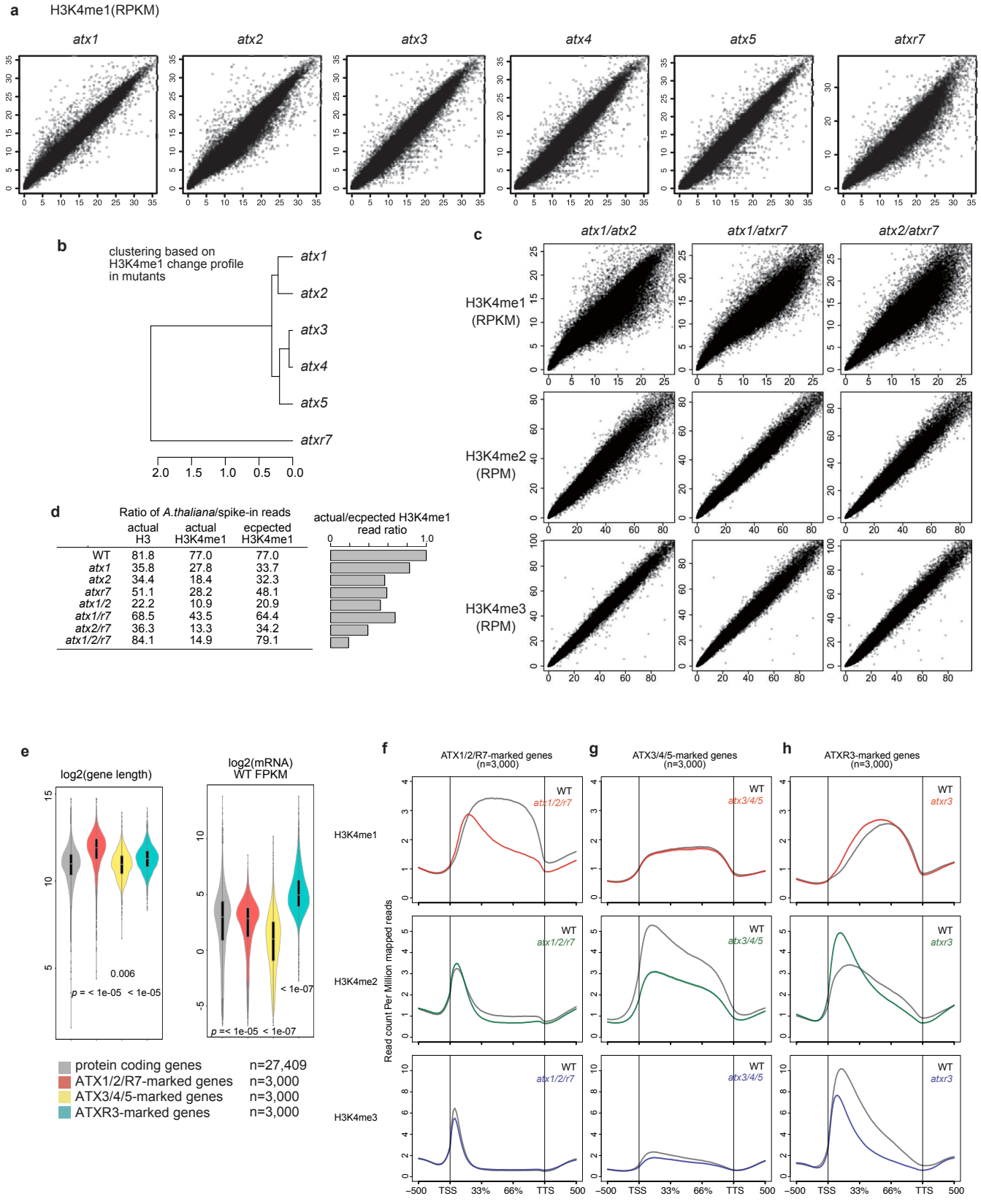
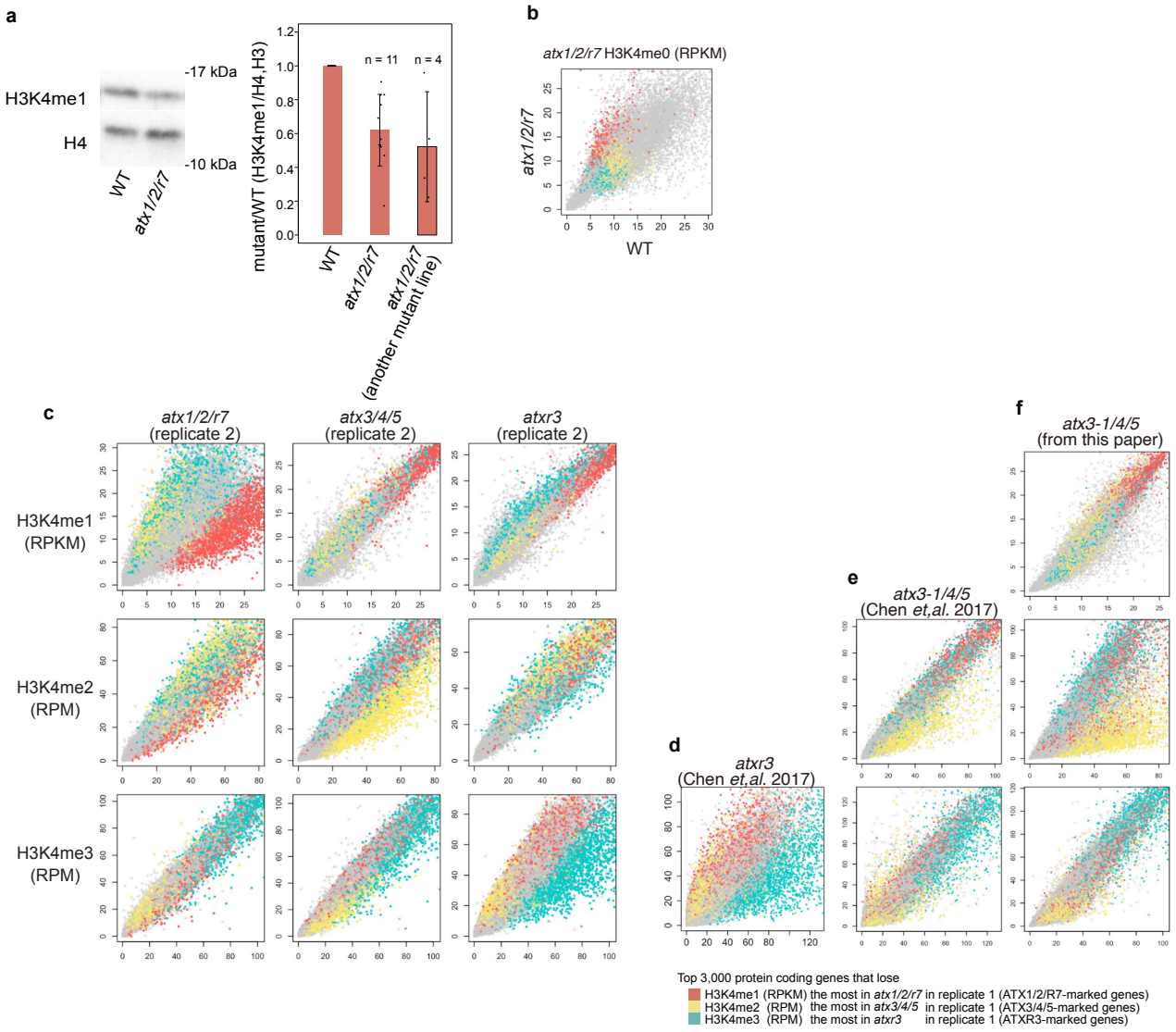


