dinucleotides that are reverse complements of each other as they have the same occurrences. Right side contains the same information, sorted by GC percentage. This table shows that State 46 has the highest ca/tg di-nucleotide occurrence frequency of any state, and that the TSS states (4-7) have the highest CpG frequency.



Supplementary Figure 27: Chromatin state enrichments for each chromosomal staining band for all human chromosomes. For each chromosome, the staining pattern is shown (top row) with gneg (no stain), gpos25, gpos50, gpos75, and gpos100 patterns shown using progressively darker shades, and brown used to represent stalk, acrocentric, and variable heterochromatic bands (see **Supplementary Fig. 28** headers for color legend). The coordinates of the bands and staining patterns were obtained from the UCSC genome browser[20,21]. This figure shows that the satellite enriched states (48-51) are enriched in centromere regions of the chromosome, that specific chromosome bands with darker stains are found with states 41 and 42, that the zinc finger enriched state (state 28) enriches on chromosome 19, and the unmappable state (state 40) enriches on regions at the beginning of several chromosomes.

state	stalk	variable heterocrhomatic	acrocentric	gneg	gpos25	gpos50	gpos75	gpos100
1	0.0	0.1	0.1	1.5	1.7	1.0	0.6	0.3
2	0.0	0.1	0.1	1.4	1.6	1.0	0.6	0.4
3	0.0	0.1	0.1	1.4	1.7	1.0	0.6	0.4
4	0.0	0.1	0.1	1.4	1.6	1.0	0.7	0.4
5	0.0	0.1	0.1	1.6	1.6	0.9	0.5	0.3
6	0.0	0.1	0.1	1.5	1.8	1.0	0.5	0.3
7	0.0	0.1	0.1	1.4	1.7	1.0	0.6	0.4
8	0.0	0.0	0.1	1.5	1.5	1.0	0.5	0.3
9	0.0	0.0	0.2	1.5	1.3	1.0	0.6	0.5
10	0.0	0.1	0.0	1.6	1.7	0.9	0.4	0.2
11	0.0	0.0	0.1	1.6	1.7	0.9	0.4	0.2
12	0.0	0.0	0.1	1.6	1.6	1.0	0.4	0.2
13	0.0	0.0	0.1	1.6	1.6	1.0	0.4	0.2
14	0.0	0.0	0.1	1.5	1.2	1.2	0.6	0.3
15	0.0	0.1	0.1	1.5	1.5	1.0	0.6	0.4
16	0.0	0.1	0.1	1.4	1.4	1.1	0.7	0.4
17	0.0	0.0	0.1	1.5	1.5	1.2	0.6	0.3
18	0.0	0.0	0.1	1.4	1.4	1.1	0.7	0.4
19	0.0	0.1	0.1	1.4	1.5	1.1	0.8	0.4
20	0.0	0.0	0.1	1.6	1.8	0.9	0.4	0.2
21	0.0	0.0	0.1	1.8	2.0	0.6	0.2	0.1
22	0.0	0.0	0.0	1.9	1.8	0.5	0.2	0.1
23	0.0	0.0	0.0	1.8	2.0	0.5	0.2	0.1
24	0.0	0.1	0.1	1.5	1.5	1.1	0.6	0.3
25	0.0	0.1	0.1	1.4	1.4	1.2	0.8	0.4
26	0.0	0.1	0.1	1.3	1.4	1.1	0.8	0.5
27	0.0	0.0	0.0	1.6	1.8	0.9	0.4	0.3
28	0.0	1.9	0.3	1.0	5.8	0.3	0.2	0.1
29	0.0	0.1	0.1	1.5	1.5	1.0	0.6	0.2
30	0.0	0.1	0.1	1.4	1.3	1.1	0.7	0.4
31	0.0	0.1	0.1	1.4	1.3	1.0	0.7	0.5
32	0.0	0.0	0.1	1.6	1.7	0.9	0.4	0.2
33	0.0	0.0	0.1	1.5	1.8	0.9	0.5	0.3
34	0.0	0.1	0.1	1.5	1.5	1.0	0.6	0.3
35	0.0	0.0	0.0	1.6	1.6	0.9	0.5	0.3
36	0.0	0.1	0.1	1.5	1.6	0.9	0.5	0.3
37	0.0	0.1	0.1	1.4	1.4	1.0	0.8	0.5
38	0.0	0.1	0.1	1.3	1.2	1.1	0.9	0.6
39	0.0	0.1	0.1	1.3	1.5	1.1	0.8	0.5
40	7.6	6.6	6.1	0.6	0.4	0.5	0.3	0.5
41	0.0	0.2	0.4	0.5	0.4	0.9	1.7	2.5
42	0.0	0.2	0.3	0.5	0.5	1.1	1.7	2.2
43	0.0	0.1	0.1	1.1	1.2	1.3	1.1	0.7
44	0.0	0.1	0.1	1.2	1.3	1.3	1.0	0.6
45	0.0	0.1	0.1	1.3	1.6	1.3	0.8	0.4
46	0.0	0.1	0.2	1.7	1.7	0.7	0.4	0.2
47	0.0	0.2	0.1	1.2	1.3	1.3	0.9	0.6
48	0.0	3.2	6.2	0.8	2.2	0.2	0.5	0.4
49	0.0	3.6	11.2	0.5	1.8	0.2	0.4	0.1
50	0.0	4.7	12.0	0.6	0.6	0.2	0.2	0.1
51	0.0	4.4	12.7	0.5	1.4	0.2	0.2	0.1
% Overall	0.6	3.9	3.6	42.1	6.8	13.6	13.1	16.2

Supplementary Figure 28: Staining band genome-wide enrichments for each state. Genome-wide fold enrichment of states for each of the staining patterns[21] shows that state 41 and 42 are the only two states enriched for the gpos100 stain.

GO Category/ State	1 (3%)	2 (2%)	3 (5%)	4 (8%)	5 (14%)	6 (13%)	7 (9%)	8 (3%)	36 (3%)	37 (5%)	40 (5%)	41 (3%)	43 (8%)	45 (4%)
tRNA metabolic process	4.44 (0.003)	0.74 (1)	1.55 (1)	0.18 (1)	1.35 (1)	1.95 (0.45)	2.44 (0.014)	1.46 (1)	0 (1)	0.42 (1)	0 (1)	0 (1)	0 (1)	0 (1)
Cell Cycle Phase	1.16 (1)	1.51 (1)	2.70 $(2x10^{-7})$	0.57 (1)	1.61 0.001	1.45 (1)	1.15 (1)	1.51 (1)	0.65 (1)	0.53 (1)	0 .12(1)	0.52 (1)	0.38 (1)	0.33 (1)
Embryonic Development	0.50 (1)	0.93 (1)	1.24 (1)	2.82 $(9x10^{-23})$	1.07 (1)	0.85 (1)	0.54 (1)	1.00 (1)	0.78 (1)	0.39 (1)	0.16 (1)	0.53 (1)	0.87 (1)	3.20 $(2.3x10^{-13})$
Chromatin	0.81 (1)	0.64 (1)	1.20 (1)	0.48 (1)	2.17 $(1.4x10^{-7})$	1.64 (1)	0.85 (1)	0.85 (1)	1.43 (1)	0.40 (1)	1.71 (1)	0 (1)	0.39 (1)	0 (1)
Response to DNA Damage Stimulus	2.04 (1)	1.14 (1)	1.20 (1)	0.35 (1)	1.55 (0.074)	2.13 $(6.5x10^{-11})$	1.97 $(1.0x10^{-4})$	0.84 (1)	0 .19 (1)	0.58 (1)	0.72 (1)	0 (1)	0.15 (1)	0 .07 (1)
RNA Processing	1.71 (1)	0.77 (1)	0.49 (1)	0.26 (1)	1.31 (1)	1.91 $(4.2x10^{-11})$	2.64 $(8.7x10^{-24})$	2.46 $(3.0x10^{-4})$	0.19 (1)	0.16 (1)	0.54 (1)	0.13 (1)	0.08 (1)	0.17 (1)
T cell Activation	1.46 (1)	2.70(1)	0.77 (1)	0.88 (1)	1.27 (1)	0.70 (1)	0.79 (1)	4.72 $(2x10^{-7})$	0.41 (1)	0.52 (1)	0.31 (1)	0 (1)	0.50 (1)	0.85 (1)
Intermediate Filament	0.20 (1)	0.51 (1)	0 .24 (1)	0.45 (1)	0.37 (1)	0.08 (1)	0.11 (1)	0 (1)	7.84 $(9.2x10^{-21})$	1.67 (1)	0 (1)	4.38 $(9.1x10^{-5})$	2.81 $(1.8x10^{-7})$	0.40 (1)
Hormone Activity	0.27 (1)	0.34 (1)	0.83 (1)	1.24 (1)	0.15(1)	0.11 (1)	0.39 (1)	0(1)	1.58 (1)	3.33 $(4.3x10^{-4})$	1.4 (1)	0.83 (1)	1.54 (1)	2.73 (1)
Male Gamete Generation	1.13 (1)	1.13 (1)	0.77 (1)	1.14 (1)	0.67 (1)	0.97 (1)	0.65 (1)	0.76 (1)	2.35 (1)	1.67 (1)	2.80 (0.002)	1.43 (1)	0.94 (1)	0.67 (1)
Olfactory Receptor Activity	0 (1)	0.12 (1)	0.11 (1)	0 (1)	0 (1)	0.02 (1)	0 (1)	0 (1)	0.70 (1)	0.93 (1)	1.53 (1)	8.04 $(2.3x10^{-49})$	5.19 $(1.1x10^{-89})$	0.66 (1)

Supplementary Figure 29: Gene Ontology (GO) enrichments for states with the most transcription start sites (TSS). The table shows the Gene Ontology (GO) enrichments for selected GO categories for the states with the largest number of RefSeq TSS assigned to them based on mostly like state assignments. On the top row, colored by state grouping, are listed the states and for each the percentage of RefSeq TSS that are assigned to that state. In each cell is the fold enrichment and Bonferroni-corrected p-value for genes of that category with a TSS in that state computed using the STEM software[22]. Cells with p-values <=0.01 are highlighted in yellow.

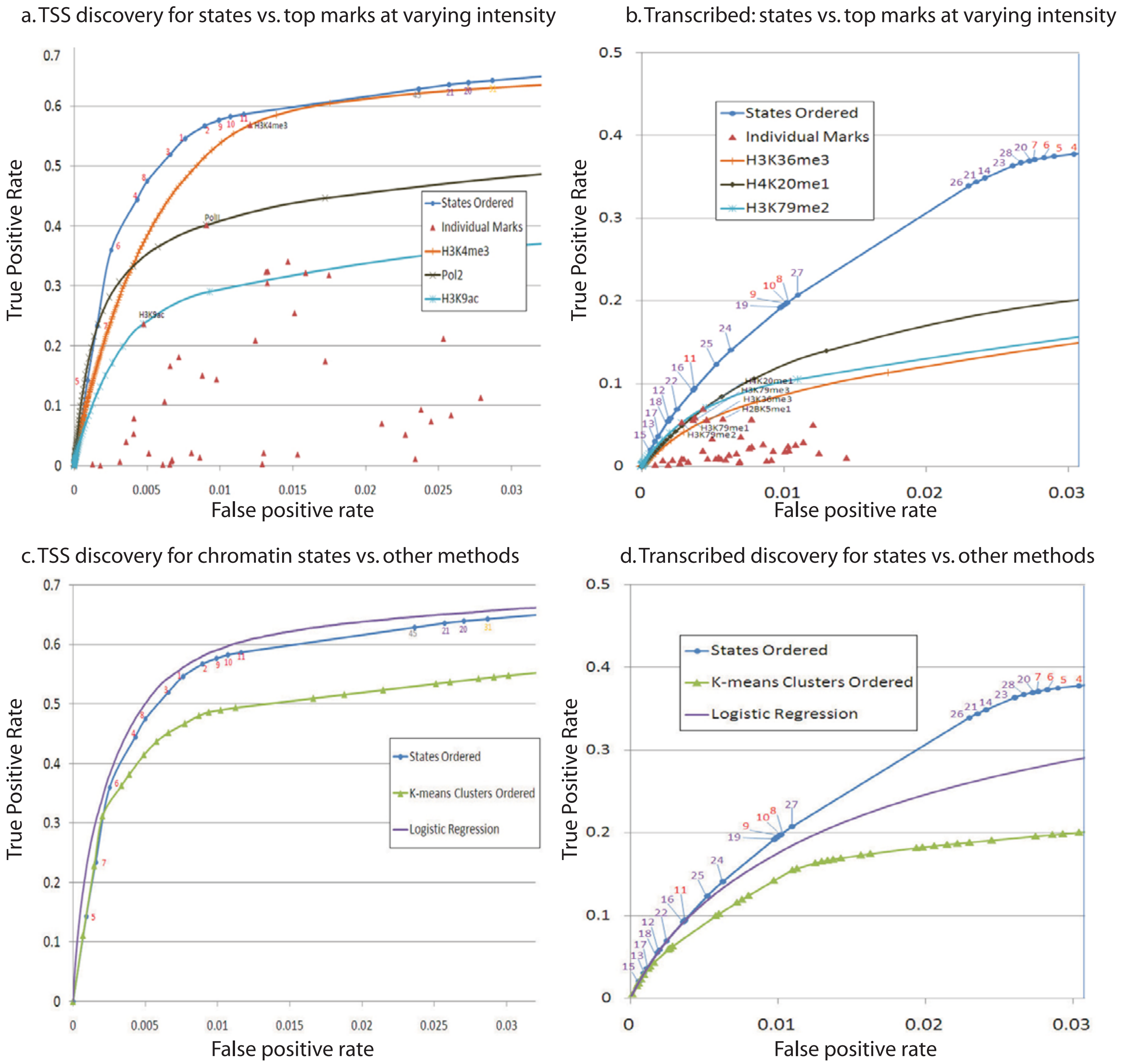
state	H3K9ac 0h	H4K16ac 0h	PolII 0h	H3K9ac: log_2(2h/0h)	H3K9ac: log_2(8h/0h)	H4K16ac: log_2(2h/0h)	H4K16ac: log_2(8h/0h)	PolII log_2(8h/0h)	PolII log_2(2h/0h)
1	20.9	16.8	6.5	1.6	1.6	1.6	1.0	-0.2	0.6
2	9.3	7.7	4.1	2.0	1.9	2.0	1.5	-0.1	0.6
3	4.6	3.2	3.5	2.3	2.2	2.7	2.3	0.0	0.8
4	1.2	0.7	1.9	1.7	1.9	2.8	2.8	0.1	0.7
5	4.8	3.0	5.5	2.2	1.7	2.7	2.0	-0.1	1.2
6	11.8	7.0	7.7	2.0	1.6	2.4	1.6	-0.1	1.2
7	35.9	22.4	10.6	1.0	0.7	1.3	0.4	-0.3	1.0
8	15.0	6.4	5.4	2.4	1.7	2.9	2.1	0.0	0.8
9	8.4	4.2	2.8	2.8	1.4	3.0	2.4	-0.2	0.5
10	10.7	8.1	5.0	2.7	2.7	2.7	2.1	0.1	0.6
11	5.7	3.8	3.5	3.3	3.1	3.4	2.9	0.2	0.7
12	2.5	4.7	2.4	3.2	3.7	2.2	2.1	0.4	0.7
13	2.1	3.7	1.9	1.9	2.5	1.2	1.5	0.2	0.6
14	2.5	4.8	1.8	2.5	2.1	1.6	1.5	0.0	0.4
15	2.0	3.6	1.5	1.0	0.8	0.4	0.8	-0.2	0.3
16	1.1	2.0	0.9	0.5	0.2	0.0	0.5	-0.2	0.3
17	2.0	3.7	1.7	0.4	1.5	-0.1	0.6	0.2	0.5
18	1.8	3.3	1.4	-0.9	-0.2	-1.1	-0.2	0.0	0.3
19	1.2	2.1	1.0	-1.4	-1.1	-1.4	-0.5	-0.2	0.2
20	3.0	4.4	2.9	2.7	3.2	2.3	2.1	0.2	0.7
21	1.2	2.0	2.2	2.9	3.8	2.3	2.5	0.3	0.9
22	0.8	1.7	1.6	0.2	2.3	0.0	1.0	0.3	0.8
23	0.6	1.1	1.1	0.3	1.8	0.2	1.0	0.3	0.7
24	1.5	2.8	1.5	-1.8	-0.2	-1.8	-0.7	0.1	0.3
25	1.3	2.2	1.3	-2.8	-1.7	-2.4	-1.3	0.0	0.1
26	1.3	2.0	1.1	-2.7	-2.1	-2.4	-1.4	-0.2	0.1
27	1.2	2.5	1.6	0.6	1.9	0.0	0.9	0.0	0.6
28	0.7	0.6	1.3	-2.4	-2.6	-0.9	-0.1	-0.1	0.1
29	4.9	7.7	3.0	2.3	2.5	1.8	1.5	0.0	0.3
30	4.3	5.7	2.2	2.2	1.6	1.7	1.4	-0.2	0.2
31	3.8	5.2	2.1	0.8	0.9	0.4	0.5	-0.1	0.1
32	1.4	3.3	1.8	1.7	2.5	1.4	1.5	0.2	0.4
33	1.9	2.8	1.8	1.6	2.1	1.3	1.5	0.1	0.4
34	1.5	2.6	1.3	1.4	1.3	1.0	1.1	-0.1	0.3
35	1.0	1.9	1.4	-0.4	0.8	0.0	0.6	0.1	0.2
36	0.9	1.3	1.0	-1.4	-0.7	-0.9	-0.1	0.0	0.0
37	0.9	1.0	0.9	-2.5	-2.4	-2.0	-1.3	0.0	-0.1
38	1.8	2.1	1.3	-0.7	-0.8	-0.6	-0.2	-0.1	-0.1
39	1.4	1.3	2.6	-0.1	0.2	0.1	0.9	-0.4	-0.1
40	0.1	0.1	0.2	-2.9	-3.6	-2.2	-2.2	-1.8	-0.3
41	0.8	0.5	1.0	-3.2	-4.5	-2.1	-2.4	-0.2	-0.3
42	0.9	0.5	1.1	-3.2	-4.1	-2.1	-2.4	0.0	-0.3
43	0.9	0.7	1.0	-3.2	-3.8	-2.4	-2.4	0.0	-0.3
44	0.9	0.7	1.1	-3.1	-3.4	-2.4	-2.3	0.1	-0.2
45	0.9	0.6	1.1	-2.4	-2.2	-1.3	-0.9	0.1	-0.2
46	0.9	1.3	1.8	-1.3	-0.2	-0.8	-0.2	0.2	0.1
47	0.9	0.7	1.1	-3.1	-4.1	-2.1	-1.9	-0.1	-0.3
48	0.7	0.4	1.0	-2.6	-3.4	-1.5	-1.1	0.1	-0.2
49	2.4	1.5	4.0	-3.1	-3.9	-2.1	-2.1	0.2	-0.2
50	6.9	4.7	14.1	-2.9	-3.7	-1.3	-2.1	0.1	-0.2
51	57.1	33.0	133.4	-2.6	-3.5	-2.0	-2.1	0.1	-0.2

Supplementary Figure 30: Histone Deacetylase (HDAC) inhibition response enrichments. For each chromatin state is shown the enrichment for H3K9ac, H4K16ac, PolII tags before HDAC inhibition (first group of columns) based on the data of (Wang et al, 2009)[23]. The next three pairs of columns show the log-base-2 fold change for these three marks at 2 hours and 8 hours after HDAC inhibition. The repressive promoter state (State 4) shows a notable increase in acetylation enrichment after HDAC inhibition, while repressive states 40-45 do not.