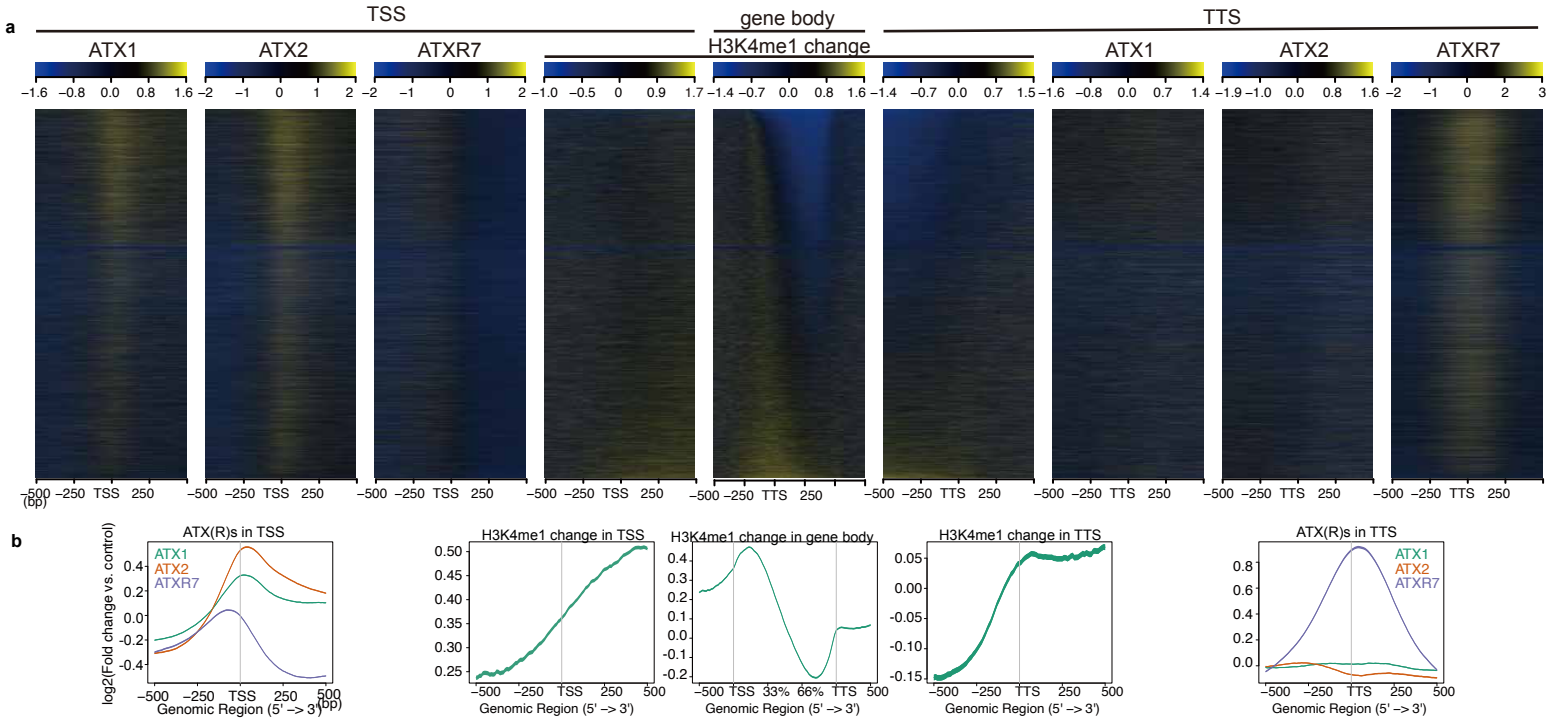
	actual H3	actual H3K4me1	ecpected H3K4me1
WT	81.8	77.0	77.0
atx1	35.8	27.8	33.7
atx2	34.4	18.4	32.3
atxr7	51.1	28.2	48.1
atx1/2	22.2	10.9	20.9
atx1/r7	68.5	43.5	64.4
atx2/r7	36.3	13.3	34.2
atx1/2/r7	84.1	14.9	79.1

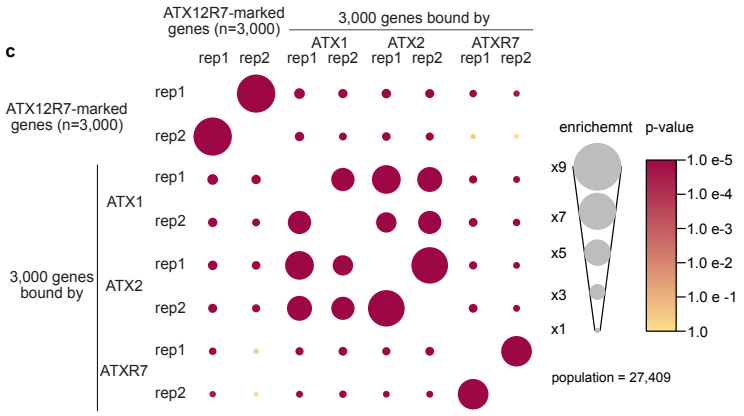
Supplementary Figure 1 Impacts of *atx(r)* mutations on H3K4me

(a) ChIP-seq of H3K4me1 in 6 mutants (y-axis) did not show large change compared to WT (x-axis). Each dot represents one gene and axes are trimmed to 0.98 quantiles. **(b)** *atx1,2,3,4,5* and *atxr7* mutants were clustered based on the H3K4me1 pattern detected in ChIP-seq (see Methods). **(c)** ChIP-seq for H3K4me1, H3K4me2, H3K4me3 in *atx1/2, atx1/atxr7 and atx2/atxr7* double mutants (y-axis) compared to WT (x-axis). For all three mutants, H3K4me1 shows the most prominent reduction. **(d)** Spike-in normalization shows mutations of *atx1*, *atx2,* and *atxr7* synergistically cause reduced H3K4me1. The raw read counts are provided in Supplementary Data 2. **(e)** ATX1/2/R7-, ATX3/4/5-, and ATXR3-marked genes have distinct characteristics in gene length and expression level. The center line of violin plot, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. Numbers within the figure represent p-values calculated with two-sided Dunnett's test comparing against protein coding genes. **(f)-(h)** Metaplots illustrating the averaged distribution of H3K4me1,2,3 over ATX1/2/R7-marked **(f)**, ATX3/4/5-marked **(g)**, and ATXR3-marked genes **(h)**.