state	LINE	SINE	LTR	DNA	Simple_repeat	Low_complexity	Satellite	Other	snRNA	rRNA	tRNA	srpRNA	scRNA	Unknown	RNA
1	0.3	0.6	0.4	0.5	0.6	0.5	0.1	0.0	3.6	0.5	6.0	0.1	0.9	1.1	0.5
2	0.4	0.7	0.6	0.7	0.9	0.6	0.1	0.1	1.1	0.7	4.8	0.5	2.4	1.6	0.7
3	0.3	0.7	0.4	0.5	1.4	1.8	0.3	0.0	0.9	1.0	15.0	1.1	1.5	0.5	0.1
4	0.1	0.3	0.2	0.2	2.6	5.0	0.8	0.1	0.5	1.1	13.9	0.6	0.2	0.0	0.0
5	0.1	0.4	0.1	0.2	4.4	9.1	0.2	0.0	1.7	0.4	20.5	0.8	0.4	0.4	0.6
6	0.1	0.3	0.1	0.2	1.8	3.7	0.2	0.0	2.5	0.3	19.7	2.2	0.6	0.0	0.1
7	0.1	0.2	0.1	0.2	0.9	1.8	0.1	0.0	2.7	0.0	6.4	1.2	0.6	0.0	0.8
8	0.2	0.4	0.1	0.3	1.3	2.6	0.1	0.0	0.2	1.2	0.3	0.3	2.5	2.2	0.0
9	0.4	0.8	0.2	0.6	1.3	2.1	0.0	0.1	1.8	0.3	0.2	0.1	1.1	0.0	1.1
10	0.3	0.7	0.2	0.5	0.7	0.4	0.0	0.1	0.6	3.5	0.6	1.2	0.5	0.0	1.9
11	0.4	0.8	0.2	0.6	0.9	0.6	0.1	0.0	0.5	3.3	0.6	0.2	2.9	1.2	0.0
12	0.5	0.8	0.5	0.8	0.6	0.2	0.0	0.0	0.3	2.6	0.8	1.9	0.7	0.8	2.0
13	0.7	1.1	0.6	1.3	0.7	0.4	0.0	0.1	1.1	3.5	0.5	2.7	1.5	0.7	3.3
14	0.7	1.2	0.7	1.2	0.8	0.8	0.1	0.1	1.3	1.3	0.4	3.5	1.4	0.4	1.5
15	1.0	1.5	0.5	1.5	0.8	0.9	0.0	0.5	2.5	1.7	0.8	2.1	3.1	0.1	1.3
16	1.0	2.4	0.4	1.2	1.0	1.2	0.0	2.0	2.2	0.9	0.7	1.4	2.6	0.3	0.9
17	0.8	0.9	0.8	1.2	0.8	0.4	0.0	0.1	0.7	1.9	0.5	2.8	1.6	0.0	2.0
18	0.9	1.3	0.8	1.5	0.8	0.7	0.0	0.7	1.5	1.3	0.8	3.2	2.0	0.1	1.0
19	1.1	1.9	0.6	1.3	1.0	1.1	0.0	3.5	2.0	0.8	0.8	1.4	2.1	0.2	0.6
20	0.5	0.8	0.5	0.7	1.1	0.4	0.1	0.0	0.4	0.4	1.2	0.8	1.2	0.3	0.5
21	0.3	0.8	0.3	0.5	1.4	0.7	0.2	0.1	0.9	0.5	0.9	1.2	1.2	0.0	0.1
22	0.4	0.7	0.4	0.6	0.9	0.5	0.1	0.6	0.5	0.7	0.6	1.3	1.5	0.0	0.0
23	0.4	1.6	0.3	0.6	1.3	1.0	0.3	3.9	0.7	0.7	0.6	1.2	1.4	0.3	0.7
24	0.7	0.9	0.7	1.1	0.6	0.5	0.2	0.4	1.1	1.5	1.3	2.2	1.2	0.0	2.7
25	0.7	1.1	0.7	1.3	0.7	0.9	0.1	0.3	1.4	1.2	1.2	1.5	1.6	0.1	1.7
26	0.9	1.6	0.5	1.2	0.9	1.2	0.1	2.1	1.7	1.1	1.0	1.2	2.3	0.2	0.9
27	0.4	1.4	0.3	1.1	0.7	0.9	0.0	0.3	2.2	1.3	1.3	2.7	2.3	0.0	0.5
28	1.1	1.3	1.4	0.9	0.7	0.7	0.7	0.3	1.1	4.2	1.3	0.0	2.1	0.0	0.0
29	0.5	0.9	0.8	0.9	0.8	0.4	0.1	0.0	0.6	0.9	1.8	1.5	1.3	1.3	1.4
30	0.6	1.2	0.8	1.0	1.0	1.0	0.2	0.1	1.7	2.9	1.5	1.6	2.8	2.6	1.3
31	0.5	0.8	1.1	1.1	0.7	0.5	0.1	0.1	1.1	0.8	2.8	1.4	1.2	2.0	1.4
32	0.5	0.9	1.0	0.9	1.0	0.5	0.1	0.1	0.2	0.4	0.2	2.1	0.7	0.7	4.0
33	0.6	1.2	0.8	1.1	1.3	0.8	0.1	0.1	1.1	1.0	1.2	2.0	1.8	0.1	1.5
34	0.7	1.8	0.8	1.1	1.2	1.2	0.1	0.5	1.9	0.9	0.9	1.5	2.5	0.4	1.3
35	0.6	1.0	1.2	1.0	0.9	0.5	0.1	0.1	0.7	1.3	0.7	1.4	1.4	1.1	1.9
36	0.8	1.6	1.1	1.2	1.2	0.8	0.1	1.1	1.2	1.1	1.0	2.1	1.6	1.0	1.1
37	1.1	1.5	1.1	1.2	1.1	0.9	0.3	2.2	1.5	1.3	1.1	1.9	1.7	0.8	0.9
38	0.7	1.0	1.2	1.2	0.9	0.8	0.3	0.3	1.2	1.1	1.7	1.4	1.4	1.4	1.8
39	0.5	0.9	0.6	1.1	1.1	1.0	0.3	0.2	1.2	0.6	2.8	1.0	1.6	0.3	0.7
40	0.6	0.4	0.6	0.4	0.5	0.5	1.1	0.5	0.5	0.4	0.7	0.6	0.4	0.3	0.4
41	1.3	0.7	1.3	1.0	1.1	1.4	0.7	0.4	0.6	0.7	0.7	0.3	0.3	1.3	1.0
42	1.1	0.6	1.5	1.0	1.1	1.1	1.0	0.4	0.5	1.0	0.8	0.6	0.3	1.4	1.4
43	1.1	1.1	1.1	1.2	1.1	0.9	0.2	1.1	1.0	1.0	0.8	0.9	0.9	1.7	1.2
44	0.8	0.9	1.3	1.1	1.2	0.7	0.3	0.6	0.7	1.2	0.8	1.0	0.6	1.3	1.6
45	0.7	1.0	0.9	1.0	1.2	1.1	0.2	0.5	0.9	0.8	2.6	0.7	0.8	1.0	1.3
46	0.4	0.5	0.6	0.5	13.9	0.4	0.8	3.3	0.6	4.3	0.1	0.3	0.1	0.0	1.0
47	1.8	0.7	1.6	1.0	0.8	0.7	0.5	0.6	1.4	0.9	2.8	0.7	1.7	0.8	1.0
48	1.0	0.8	1.2	0.6	0.8	0.6	63.4	3.1	0.5	5.5	1.1	0.9	0.4	0.0	0.4
49	0.4	0.3	0.5	0.5	1.3	0.5	138	0.3	0.8	17.2	1.7	1.1	0.0	0.0	1.0
50	0.2	0.4	0.3	0.3	2.9	0.4	158	0.6	0.9	47.9	6.7	0.0	0.0	0.0	2.7
51	0.2	0.3	0.3	0.3	1.3	0.8	159	0.0	0.0	365	12.2	0.0	0.0	0.0	0.0
% Overall	26.8	21.7	11.2	5.03	3.29	2.82	0.42	0.15	0.04	0.02	0.02	0.01	0.01	0.01	0.01