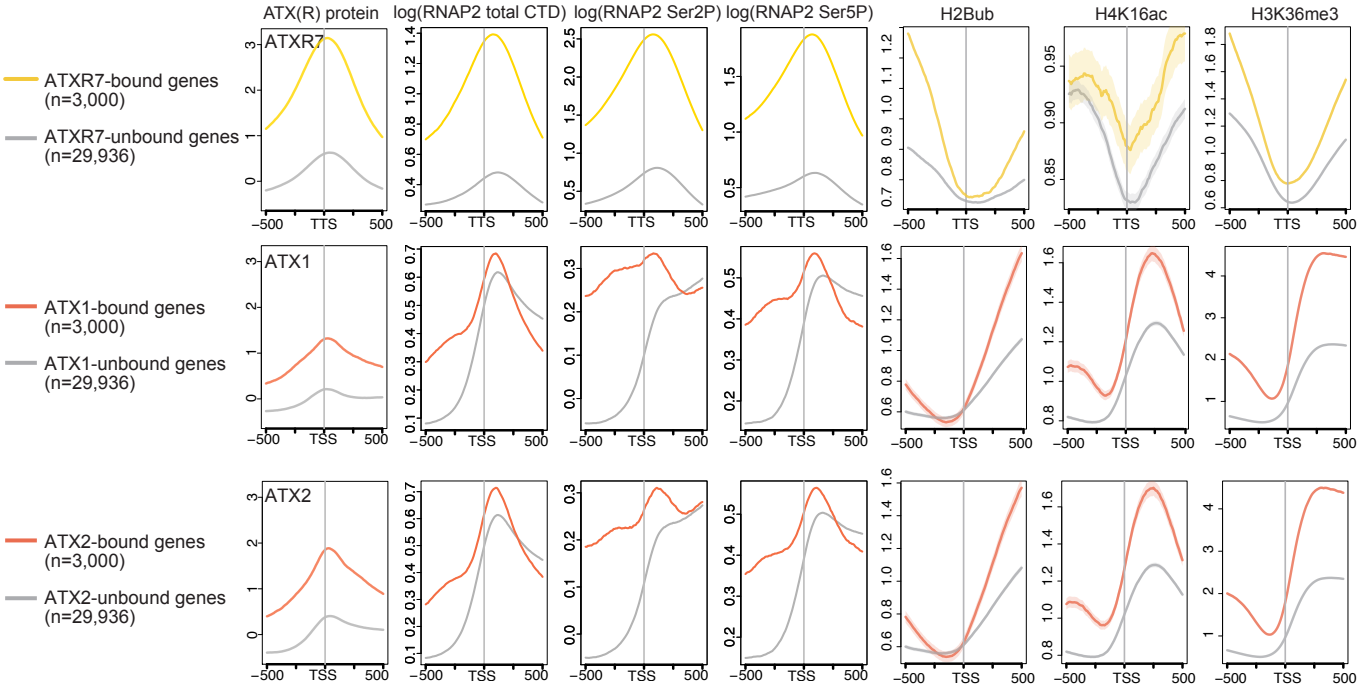


Supplementary Figure 2 Effects on H3K4me in mutants agree with previous reports and replicates.
(a) Western blot for H3K4 methylations on bulk histone extracted from the mutants. Left panel shows example blots for H4 and H3K4me1 obtained from the same membrane. Right panel shows H3K4me1 signals normalized to H3 or H4 histone signals in two lines of *atx1/2/r7* mutants compared to WT. Data are presented as mean±SD and each dot represents data of each replicate. **(b-f)** Mutant (y-axis) ~ WT (x-axis) comparison of H3K4me profiles. In all panels, genes are colored with the same rules defined in Figure. 1d. **(b)** ChIP-seq targeting H3K4me0 shows that in *atx12r7*, hypo-H3K4me1 genes gain H3K4me0, consistent with the view that ATX1,2, and ATXR7 are responsible for H3K4me0-to-H3K4me1 conversion. **(c)** Biological replicates of H3K4me ChIP-seq in *atx1/2/r7, atx3/4/5 and atxr3*. **(d-f)** Previously reported ChIP-seq datasets in *atxr3* (**d**), *atx3-1/4/5* (**e**) mutants, and our datasets on *atx3-1/4/5* (**f**). Our datasets show consistent trends with datasets from ref. 20; in both datasets, *atxr3* mutants lose H3K4me3 in blue-colored genes while *atx3(-1)/4/5* mutants lose H3K4me2 in yellow-colored genes. Axes are trimmed to 0.98 quantiles.