state	LINE: L1	LINE: L2	LINE: CR1	LINE: RTE	SINE: Alu	SINE: MIR	LTR: MaLR	LTR: ERVL	LTR: ERVK	LTR: ERV1	Satellite: centr	Satellite: Satellite	Satellite: telo	Satellite: acro	Simple_repeat: (TG)n	Simple_repeat: (CA)n	Simple_repeat: (CATG)n	Low_complexity: GC_rich
1	0.1	1.0	0.9	0.3	0.2	1.8	0.3	0.4	0.3	0.0	0.0	0.2	0.0	0.0	0.9	0.9	0.0	4.8
2	0.2	1.2	0.8	0.5	0.3	1.6	0.4	0.6	0.5	2.2	0.0	0.3	1.0	0.0	1.4	1.2	0.4	5.0
3	0.2	0.7	0.5	0.5	0.5	1.2	0.2	0.4	0.6	0.9	0.2	0.4	2.1	0.0	1.4	1.3	0.3	25.0
4	0.1	0.3	0.3	0.2	0.2	0.6	0.1	0.1	0.3	0.0	0.6	0.7	7.3	0.0	1.7	1.7	0.7	103.9
5	0.1	0.2	0.1	0.1	0.5	0.3	0.0	0.1	0.4	0.6	0.1	0.3	2.7	0.0	0.6	0.6	0.0	197.1
6	0.1	0.3	0.2	0.2	0.3	0.5	0.0	0.1	0.2	0.6	0.0	0.4	1.1	0.0	0.6	0.7	0.3	82.5
7	0.0	0.2	0.1	0.1	0.1	0.5	0.0	0.0	0.1	1.7	0.0	0.3	0.0	0.0	0.3	0.4	0.0	39.7
8	0.1	0.4	0.6	0.6	0.2	1.0	0.1	0.1	0.1	0.0	0.0	0.3	0.0	0.0	0.8	0.9	0.8	54.3
9	0.3	0.6	0.9	0.9	0.7	1.1	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.8	1.1	0.0	28.8
10	0.1	1.2	1.3	0.4	0.1	2.1	0.2	0.2	0.1	0.1	0.0	0.1	0.0	0.0	1.4	1.6	7.1	4.1
11	0.1	1.2	1.1	0.5	0.4	1.9	0.2	0.3	0.1	2.5	0.0	0.2	0.1	0.0	1.5	1.4	10.6	5.2
12	0.2	1.8	1.4	0.7	0.3	2.3	0.7	0.6	0.0	1.5	0.0	0.0	0.4	0.0	1.0	1.2	1.7	0.1
13	0.4	1.8	1.3	0.9	0.9	1.7	0.7	0.7	0.2	0.7	0.0	0.0	0.0	0.0	1.1	1.0	1.3	0.3
14	0.5	1.5	1.6	1.2	1.1	1.5	0.8	0.7	0.3	0.3	0.0	0.1	0.3	0.0	0.9	0.7	0.0	0.2
15	0.9	1.1	1.2	1.3	1.7	1.1	0.5	0.3	0.4	0.5	0.0	0.0	0.0	0.0	0.7	0.7	0.5	0.2
16	1.1	0.7	0.8	0.9	3.1	0.7	0.4	0.2	0.7	0.6	0.0	0.0	0.0	0.0	0.8	0.7	0.0	0.2
17	0.4	2.1	1.3	0.7	0.6	1.7	0.9	1.0	0.2	2.1	0.0	0.1	0.2	0.0	1.7	1.4	5.2	0.2
18	0.7	1.7	1.4	1.2	1.3	1.4	0.8	0.7	0.4	1.0	0.0	0.0	0.1	0.0	0.9	0.8	0.9	0.1
19	1.1	0.9	1.0	1.2	2.3	0.9	0.5	0.3	0.8	0.7	0.0	0.0	0.1	0.0	0.7	0.8	0.2	0.1
20	0.1	1.5	1.3	0.8	0.3	2.0	0.5	0.8	0.3	0.1	0.0	0.4	0.3	0.0	2.8	2.1	2.8	1.8
21	0.1	1.1	0.6	0.4	0.6	1.6	0.3	0.3	0.2	0.7	0.0	0.6	1.6	0.0	2.7	2.5	4.3	6.4
22	0.2	1.0	0.7	0.5	0.6	1.2	0.4	0.3	0.1	0.9	0.0	0.3	0.9	0.0	1.3	1.4	4.4	2.2
23	0.3	0.8	0.5	0.4	1.8	1.0	0.3	0.3	0.5	0.8	0.0	0.7	1.9	0.0	1.0	1.0	2.6	5.9
24	0.4	1.6	1.3	1.0	0.7	1.5	0.8	0.9	0.3	0.4	0.0	0.6	0.1	0.0	0.8	0.9	1.5	0.1
25	0.6	1.1	1.1	1.2	1.1	1.1	0.8	0.7	0.9	0.4	0.0	0.4	0.0	0.0	0.6	0.7	0.4	0.0
26	0.9	1.0	1.1	1.2	1.9	0.9	0.5	0.4	0.8	0.4	0.0	0.2	0.1	0.0	0.7	0.7	0.3	0.1
27	0.3	0.8	0.8	0.8	1.5	1.0	0.3	0.3	0.4	0.6	0.0	0.1	0.4	0.0	0.5	0.6	0.0	0.7
28	1.2	0.7	0.3	0.7	1.6	0.4	0.9	1.4	5.8	1.9	0.9	0.2	0.5	13.6	0.6	0.7	0.0	0.0
29	0.1	1.6	1.3	1.2	0.4	2.2	0.9	1.0	0.3	1.0	0.0	0.2	0.0	0.0	1.1	1.3	0.2	0.3
30	0.4	1.3	1.4	1.5	1.0	1.6	0.7	0.6	0.6	2.5	0.2	0.3	0.0	0.0	1.1	1.1	0.2	1.1
31	0.3	1.5	1.4	0.9	0.4	1.8	1.0	1.3	0.5	1.3	0.0	0.3	0.4	0.0	1.2	1.1	0.5	0.6
32	0.2	1.7	1.3	0.9	0.5	2.1	1.1	1.3	0.2	0.9	0.0	0.2	0.0	0.0	1.9	1.8	2.6	0.2
33	0.3	1.5	1.2	0.9	1.1	1.8	0.8	1.0	0.5	0.8	0.0	0.2	1.0	0.0	2.1	1.8	2.1	1.3
34	0.5	1.2	1.2	1.1	1.9	1.3	0.7	0.6	0.8	0.4	0.0	0.3	1.1	0.0	1.0	1.0	0.1	1.5
35	0.3	1.7	1.5	0.9	0.7	2.0	1.3	1.6	0.3	0.3	0.0	0.2	0.2	0.0	1.2	1.2	0.6	0.2
36	0.6	1.3	1.3	1.1	1.7	1.5	1.1	1.1	0.8	0.7	0.0	0.3	0.8	0.0	1.1	1.2	0.6	0.3
37	1.1	1.0	1.0	1.1	1.7	1.0	1.0	0.9	1.4	1.1	0.1	0.6	1.4	0.1	0.9	1.0	0.7	0.3
38	0.4	1.5	1.5	1.3	0.8	1.6	1.2	1.4	0.7	1.7	0.1	0.6	0.6	0.0	1.2	1.1	0.5	0.6
39	0.4	0.9	1.1	1.0	0.7	1.4	0.5	0.9	1.4	11.9	0.0	0.3	7.5	0.0	1.0	1.0	0.5	2.5
40	0.7	0.3	0.4	0.3	0.5	0.3	0.5	0.5	1.0	0.8	0.7	1.8	2.6	0.8	0.5	0.5	0.6	0.2
41	1.4	1.0	1.0	1.1	0.6	0.9	1.3	1.3	1.1	1.3	0.7	0.7	0.6	1.3	1.1	1.1	0.4	0.0
42	1.0	1.2	1.0	1.0	0.4	1.0	1.4	2.3	0.8	1.8	0.9	1.4	1.1	4.1	1.3	1.2	1.7	0.1
43	1.1	1.2	1.2	1.2	1.0	1.2	1.2	1.2	0.9	0.9	0.0	0.5	0.6	0.1	1.0	1.0	0.7	0.2
44	0.6	1.5	1.4	1.1	0.6	1.6	1.4	1.9	0.5	1.3	0.0	0.6	0.8	0.0	1.8	1.8	4.3	0.2
45	0.5	1.4	1.4	1.0	0.7	1.7	1.0	1.2	0.5	1.2	0.1	0.4	0.4	0.0	1.5	1.4	1.0	4.9
46	0.3	0.7	0.5	0.3	0.3	0.9	0.5	0.7	0.2	0.0	0.2	0.9	15.5	0.0	44.8	43.8	301.6	1.0
47	2.2	0.7	0.7	0.9	0.7	0.7	1.7	1.2	2.7	1.2	0.1	1.1	0.3	0.1	0.8	0.8	0.8	0.2
48	1.2	0.4	0.2	0.4	1.0	0.3	0.7	0.8	5.7	1.4	91.1	15.5	5.9	45.8	0.6	0.6	1.0	0.6
49	0.4	0.4	0.3	0.4	0.3	0.3	0.4	0.5	0.6	1.4	108.8	200.2	9.6	223.4	0.7	0.6	12.6	0.4
50	0.2	0.3	0.2	0.8	0.4	0.2	0.3	0.4	0.1	0.0	75.0	324.0	7.3	124.5	0.7	1.0	35.6	0.7
51	0.2	0.2	0.1	0.0	0.3	0.0	0.1	0.2	1.7	0.0	128.9	227.8	0.0	100.8	0.3	1.5	0.0	3.7
% Overall	20.69	5.469	0.662	0.210	16.09	6.211	5.300	2.203	0.335	0.010	0.269	0.143	0.010	0.001	0.437	0.436	0.004	0.095

Supplementary Figure 31: RepeatMasker class and family enrichments. Left: The table reports the fold enrichment for each RepeatMasker[24] class of repeats obtained from the UCSC genome browser[20]. The bottom row reports the total percentage of 200bp intervals containing one of these repeat elements. The columns were ordered based on the percentage in the bottom row. For example, while 3.29% of 200-bp intervals (bottom row) intersect a simple repeat element, there is a 13.9 fold enrichment in State 46 (since 45.7% of intervals in State 46 intersect a simple repeat element). Color scale adjusted for each column differently. **Right:** The table reports the fold enrichment for each RepeatMasker family of repeats of the class LINE, SINE, LTR, or Satellite, and the enrichments for (TG)n and (CA)n Simple Repeats, and the GC_rich repeat of the Low complexity class. These specific additional enrichments were selected since they cover at least 5% of at least one state.



a. TSS discovery for states vs. top marks at varying intensity

b. Transcribed: states vs. top marks at varying intensity

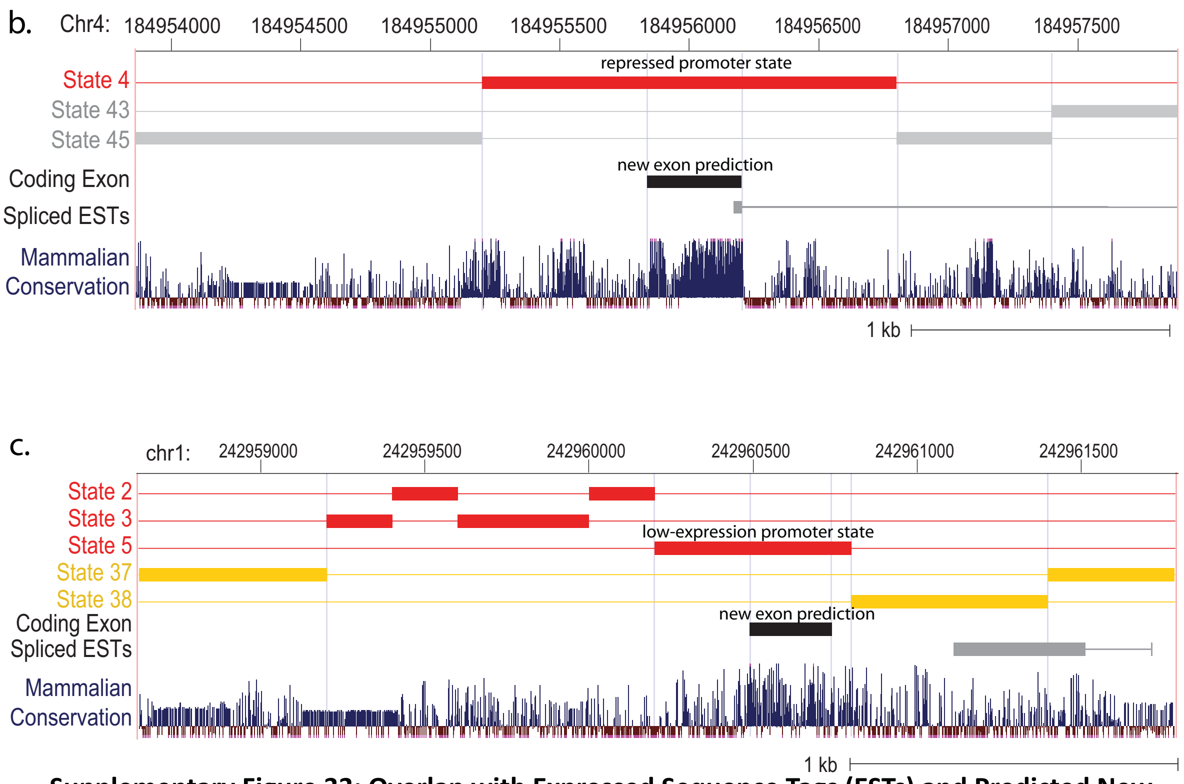
c. TSS discovery for chromatin states vs. other methods