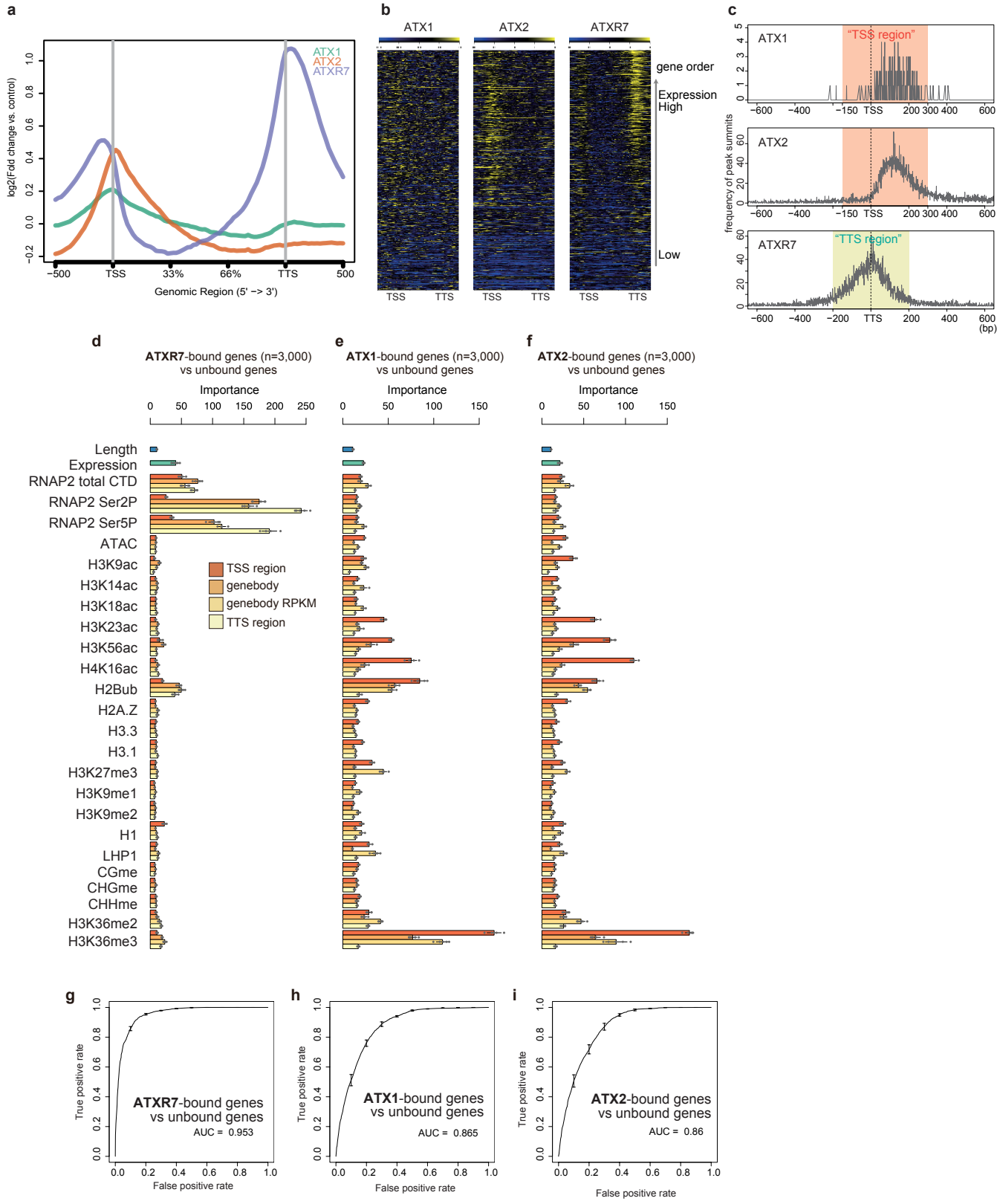




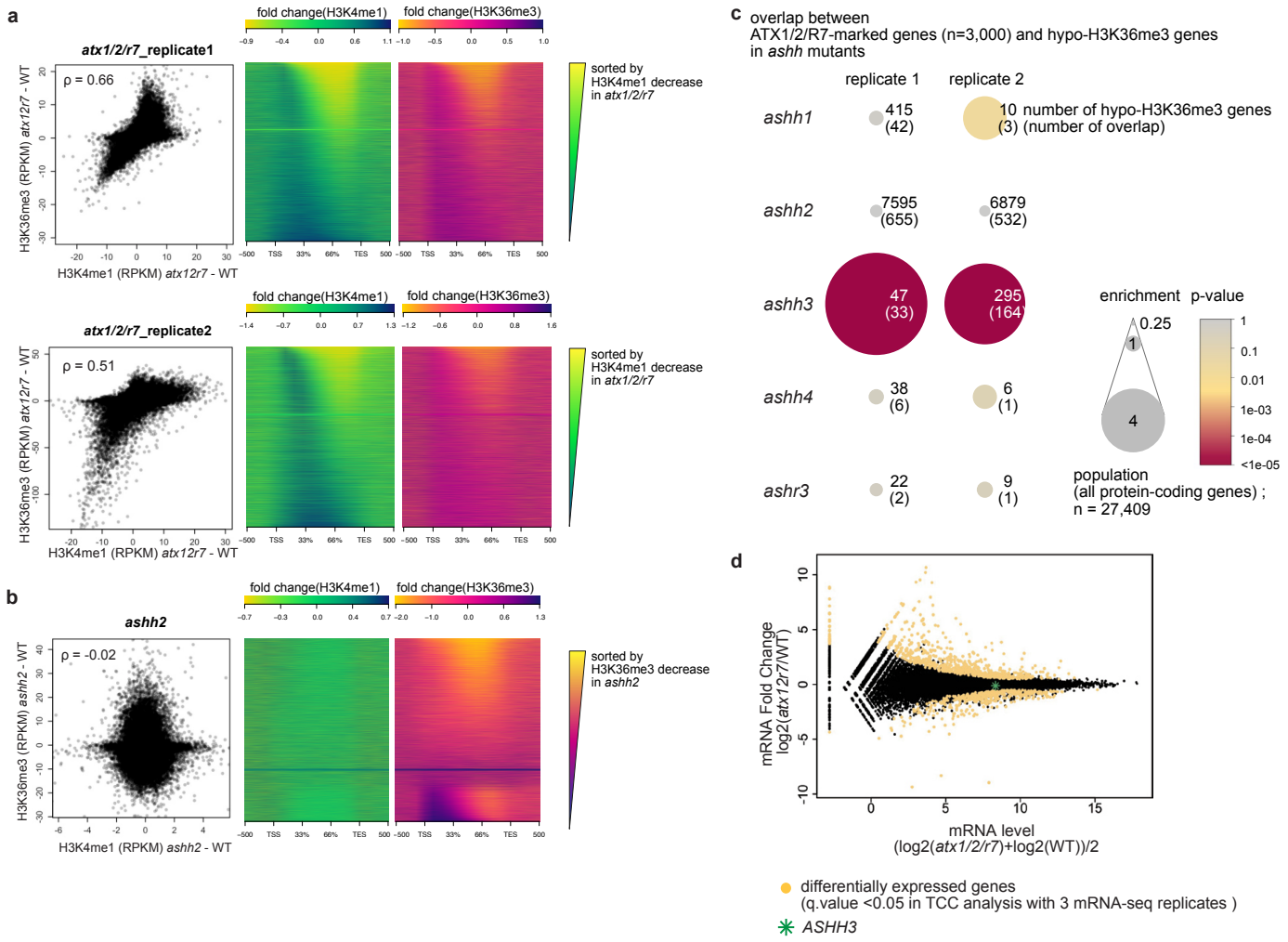
Supplementary Figure 3 ATX1, ATX2, and ATXR7 occupied genes tend to lose H3K4me1 in *atx1/2/r7*.
(a) Heat maps showing H3K4me1 change in *atx1/2/r7* and ATX(R)s distribution. All heat maps are sorted according to the H3K4me1 change in *atx1/2/r7* (the heat map in the middle is the same data as Fig. 1c, shown for reference) so that genes that lost H3K4me1 the most come to the top. **(b)** All genes metaplot corresponding to the panel **(a)** above. **(c)** Overlap between ATX12R7-marked genes and ATX1, ATX2, ATXR7-bound genes. Enrichment is the actual/expected number of overlapped genes, p-values are calculated with hyper-geometric test.



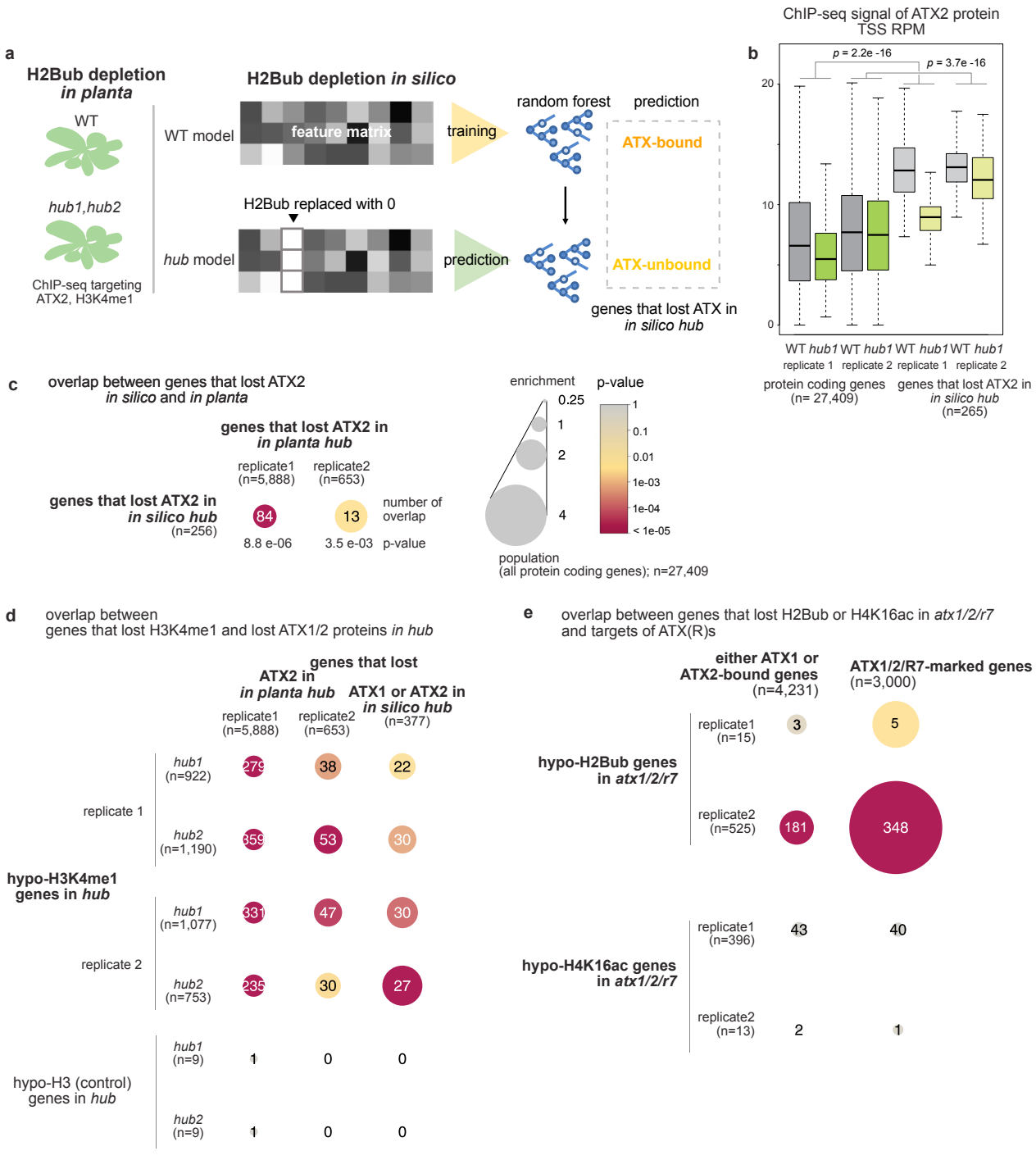
Supplementary Figure 4 Distribution profiles of predictively 'important' chromatin modifications for ATX1, ATX2, or ATXR7 localization.
Averaged distribution profiles of ATX(R)s proteins and features that had high importance in the random forest models to predict ATXR7 or ATX1,2-bound genes (see Fig. 3). Profiles around TTS are shown for ATXR7, and profiles around TSS are shown for ATX1 and ATX2.