d. Transcribed discovery for states vs. other methods

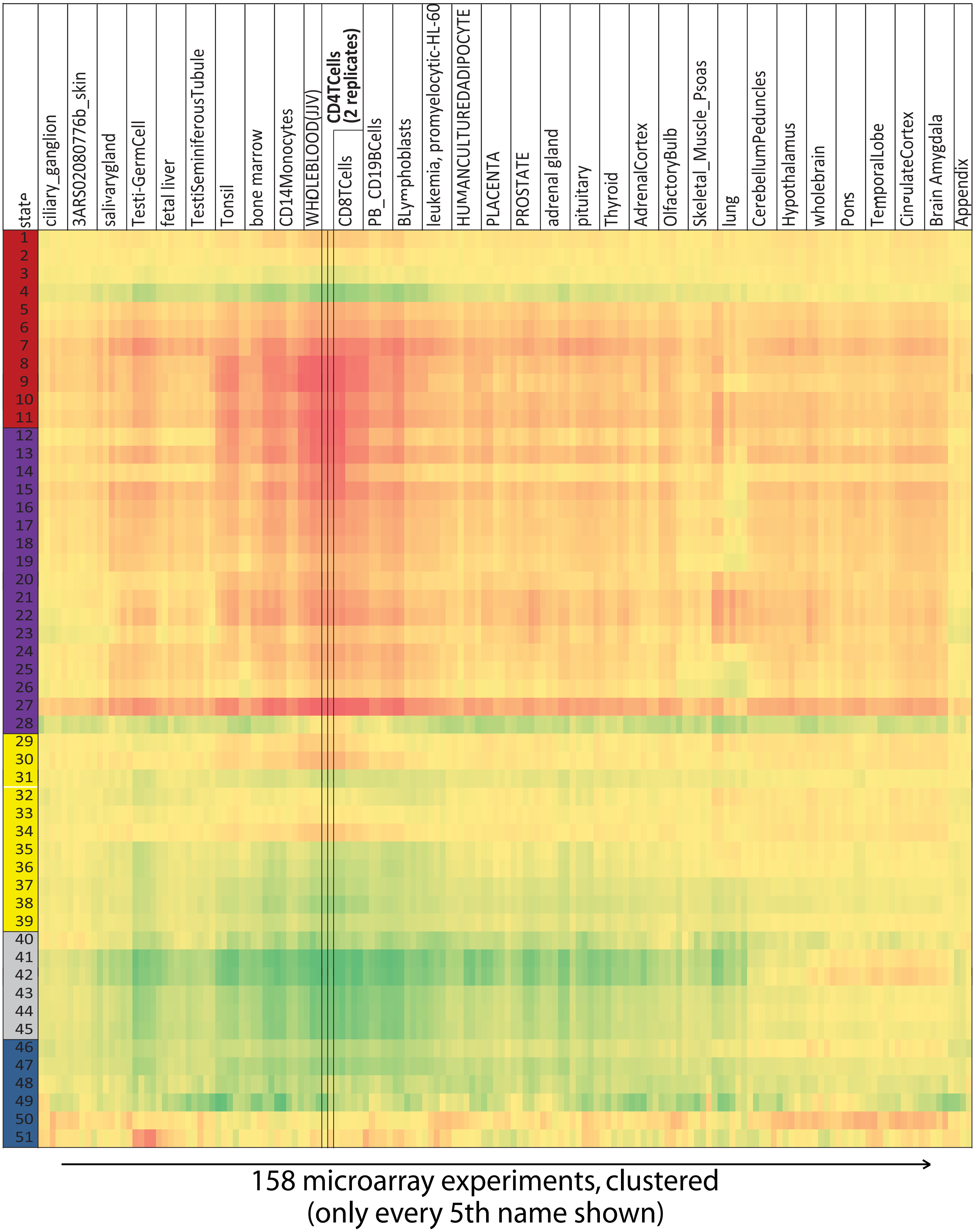
Supplementary Figure 32: Comparison of TSS recovery with individual marks at varying intensity thresholds, k-means, and logistic regression. a-d. Receiver Operating Characteristic (ROC) curves for recovery of RefSeq Transcription Start Sites (TSS) and RefSeq Transcribed Regions for chromatin states, best-performing chromatin marks at varying read-count signal intensity thresholds, and two alternate methods: a 51-cluster k-means clustering (green) of the same binarized input features used by chromatin states, but without any spatial information (also ordered by their TSS enrichment), and a supervised learning logistic regression classifier (purple) given chromatin mark signal intensity information and labeled gene annotation data but lacking spatial information. **a. TSS recovery for chromatin states and individual chromatin marks.** ROC curves shown for the three top-performing marks at all varying intensity levels obtained by varying thresholds in the number of reads within a bin necessary for a presence call for a given mark. Even though H3K4me3 performs similar to chromatin states at a 1% false positive rate, it significantly underperforms states at more stringent false positive rates (20% lower true positive rate at a false positive rate of 0.5%). **b. Transcribed region recovery for chromatin states and individual chromatin marks.** ROC curves shown for chromatin states and the top three chromatin marks at varying intensity thresholds. No single mark performs comparably to chromatin states that are able to use both combinations of marks and spatial context information. **c. TSS recovery for chromatin states and alternative methods.** Chromatin states outperform a k-means clustering approach, showing that even for the identification of TSS that are very punctate, spatial context information can play an important role, likely capturing transitions between upstream and downstream promoter states. We also compared to logistic regression, a supervised classifier that specifically learns predictive chromatin combinations by training on known TSS, while our chromatin states were learned de novo without any previous annotation information. Logistic regression slightly out-performed chromatin states (~3% increase in performance), benefiting from the supervised learning approach, and also having access to mark the full spectrum of mark intensity information. **d. Transcribed region recovery for chromatin states and alternate methods.** In contrast to TSS, transcribed regions are much harder to recover without spatial information, and chromatin states strongly outperform both k-means clustering and logistic regression with locally defined features.

a.

state	% RefSeq	% EST	% EST for non-RefSeq	Lin et al, Exons	Lin et al, Exons | Not RefSeq Exon
1	45	82	68	2	3
2	44	77	61	2	3
3	49	79	61	4	6
4	52	80	62	6	14
5	63	90	75	5	11
6	61	91	78	6	6
7	70	96	86	8	6
8	87	98	85	5	4
9	89	98	87	3	2
10	87	97	80	4	2
11	92	98	89	4	2
12	93	97	79	2	1
13	95	99	91	2	1
14	82	94	69	2	1
15	95	99	92	1	0
16	92	98	84	1	0
17	94	99	85	2	1
18	93	99	85	2	1
19	89	97	79	1	0
20	70	87	58	3	2
21	83	93	67	9	5
22	92	98	82	11	7
23	81	93	68	7	5
24	90	97	74	6	3
25	91	97	73	8	2
26	86	95	70	3	1
27	88	96	72	8	2
28	77	96	86	2	2
29	44	71	51	2	1
30	46	74	54	1	1
31	32	65	50	1	1
32	42	67	46	2	2
33	51	75	52	2	2
34	52	77	55	1	1
35	34	62	45	2	2
36	33	61	45	1	2
37	38	65	47	1	1
38	32	61	45	1	1
39	37	64	46	1	1
40	10	25	19	0	0
41	22	46	33	0	0
42	23	47	34	0	0
43	33	58	41	1	1
44	34	59	41	1	2
45	34	60	43	1	2
46	42	63	41	2	2
47	29	57	42	0	1
48	17	52	44	0	1
49	9	40	34	0	0
50	6	42	38	0	0
51	6	47	43	1	0
Overall	36	58	37	2.0	0.1



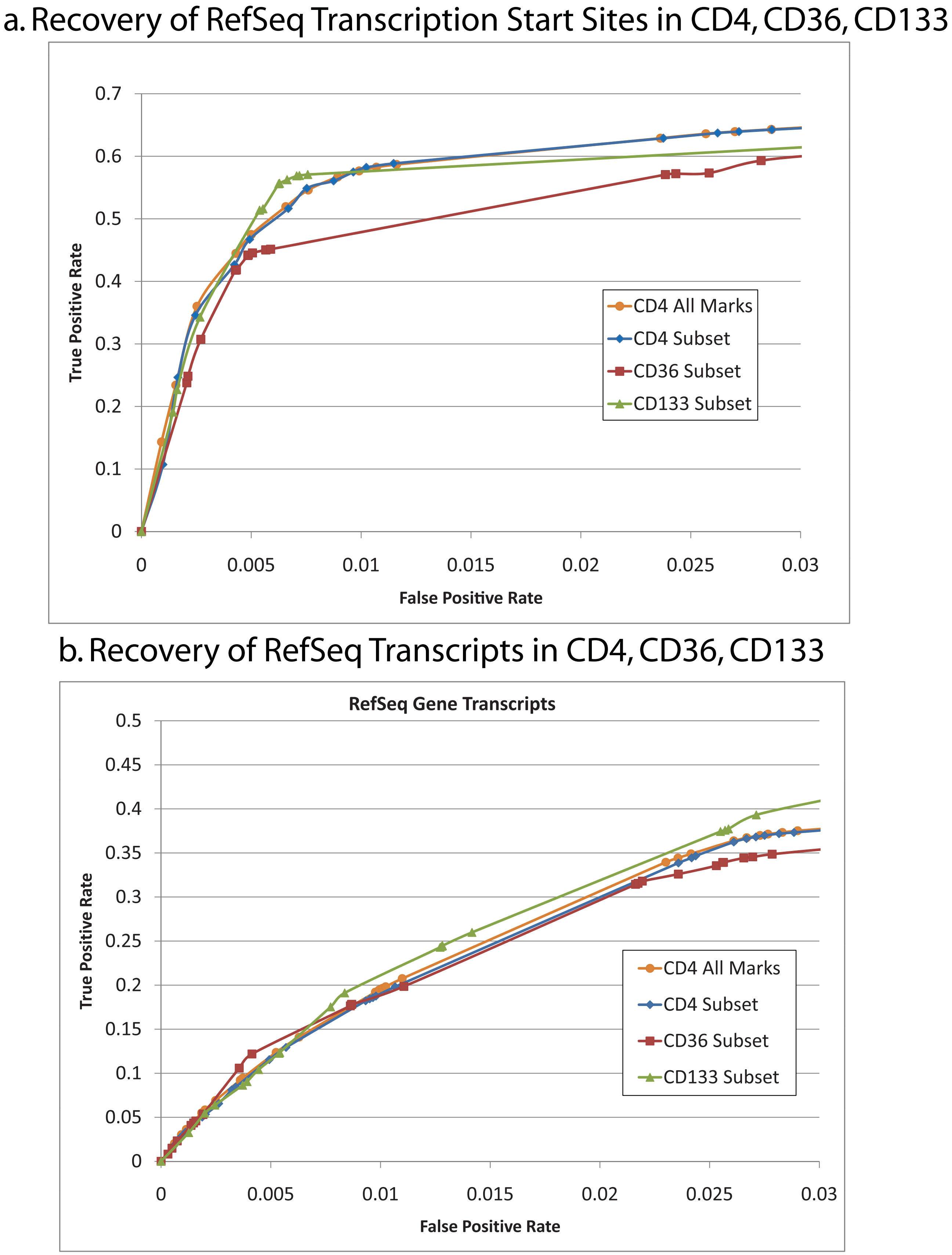
Supplementary Figure 33: Overlap with Expressed Sequence Tags (ESTs) and Predicted New Exons. Independent experimental and comparative information provides support that a significant fraction of false positives in Figures 5a and 5b are genuine novel unannotated TSS and transcribed regions currently missing from RefSeq. **a.** In the left table, for each state is shown the percentage of the state falling within a RefSeq annotated transcribed region. In the center table is the percentage of 200 bp intervals associated with a state intersecting expressed sequence tag (EST), and the percentage of the state that overlaps with EST data when restricting to RefSeq transcribed regions. Bottom row indicates the genome-wide percentages of these quantities. This table shows that chromatin states are predictive of function even outside known annotations, and many non-RefSeq-annotated regions falling in transcribed states are indeed supported by EST data for being transcriptionally active. In the right table are fold enrichment for protein-coding exons predicted by evolution conservation using 29 mammals (Lin and Kellis, in preparation), and the same fold enrichment specifically outside RefSeq exons. The shift in enrichment from transcribed states (states 21-27) to repressed and low-expression promoter states (state 4 and 5) suggests that novel exons missing from RefSeq are likely to be short and of low expression. **b**. Example of candidate novel exon repressed in CD4 T-cells. Highly-conserved protein-coding exon (black) is annotated in repressed promoter state 4 (red) and surrounded by repressed states (grey), likely due to its short length and repression in CD4 T-cells. **c**. Example of candidate novel exon active in CD4 T-cells. This evolutionarily-predicted new protein-coding exon (black) lies in a low-expression promoter state (state 5) and is associated with several other promoter states (2 and 3) and flanked by active intergenic regions.



Supplementary Figure 34: Expression enrichments for numerous cell types. Average expression level corresponding to each state across 158 microarray experiments[7]. Experiments are clustered hierarchically and ordered based on the optimal leaf ordering[5]. The name of every fifth experiment is listed for compactness. The two CD4 T experiments are boxed and marked. Heatmap shows that the chromatin states defined here based on CD4 T chromatin marks show the highest and lowest expression levels in CD4 T and related cell types. Average expression levels of each state were computed as described for the CD4 T data in the Online Methods section, but all replicates were kept separate.

state	Transcript top 2000	Transcript bottom 2000	5' end top 2000	5' end bottom 2000
1	0.8	1.4	14.4	8.7
2	0.5	1.8	6.4	6.5
3	0.4	2.6	5.9	21.0
4	0.3	4.5	4.8	59.4
5	2.0	1.8	131.0	41.8
6	1.8	1.4	140.2	27.1
7	3.2	1.0	210.7	20.7
8	8.1	1.1	169.7	7.5
9	9.9	1.0	39.1	5.0
10	5.5	1.9	17.5	2.1
11	7.0	1.4	9.1	2.5
12	7.3	2.4	5.0	0.6
13	6.5	1.9	3.8	0.8
14	4.6	1.4	2.1	0.4
15	5.0	1.6	1.6	0.4
16	3.3	1.6	0.9	0.3
17	3.0	2.4	1.2	0.4
18	2.4	2.0	0.9	0.3
19	1.7	2.0	0.6	0.5
20	1.9	1.8	2.8	3.1
21	3.4	2.1	6.5	3.3
22	3.8	2.6	4.6	2.1
23	1.9	2.9	1.4	3.5
24	2.1	1.7	1.1	1.0
25	1.5	1.7	0.9	0.8
26	1.0	2.0	0.6	0.8
27	8.6	0.9	8.8	1.6
28	0.5	0.9	1.8	1.7
29	1.0	1.8	0.7	1.1
30	1.4	1.5	1.7	1.3
31	0.3	2.1	0.7	2.3
32	0.3	1.9	0.3	1.2
33	0.6	2.2	0.5	1.6
34	1.0	1.7	0.7	1.4
35	0.2	2.1	0.1	1.5
36	0.2	2.2	0.1	1.3
37	0.2	2.6	0.1	1.0
38	0.1	2.5	0.1	1.8
39	0.3	2.6	0.4	1.5
40	0.2	1.1	0.3	0.5
41	0.0	1.9	0.0	0.2
42	0.0	1.9	0.0	0.3
43	0.1	3.1	0.0	0.7
44	0.1	3.2	0.0	0.9
45	0.1	3.0	0.1	4.6
46	0.2	3.5	0.0	1.2
47	0.1	2.0	0.0	0.4
48	0.1	0.6	0.2	1.1
49	0.1	0.4	0.0	1.0
50	0.1	0.5	0.0	0.0
51	0.1	0.5	0.0	0.0
% Overall	1.53	3.36	0.011	0.012

Supplementary Figure 35: State enrichments for most expressed and most repressed genes. For each state is shown the enrichment for the 2000 affymetrix probe sets with the highest and lowest expression in CD4 T cells[7], and the enrichments based only on the intervals which intersect a 5′ end of a probe set. The bottom row indicates the percentage of 200 bp intervals each category represents. The figure indicates that both highly- and lowly-expressed genes show specific state enrichments, though these were stronger for genes of higher expression. The genomic coordinates of probe sets were obtained from the UCSC genome browser[20].



Supplementary Figure 36. Transcription Start Site (TSS) and Transcribed Region recovery in additional cell types. This figure shows the recovery of RefSeq TSS and genes when applying the CD4 model learned on 41 marks to the subset of 10 marks inferred from the CD36 and CD133 data[25]. The states are ordered in the same way as used in the analysis based on the CD4 model with all the marks as shown in **Figure 5**. The figure indicates the functional enrichment of these states are relatively robust across these cell types. The recovery of TSS in CD36 compared to CD133 was somewhat lower, consistent with the previous observation[25] that in the more differentiated CD36 cells fewer repressed gene promoters are marked with H3K4me3.

a. % State overlap at varying distances within TSS

state	TSS +-2kb	TSS +-5kb	TSS +-10kb	TSS +-20kb	TSS +-50kb	TSS +-100kb
1	50.7	54.4	59.5	67.3	81.3	90.6
2	41.0	45.8	51.1	60.0	76.0	87.1
3	51.6	57.1	62.7	70.3	82.8	91.0
4	56.7	65.0	70.8	77.0	85.7	92.1
5	74.3	79.1	82.7	86.8	92.7	96.6
6	77.8	81.0	84.0	88.1	93.7	97.2
7	88.7	90.6	92.2	94.3	97.4	99.0
8	71.5	80.1	83.9	88.7	94.8	98.3
9	40.9	69.0	79.0	85.3	92.9	96.9
10	48.4	56.2	65.0	75.2	89.6	96.6
11	54.0	72.6	80.7	87.6	94.8	98.4
12	8.3	18.7	32.4	51.8	79.4	94.6
13	6.3	21.6	39.6	60.3	84.7	95.7
14	9.2	20.2	32.6	49.1	75.1	91.3
15	3.4	17.8	37.2	59.8	85.1	95.6
16	3.7	17.0	35.1	56.4	82.7	94.9
17	6.6	16.8	27.5	42.2	68.3	86.4
18	2.7	11.3	23.1	39.6	67.1	86.0
19	2.1	10.4	23.1	40.8	68.9	87.5
20	21.1	30.0	39.8	53.8	76.2	89.5
21	27.8	45.4	58.5	71.9	88.4	95.4
22	4.4	19.3	38.6	60.5	84.0	94.1
23	10.2	25.7	43.3	63.4	84.7	94.2
24	1.3	5.9	15.6	32.9	65.3	85.9
25	0.6	2.8	9.6	26.0	60.3	83.7
26	1.0	4.0	10.8	25.9	57.5	80.9
27	3.2	12.2	27.8	51.6	82.7	96.3
28	4.2	12.5	30.6	61.0	85.7	92.7
29	11.0	16.3	24.4	38.2	64.7	81.1
30	15.0	21.4	29.1	42.0	65.9	82.3
31	14.7	20.2	26.9	37.9	59.4	77.0
32	8.5	15.4	23.8	36.6	61.2	79.2
33	14.4	22.8	32.1	45.2	67.1	82.9
34	13.3	24.0	33.4	46.7	69.9	85.3
35	5.3	13.3	21.9	34.3	56.3	74.9
36	4.4	13.1	23.4	36.6	58.8	76.1
37	2.6	8.7	18.3	33.2	58.8	77.0
38	8.4	14.3	21.2	32.3	54.8	73.0
39	4.1	10.5	20.2	35.1	61.8	79.8
40	0.9	2.2	4.3	8.1	16.8	25.6
41	0.3	0.9	1.8	3.7	9.2	18.0
42	0.5	1.0	1.9	3.8	9.2	18.1
43	1.3	3.8	8.5	17.5	38.0	58.3
44	1.5	3.9	8.5	17.3	37.3	57.7
45	11.0	22.4	31.9	43.0	61.1	76.7
46	4.2	9.4	17.8	30.1	53.6	71.3
47	1.6	5.1	10.8	21.6	43.9	63.4
48	2.7	6.0	11.0	20.2	35.8	46.9
49	1.9	3.1	4.9	8.4	17.1	24.7
50	0.9	2.0	3.7	6.8	14.1	21.8
51	1.1	2.1	2.6	5.7	9.9	18.6
% Overall	2.7	6.3	11.6	20.4	37.4	52.3

b. State overlap at varying distances from any gene

state	intergenic	intergenic >2kb away	intergenic >5kb away	intergenic >10kb away	intergenic >20kb away	intergenic >50kb away	intergenic >100kb away
1	55.2	27.9	24.6	20.9	15.6	9.0	4.6
2	56.2	35.2	30.9	26.4	20.2	11.7	6.3
3	51.2	30.3	26.2	21.4	15.9	8.5	4.5
4	47.9	28.7	23.9	20.3	15.4	8.5	4.6
5	37.4	14.5	12.0	9.7	7.2	3.6	1.8
6	39.1	13.3	10.8	8.8	6.0	2.9	1.3
7	30.4	7.2	5.9	4.8	3.3	1.3	0.4
8	12.5	6.8	5.5	4.9	3.7	1.2	0.3
9	10.7	7.7	6.3	5.6	4.3	2.0	0.6
10	12.7	7.1	5.6	4.5	3.0	1.1	0.4
11	7.9	4.4	3.5	2.9	1.8	0.6	0.2
12	7.2	5.9	5.0	4.2	3.1	1.1	0.2
13	5.3	4.8	4.2	3.5	2.6	0.9	0.3
14	17.7	14.0	10.7	8.2	5.8	2.6	1.0
15	5.1	4.7	3.9	3.2	2.3	0.8	0.1
16	7.8	7.0	5.6	4.4	3.0	1.2	0.3
17	6.0	4.8	3.6	2.5	1.8	0.6	0.2
18	6.7	5.7	4.4	3.2	2.3	0.9	0.3
19	11.0	9.7	7.6	5.6	3.9	1.9	0.8
20	29.6	20.7	16.0	12.4	8.6	4.4	2.2
21	16.8	8.7	5.1	3.3	2.1	1.0	0.4
22	7.6	5.0	3.6	2.9	2.2	1.2	0.6
23	18.6	11.4	7.2	4.9	3.1	1.4	0.6
24	9.9	6.6	4.1	2.4	1.3	0.6	0.2
25	9.0	5.8	4.1	3.0	2.0	1.0	0.5
26	14.4	11.0	7.4	4.7	3.0	1.5	0.6
27	12.3	4.7	1.7	0.6	0.4	0.2	0.0
28	23.3	17.6	14.1	10.8	7.0	3.9	2.5
29	55.6	46.6	41.9	36.2	27.9	16.7	9.0
30	54.2	42.2	36.9	31.7	25.2	15.7	9.1
31	68.1	56.2	51.0	45.0	36.8	22.9	13.3
32	57.7	51.2	45.3	39.3	31.1	18.7	9.9
33	48.6	39.4	33.1	26.9	20.1	11.5	6.4
34	48.4	38.1	30.4	24.5	18.4	10.4	5.6
35	65.5	61.0	54.6	47.9	38.9	24.4	14.2
36	67.2	63.1	56.0	48.0	38.1	23.5	13.8
37	61.8	58.9	53.1	44.5	33.4	19.2	11.1
38	68.4	61.9	56.5	50.4	41.6	26.7	16.0
39	63.5	58.6	51.7	43.0	32.0	17.7	9.8
40	90.3	89.5	88.4	86.6	83.5	76.7	69.6
41	78.4	78.1	77.6	76.7	75.1	70.4	63.4
42	77.3	76.9	76.4	75.5	73.8	69.1	62.0
43	67.1	65.8	63.5	59.6	52.3	37.6	24.7
44	65.5	64.2	62.0	58.1	51.2	37.4	24.6
45	65.9	59.5	53.1	46.6	38.0	24.6	14.0
46	57.8	54.4	50.6	44.3	37.1	24.6	15.2
47	71.2	69.5	66.2	60.4	50.8	34.3	21.9
48	83.4	79.9	75.0	69.2	62.5	53.4	46.4
49	91.3	89.7	88.4	86.7	83.4	76.3	68.1
50	94.2	93.3	92.4	90.6	86.9	79.4	69.7
51	93.9	93.2	93.1	92.0	87.5	81.1	74.4
% Overall	63.7	61.6	59.1	55.7	50.7	41.8	34.0
% of Pol2	28.5	14.9	11.4	9.1	6.9	4.1	2.5