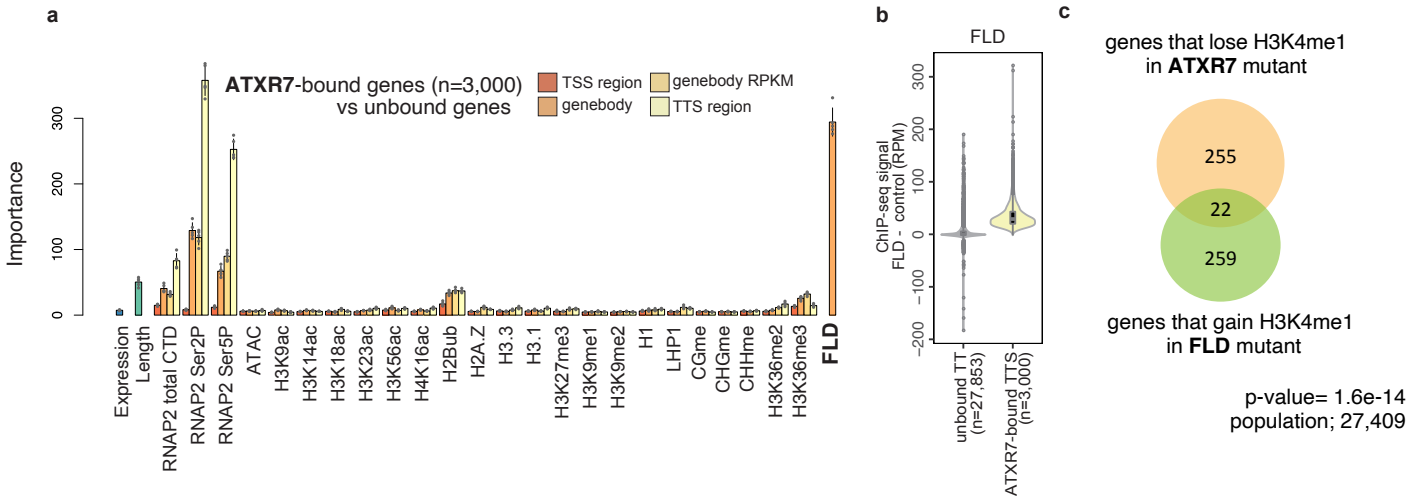
Supplementary Figure 5 Random forest analyses using the biological replicates of ATX(R)s ChIP-seq.
The ChIP-seq data sets of the biological replicates, which were independently grown, were processed in the same manner as in Fig. 2 and Fig. 3. **(a)** Metaplot and **(b)** heat maps illustrating ATX1, ATX2, and ATXR7 distributions in the gene body region corrected with non-transgenic control. The heat maps were sorted so that highly transcribed genes (measured as mRNA-seq in WT) come to the top. **(c)** Position of ATX1 and ATX2 ChIP-seq peaks relative to TSS and ATXR7 peaks relative to TTS (x-axis), visualized as a frequency of peak summits (y-axis). Consistent with the other replicates (Fig. 2), most of ATX1 and ATX2 peaks belong to the 'TSS region', while most of ATXR7 peaks belong to the 'TTS region'. **(d-f)** 'Importance' of features and **(g-i)** ROC plot of random forest model trained to predict ATX1/2/R7-bound genes. Average and standard deviation of the 5 repeats of training are plotted.



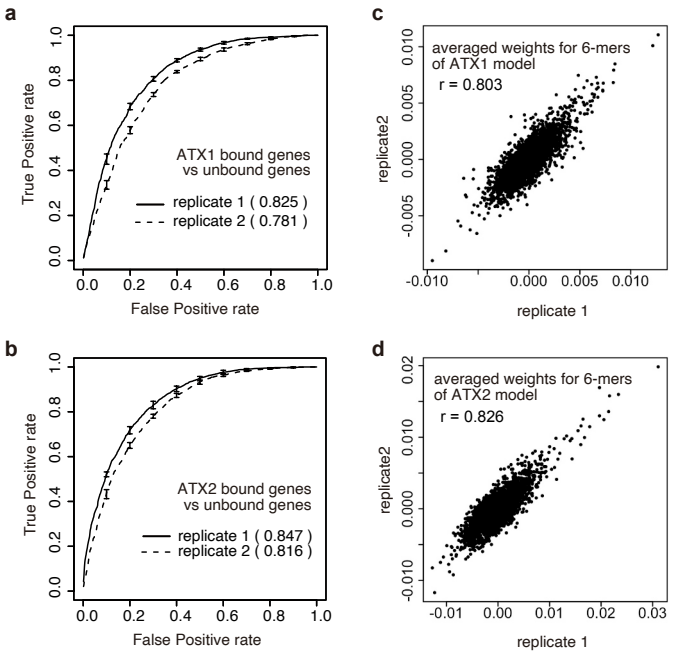
Supplementary Figure 6 H3K4me1 may promote H3K36me3 via ASHH3.
(a) In *atx1/2/r7*, hypo-H3K4me1 correlates with H3K36me3 loss, while change of H3K36me3 levels in *ashh2* **(b)** does not correlate with change of H3K4me1. Each dot in the scatter plot represents one protein coding gene and axes are trimmed to 0.0005 and 0.9995 quantiles. Heat maps in **(a)** are sorted according to the H3K4me1 change in *atx1/2/r7* (the heat map for H3K4me1 of replicate 1 is the same data as Fig. 1c, shown for reference) so that genes that lost H3K4me1 the most come to the top, while heat maps in **(b)** are sorted according to the H3K36me3 change in *ashh2*. ρ is the Spearman's correlation coefficient. **(c)** In search for a methyltransferase that catalyzes H3K36me3 in response to ATX(R)s-catalyzed H3K4me1, we analyzed H3K36me3 change by ChIP-seq in mutants of ASHH2, known major H3K36me3 methyltransferase [31,32] and ASHH1, ASHH3, ASHH4, and ASHR3, whose catalytic domain SET are similar with ASHH2 [17,18]. Overlap between ATX1/2/R7-marked genes and hypo-H3K36me3 genes detected by MACS2 in each *ashh* mutants are shown as bubble plots. The size of a bubble shows enrichment value against all protein-coding genes (n = 27,409) and the p-value represented by color is based on unadjusted one-sided hypergeometric test. *ashh3* lose H3K36me3 at ATX1/2/R7-marked locus, suggesting that ASHH3 may functions downstream of ATX1/2/R7-marked H3K4me1 to mediate H3K36me3. **(d)** mRNA expression of *ASHH3* was not altered in *atx1/2/r7*, precluding the possibility that H3K36me3 change in *atx1/2/r7* was due to reduced expression of *ASHH3*.



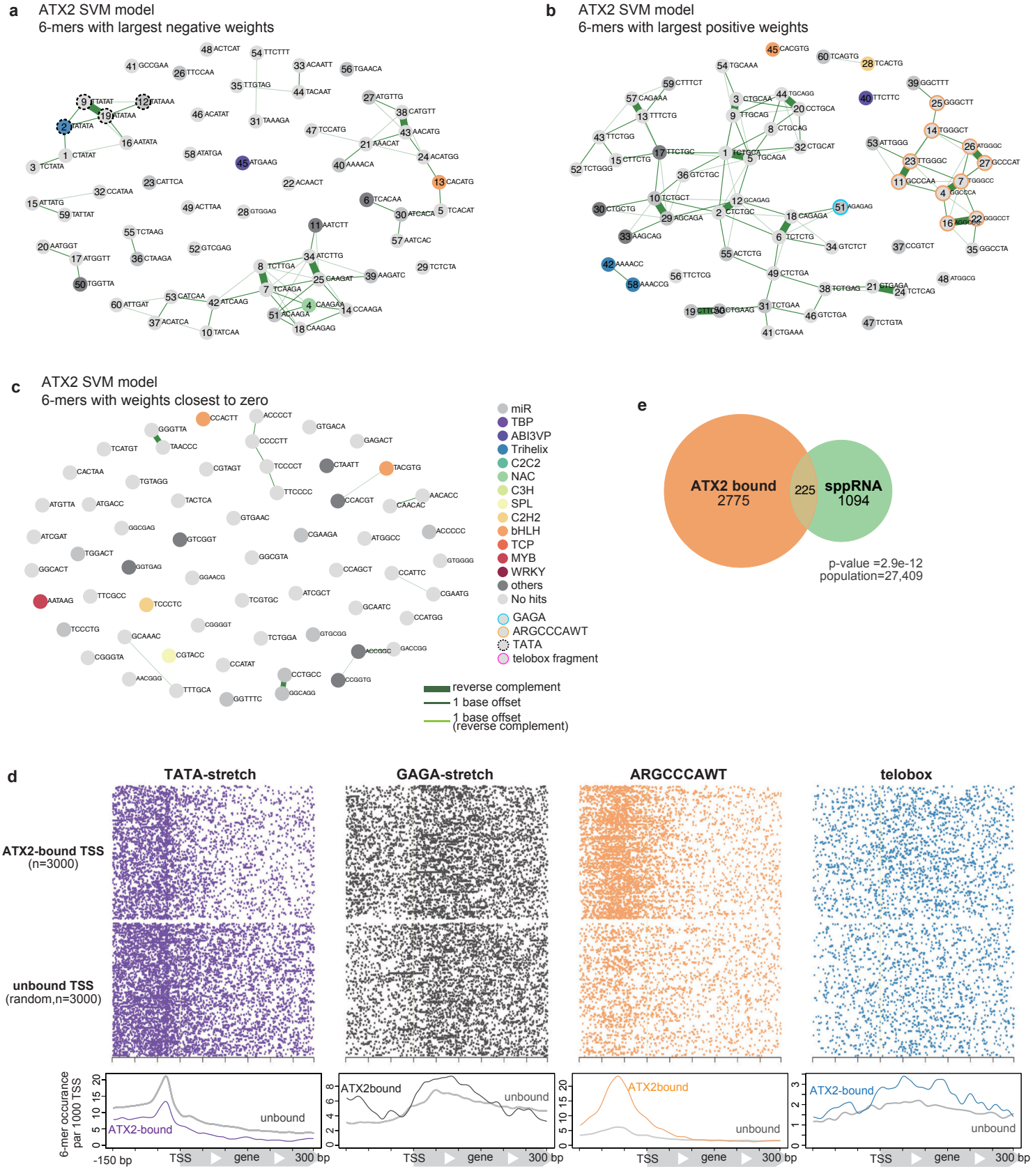
Supplementary Figure 7 Loss of H2Bub results in loss of ATX2 localization and decreased H3K4me1.
(a) Overview of the analysis. To test if H2Bub promotes presence of ATX, ChIP-seq analysis of ATX2 localization was performed in *hub2* and WT (control) background in two replicates. Mutation in either *HUB1 or HUB2* is known to depletes H2Bub to undetectable levels[42,102–104]. ATX localization in the absence of H2Bub was simulated by letting the trained random forest model (WT model) predict ATX-bound genes using a feature table in which H2Bub are replaced with 0 (*hub* model). 'Genes that lose ATX in *in silico hub*' are the genes that are predicted to be ATX-bound in the WT model and unbound in the *hub* model. **(b)** *In planta* ChIP-seq signal over the TSS region of 'genes that lose ATX in *in silico hub*' is largely reduced in *hub2* compared to WT. The ratio of wild-type to hub ATX2 signal was significantly different for all genes vs 'Genes that lose ATX in *in silico hub*', as indicated by the p-value of Wilcoxon's rank sum test. Center line of boxplot, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. **(c)** Genes that are detected to lose ATX2 *in planta* matches *in silico* prediction. Genes that correspond to hypo-ATX regions, which are detected by comparing ATX2 ChIP-seq profile in *hub2* mutant and WT background using peak caller MACS2, significantly overlap with genes predicted to lose ATX in the *in silico hub*. Numbers within the circles are the number of overlaps and the p-value represented by color is based on unadjusted one-sided hypergeometric test. Size and color scales are the same in **(c)**-**(e)**. **(d)** To further test if H2Bub promotes ATX1,2, thereby H3K4me1, H3K4me1 ChIP-seq was performed in *hub1*, *hub2* and WT. Genes hypo-H3K4 monomethylated in the *hub* significantly overlap with genes that lose ATX2 in the *in planta hub2* (2 replicates) and genes that lose ATX1 or ATX2 in the *in silico hub*. **(e)** To test if H3K4me1 promotes H2Bub or H4K16ac, ChIP-seq targeting H2Bub and H4K16ac was performed in *atx1/2/r7* and WT. Genes that are hypo-H2B ubiquitinated in *atx1/2/r7* significantly overlap with hypo-H3K4me1 genes (= ATX1/2/R7-marked genes), while genes detected to be hypo-H4K16 acetylated did not overlap with ATX1/2/R7-marked genes nor with ATX1,2 localization, suggesting that H2Bub but not H4K16ac is regulated downstream of H3K4me1.