c. Pol2 detection away from genes

all	intergenic	intergenic >2kb away	intergenic >5kb away	intergenic >10kb away	intergenic >20kb away	intergenic >50kb away	intergenic >100kb away
51.6	51.9	47.6	46.1	44.5	41.3	36.6	32.9
18.1	17.6	15.2	14.1	13.4	12.2	10.7	9.2
11.2	11.0	8.9	8.2	7.8	7.2	6.4	6.0
1.9	1.6	1.3	1.2	1.2	1.1	1.0	0.7
53.5	52.9	42.1	40.9	38.9	36.3	33.3	26.8
69.2	67.6	63.8	62.5	60.8	57.3	52.2	44.3
88.0	89.4	86.9	86.7	85.8	85.1	80.7	80.9
62.1	63.7	62.3	61.5	60.6	60.5	50.5	60.4
34.7	39.1	33.5	32.4	33.0	30.8	32.1	19.3
46.6	47.8	49.4	48.1	45.0	41.6	35.0	36.7
30.9	34.0	27.2	27.6	27.3	27.5	21.1	15.9
18.0	13.9	11.8	11.4	10.3	9.7	7.0	3.6
10.2	10.1	8.5	8.0	7.7	6.8	4.8	4.9
6.9	4.7	4.2	3.4	3.3	3.6	2.6	2.0
5.0	5.1	4.2	3.4	3.1	2.9	2.0	2.0
0.6	0.6	0.5	0.4	0.4	0.4	0.4	0.7
3.5	5.7	5.0	4.2	3.3	3.3	3.8	0.1
1.6	1.8	1.7	1.3	1.3	1.3	0.9	1.0
0.4	0.5	0.4	0.3	0.4	0.4	0.3	0.4
14.5	15.8	12.8	10.2	9.6	9.1	8.6	8.7
15.1	20.0	13.4	8.7	9.3	10.9	11.1	3.3
8.5	11.8	8.3	5.4	4.9	5.5	7.4	0.8
1.4	1.9	1.3	0.8	0.6	0.7	0.7	0.6
2.2	4.3	3.3	1.9	1.3	0.9	1.0	0.6
0.8	1.3	0.8	0.4	0.3	0.2	0.2	0.2
0.6	1.1	0.8	0.5	0.3	0.2	0.2	0.1
21.7	40.5	38.9	33.5	26.6	27.7	35.6	5.2
1.3	1.4	1.2	0.7	0.7	0.6	0.7	1.0
13.6	12.6	12.2	11.6	11.4	10.1	9.6	9.1
9.1	8.4	7.7	7.2	7.0	6.6	6.2	5.5
1.2	1.2	1.1	1.0	0.9	0.8	0.6	0.5
1.6	1.5	1.2	1.0	0.9	0.8	0.8	0.9
2.6	2.6	2.2	1.8	1.6	1.5	1.3	1.3
1.5	1.5	1.2	1.0	0.9	0.9	0.8	0.9
0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.1
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1
0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1
4.1	4.2	3.7	3.3	3.0	2.6	2.1	2.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.6
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
14.1	13.6	13.3	13.4	13.2	13.4	13.8	14.6
69.6	69.1	68.9	68.9	69.1	69.3	68.8	67.8

Supplementary Figure 37: State overlap of varying distances from TSS and genes, and detection of PolII away from genes a. The table shows the percentage of each state which is within fixed distances of 2kb, 5kb, 10kb, 20kb, 50kb, and 100kb from a RefSeq transcription start site. The bottom row are the genome-wide percentages. One can see in this table the majority of States 29-33 are more than 20kb from an annotated RefSeq TSS, despite enriching for open chromatin and transcription factor binding making many locations within these states candidates for distal enhancers. While a majority of States 1-3 fall within 20kb of an annotated TSS between 30-40% does not, and are possible candidate distal enhancers. **b.** The table on the left shows the percentage of each state that is outside a RefSeq transcribed region, and at least 2kb, 5kb, 10kb, 20kb, 50kb, and 100kb from a RefSeq transcribed region. c. The percentage of PolII '1' calls for the state outside of the same regions shown on the table at left. This table indicates there is Pol2 present in location away from known genes. Some of this may correspond to unannotated genes, while in other cases it could be for other reasons such as enhancer looping.

state	H3K9me3	H4K20me3	H3K36me3		Satellite repeats	ZNF Gene Fold	Combination
47	32	2	2		0.5	0.8	**H3K9me3**
48	12	38	1		63.4	10.8	**+H4K20me3**
28	43	75	69		0.7	111.8	**+H3K36me3**
25	2	0	60		0.1	3.6	**H3K36me3 alone**

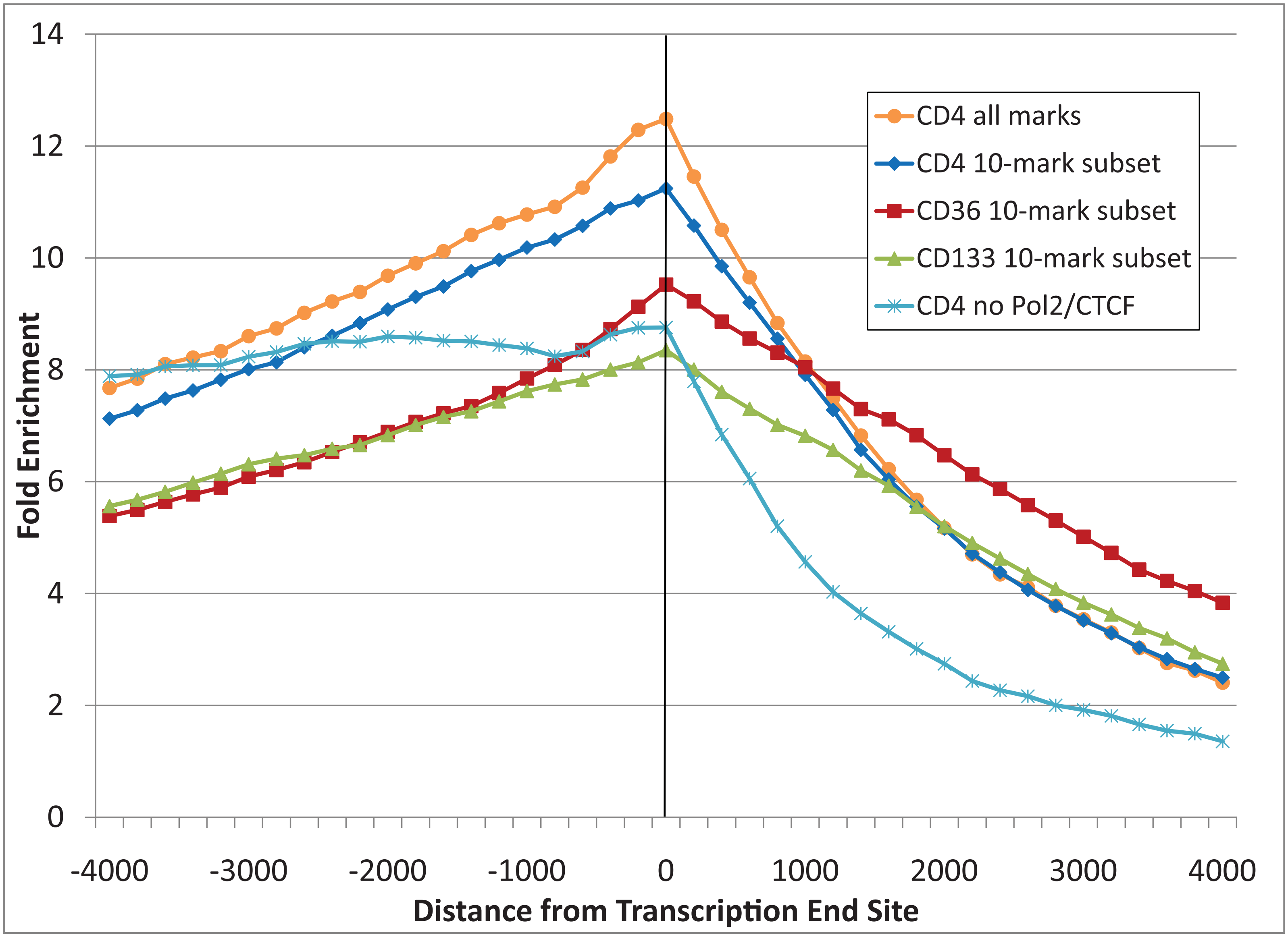
Supplementary Figure 38: Example of combinatorial mark relationships. In this example the enrichment for Satellite repeats occurs when H3K9me3 and H4K20me3 co-occur in a State 48 without H3K36me3, but not necessarily with H3K36me3 (State 28), for H3K9me3 alone (State 47), or for H3K36me3 alone (State 25). The enrichment for ZNF genes is an order of magnitude greater in State 28 which is also associated with H3K36me3 as compared to State 48 which is not.