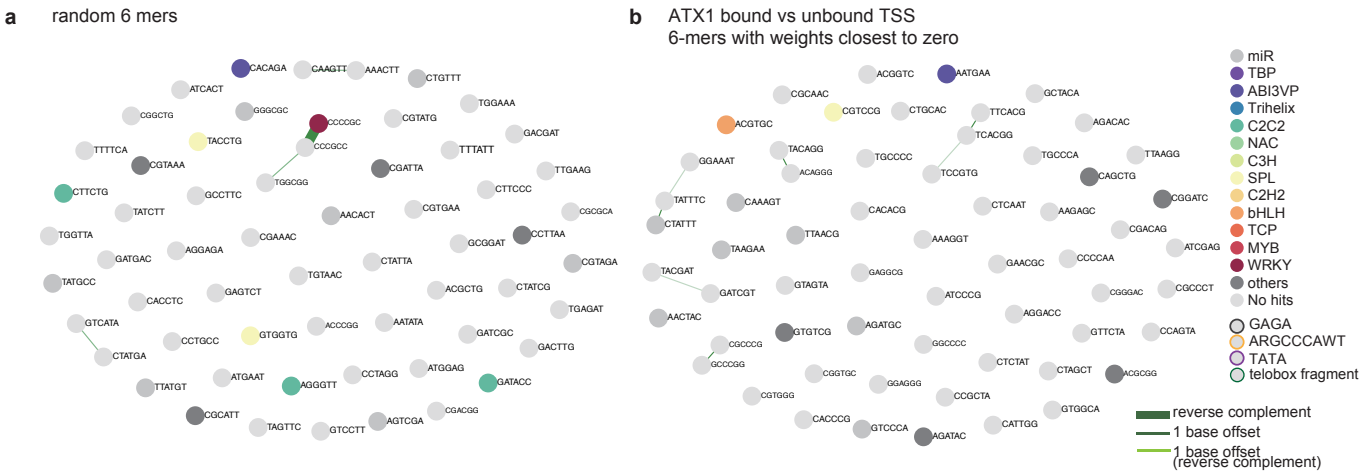
Supplementary Figure 8 ATXR7 and FLD share target genes.
(a) ATXR7 and FLD colocalize. When the Random Forest model is trained to predict ATXR7-bound and -unbound genes in the same manner as Fig. 3a except for the inclusion of the amount of FLD protein in the TTS region showed high importance. Error bars represent the standard deviation of the 5 repeats of training. **(b)** ATXR7-bound TTS regions are highly enriched in FLD. **(c)** There is a significant overlap between genes that lose H3K4me1 in the *atxr7* mutant (RPKM WT-*atxr7* > 6.5) and genes that gain H3K4me1 in the *fld* mutant. p-value is calculated with a hypergeometric test. The center point of violin plot, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range.



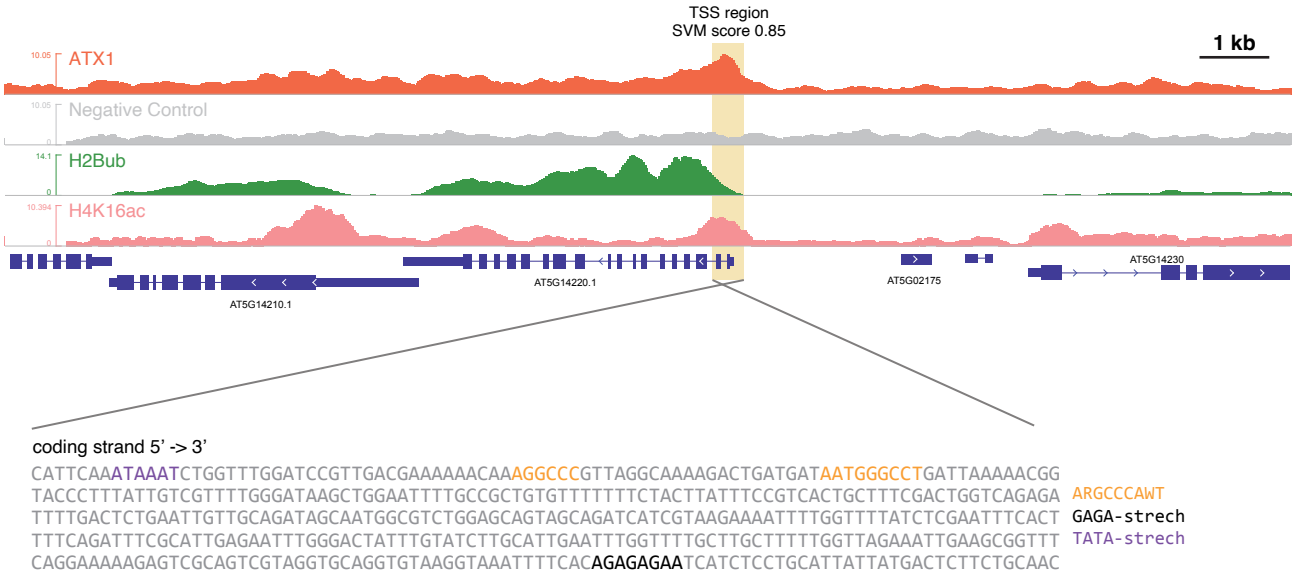
Supplementary Figure 9 Biological replicates confirmed that DNA motifs also rule ATX1 and ATX2 localization. The ChIP-seq data sets of the biological replicates (Supplementary Fig. 5) were analyzed in the same manner as Fig. 4. Biological replicates also achieved accurate prediction for both **(a)** ATX1 and **(b)** ATX2 as demonstrated by ROC plots. Error bars represent the standard deviation of the 5 repeats of training. **(c,d)** The learned trends of weights highly correlated between the replicates. r represents Pearson's correlation coefficient.