																																																					By State	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	State	Sensitivity	PPV
1	40	21	11	0	2	9	3	1	0	4	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	40%	42%
2	15	42	20	1	1	2	2	0	0	2	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	3	1	5	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	42%	44%
3	7	15	49	6	1	4	1	1	1	3	3	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	1	3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	49%	49%
4	0	1	6	72	2	1	0	1	3	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	4	72%	74%
5	1	2	3	8	41	13	8	10	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	41%	43%
6	6	2	7	2	15	35	18	7	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	35%	38%
7	3	2	2	0	11	23	47	9	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	47%	44%
8	3	2	4	2	13	9	9	35	15	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	35%	34%
9	0	1	2	6	9	3	1	13	44	2	6	1	1	2	1	1	0	0	0	1	2	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	44%	48%
10	7	5	3	0	0	1	1	3	3	34	26	5	2	0	0	0	0	0	0	5	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	34%	38%
11	2	2	2	0	0	0	0	2	3	16	62	3	2	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	62%	49%
Promoter states																																																				Average-by-state	46%	46%
																																																				Genomic average	48%	48%
12	0	0	0	0	0	0	0	0	0	1	2	38	28	1	3	0	6	1	0	4	10	1	0	0	0	0	0	0	2	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	38%	33%
13	0	0	0	0	0	0	0	0	0	0	1	12	59	1	6	0	7	2	0	1	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	59%	58%
14	0	0	0	0	0	0	0	0	0	0	0	1	2	16	8	6	2	7	5	2	3	2	2	3	1	2	2	0	3	3	1	5	7	8	2	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	14	16%	24%
15	0	0	0	0	0	0	0	0	0	0	0	1	4	3	46	9	2	9	6	0	3	6	5	1	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	46%	54%
16	0	0	0	0	0	0	0	0	0	0	0	0	0	2	8	36	0	6	21	0	0	1	7	0	0	3	1	0	0	0	0	0	0	3	1	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	36%	54%
17	0	0	0	0	0	0	0	0	0	0	0	6	18	1	8	1	47	7	1	2	2	1	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	47%	50%
18	0	0	0	0	0	0	0	0	0	0	0	0	1	1	8	5	2	51	9	0	0	1	2	4	1	5	1	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	51%	51%
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	3	56	0	0	0	2	1	1	12	0	0	0	0	0	0	0	2	2	6	7	0	0	0	0	0	1	0	0	0	0	0	0	0	0	19	56%	57%
20	1	1	1	0	0	0	0	0	1	4	5	8	6	2	1	0	6	1	0	30	7	0	0	1	0	0	0	0	9	1	1	1	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	30%	36%
21	0	0	0	0	0	0	0	0	1	1	2	5	8	2	6	1	2	2	0	2	49	9	5	0	0	0	2	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	49%	42%
22	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6	1	1	2	1	0	3	67	10	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	67%	58%
23	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	4	0	3	8	0	2	9	56	0	0	4	3	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	56%	56%
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	4	0	0	0	0	57	13	14	2	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	57%	58%
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	6	86	3	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	86%	83%
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	2	1	83	0	0	0	0	0	0	0	1	1	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	83%	81%
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	9	5	2	3	3	72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	72%	65%
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	2	1	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	2	0	0	28	76%	71%
Transcribed states																																																				Average-by-state	62%	60%
																																																				Genomic average	71%	70%
29	1	2	1	0	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	8	2	0	0	0	0	0	0	0	32	7	8	7	18	4	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	29	32%	30%
30	0	1	1	1	2	0	0	0	1	0	0	0	0	2	0	1	0	1	1	1	0	0	0	1	0	2	1	0	8	23	4	6	10	12	3	4	3	6	1	0	0	0	1	0	0	0	0	0	0	0	0	30	23%	28%
31	0	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4	2	45	2	5	2	1	2	3	22	2	0	0	0	1	0	0	0	0	0	0	0	0	31	45%	43%
32	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	2	1	1	0	2	2	1	3	1	0	5	3	2	22	15	11	4	12	5	3	0	0	0	0	1	0	0	1	0	0	0	0	0	32	22%	22%
33	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	2	3	1	4	1	0	0	1	0	1	0	0	7	4	3	9	51	2	1	1	1	2	0	0	0	0	0	0	0	1	0	0	0	0	0	33	51%	49%
34	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	6	0	0	0	1	1	1	8	1	0	1	3	1	5	1	37	5	9	10	3	1	0	0	0	2	0	0	0	0	0	0	0	0	34	37%	43%
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	1	1	5	0	0	0	1	0	2	1	4	22	19	27	3	1	0	0	0	8	1	1	0	0	0	0	0	0	35	22%	23%
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	3	41	26	1	0	1	0	0	17	1	1	0	0	0	0	0	0	36	41%	42%
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	7	67	1	0	3	2	0	14	2	0	0	0	0	0	0	0	37	67%	66%
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	4	1	1	1	2	3	7	64	2	1	1	0	10	2	1	0	0	0	0	0	0	38	64%	60%
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	5	0	0	0	1	2	1	1	2	3	8	23	10	9	2	1	0	23	3	2	1	0	0	0	0	0	39	9%	10%
Active Intergenic states																																																				Average-by-state	37%	38%
																																																				Genomic average	57%	56%
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	94	2	0	1	0	0	0	0	0	0	0	0	40	94%	90%
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	90	2	3	0	0	0	0	1	0	0	0	41	90%	92%
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	32	56	6	0	0	0	0	1	0	0	0	42	56%	57%
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7	0	0	1	3	0	83	2	1	0	0	0	0	0	0	43	83%	83%
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	2	9	1	0	1	3	0	21	57	3	0	1	0	0	0	0	44	57%	58%
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	7	4	83	0	0	0	0	0	0	45	83%	81%
Repetitive states																																																				Average-by-state	61%	60%
																																																				Genomic average	86%	86%
46	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	1	2	1	3	1	1	3	0	0	1	1	1	5	4	3	5	7	13	3	1	1	2	1	11	3	4	16	1	0	0	0	0	46	16%	21%
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	2	0	3	1	1	0	85	5	1	0	0	47	85%	85%
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	1	4	0	0	2	27	2	5	0	1	0	8	43	2	0	0	48	43%	46%
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	1	0	0	1	4	1	1	0	0	0	4	23	53	4	0	49	53%	47%
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	17	63	6	50	63%	54%
51	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	87	51	87%	81%
Repetitive states																																																				Average-by-state	65%	64%
																																																				Genomic average	69%	69%
All states																																																				Average-by-state	52%	52%
																																																				Genomic average	77.2%	76.9%

Supplementary Figure 39: Chromatin State Recovery with Subset of 10 Chromatin Marks. (**Left**) The figure shows a "confusion matrix" between posterior assignments when using a subset of 10 marks (H3K4me1, H3K4me3, H3K9me1, H3K9me3, H3K27me1, H3K27me3, H3K36me3, H4K20me1, H2A.Z, and PolII), previously used in (Cui et al, 2009)[25] as compared to all 41 marks. An entry in the grid indicates the percentage of the row state based on all the marks assigned to the column state when using the subset of 10 marks to determine posterior state assignments (see **Online Methods**). For example, only 9% of the CTCF/insulator island state (state 39) is correctly assigned using this set of marks (which does not include CTCF) while 23% would be assigned to states 37 and 43 each. (**Right**) The sensitivity and positive predictive-value of the state assignments based on the subset of marks compared to the full set of marks for each state. Also shown for each group of states and overall are the average sensitivity and positive predictive value where the averages are given both by weighting based on state size and by considering all states equally.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1	94	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	93	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	93	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	95	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	3	4	88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	4	1	0	0	91	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	4	0	0	0	0	1	93	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	1	0	94	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	1	0	0	1	93	0	0	0	1	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	0	95	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	1	0	0	95	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	1	0	97	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	98	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	1	0	0	0	0	97	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	96	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	93	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	96	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	94	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	5	0	0	1	6	3	3	8	69	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	95	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	95	0	0	1	0	0	3	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	93	4	0	0	0	1	0	0	0	0	0	0	0	0
39	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	5	0	0	0	0	0	0	1	2	1	11	28	13	10	1	1	0	19	2	2	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	99	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	98	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	98	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	97	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	97	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	98	1	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	96	1
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	95

Supplementary Figure 40: Chromatin State Recovery with all marks except CTCF and Pol2. The figure shows the percentage of each state that would be recovered when excluding CTCF and Pol2, but using the 38-histone modifications and H2A.Z to determine the posterior state assignments. We find that nearly all individual states are recovered at rates greater than 90%. The only three exceptions are insulator state 39 that heavily relies on CTCF (10% recovery), transcription end state 27 that relies on Pol2 (69% recovery), and promoter state 5 (88% recovery).



Supplementary Figure 41: Enrichment of State 27 Relative to the Transcription End Sites across Cell Types. The figure shows that state 27 learned based on 41 marks in CD4 T cells still shows enrichment relative to the transcription end site in two additional cell types, CD36 and CD133, based on a subset of 10 marks[25]. Even though the model was not learned with CD36 and CD133 the state still show an enrichment profiling peaking over the transcription end site. Also shown is the enrichment of the state based on the same subset of 10 marks in CD4 T as well as all marks in CD4 T cells except CTCF and PolII.

Supplementary References

1. Schones, D. E. *et al*. Dynamic regulation of nucleosome positioning in the human genome. *Cell* 132, 887-898 (2008).
2. Elkan, C. *Using the triangle inequality to accelerate k-means* (*Twentieth International Conference on Machine Learning (ICML'03)* Ser. 20, 2003).
3. Komarek, P. &Moore, A. W. *Making logistic regression a core data mining tool with tr-irls* (Proceedings of the 5th International Conference on Data Mining, 2005).
4. Hon, G., Ren, B. & Wang, W. ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput. Biol.* **4**, e1000201 (2008).
5. Bar-Joseph, Z., Gifford, D. K. & Jaakkola, T. S. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* **17**, S22 (2001).
6. Zang, C. *et al*. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958 (2009).
7. Su, A. I. *et al*. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6062-6067 (2004).
8. Zeller, K. I. *et al*. Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proceedings of the National Academy of Sciences* **103**, 17834 (2006).
9. Lupien, M. *et al*. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-970 (2008).
10. Lin, C. Y. *et al*. Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet* **3**, e87 (2007).
11. Valouev, A. *et al*. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nature methods* **5**, 829-834 (2008).
12. O'Geen, H. *et al*. Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet.* **3**, e89 (2007).
13. Lim, C. A. *et al*. Genome-wide mapping of RELA (p65) binding identifies E2F1 as a transcriptional activator recruited by NF-κB upon TLR4 activation. *Mol. Cell* **27**, 622-635 (2007).
14. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497 (2007).
15. Wei, C. L. *et al*. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207-219 (2006).
16. Yang, A. *et al*. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol. Cell* **24**, 593-602 (2006).
17. Robertson, G. *et al*. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods* **4**, 651-657 (2007).
18. Rada-Iglesias, A. *et al*. Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. *Genome Res.* **18**, 380 (2008).
19. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732-1741 (2009).
20. Karolchik, D. *et al*. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* **36**, D773-9 (2008).
21. Furey, T. S. & Haussler, D. Integration of the cytogenetic map with the draft human genome sequence. *Hum. Mol. Genet.* **12**, 1037-1044 (2003).
22. Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**, 191 (2006).
23. Wang, Z. *et al*. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019-1031 (2009).
24. Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-3.0* (1996).
25. Cui, K. *et al*. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell. Stem Cell.* **4**, 80-93 (2009).