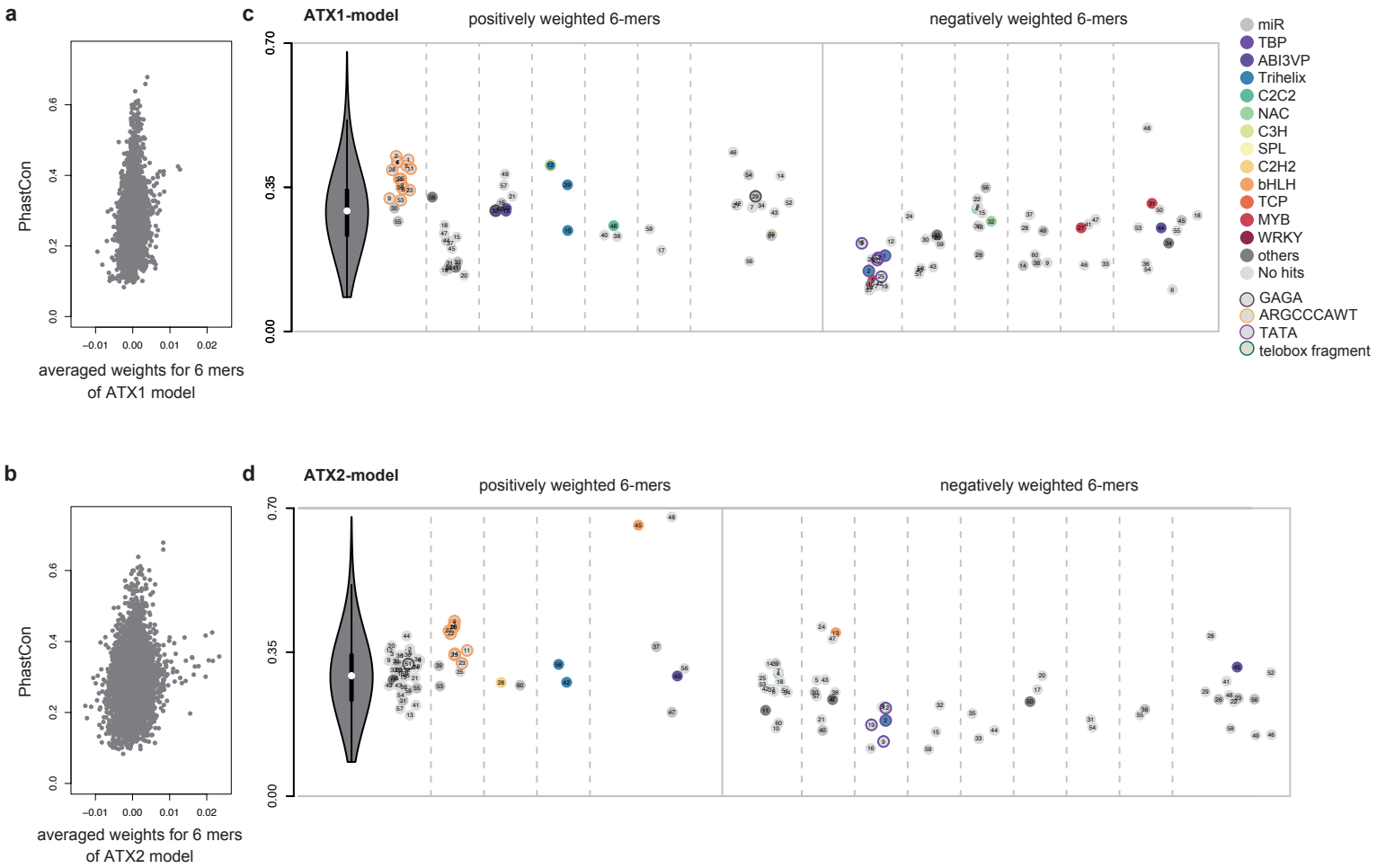
Supplementary Figure 10 Annotations of predictive and non-predictive 6-mers in the ATX2 model. Predictive and non-predictive 6-mers in the ATX2 models were selected and analyzed in the same manner as in ATX1 models (Fig. 4). **(a-c)** Clustering and annotation of predictive and non-predictive 6-mers. Each circle represents the top sixty 6-mers with negative (**a**), positive (**b**), or nearly zero (**c**) SVM weights in the ATX2 model. **(d)** Positional distributions of highly weighted motifs in the TSS region. **(e)** ATX2-bound TSS significantly overlaps with sppRNA-harboring TSS detected in the *hen2-2* background. The significance of the overlap was tested using a hypergeometric test.



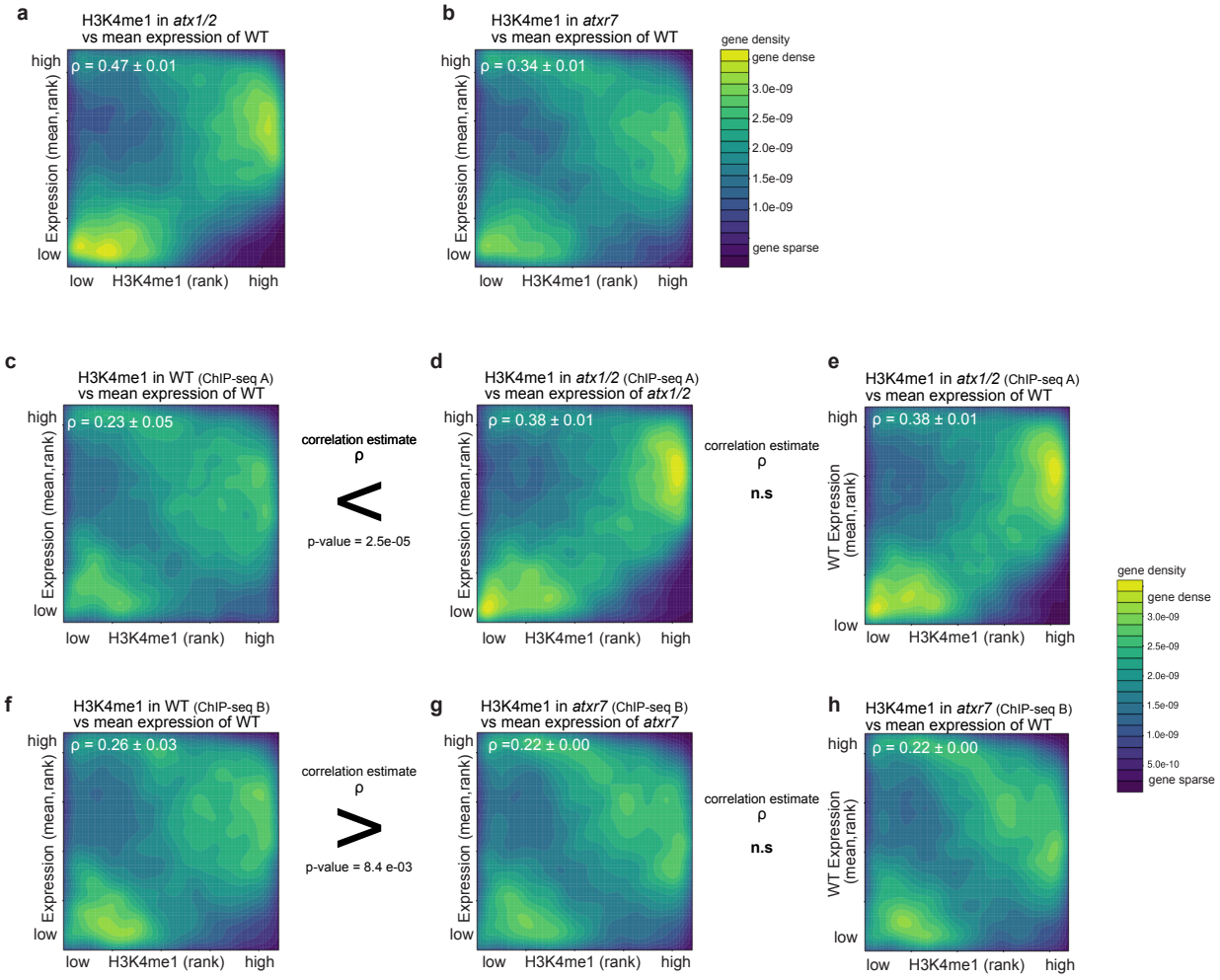
Supplementary Figure 11 Non-predictive 6-mers.
(a,b) Clustering and annotation of non-predictive sixty 6-mers. Random (**a**) and near-zero-weighted (**b**) sixty 6-mers in the ATX1 model.



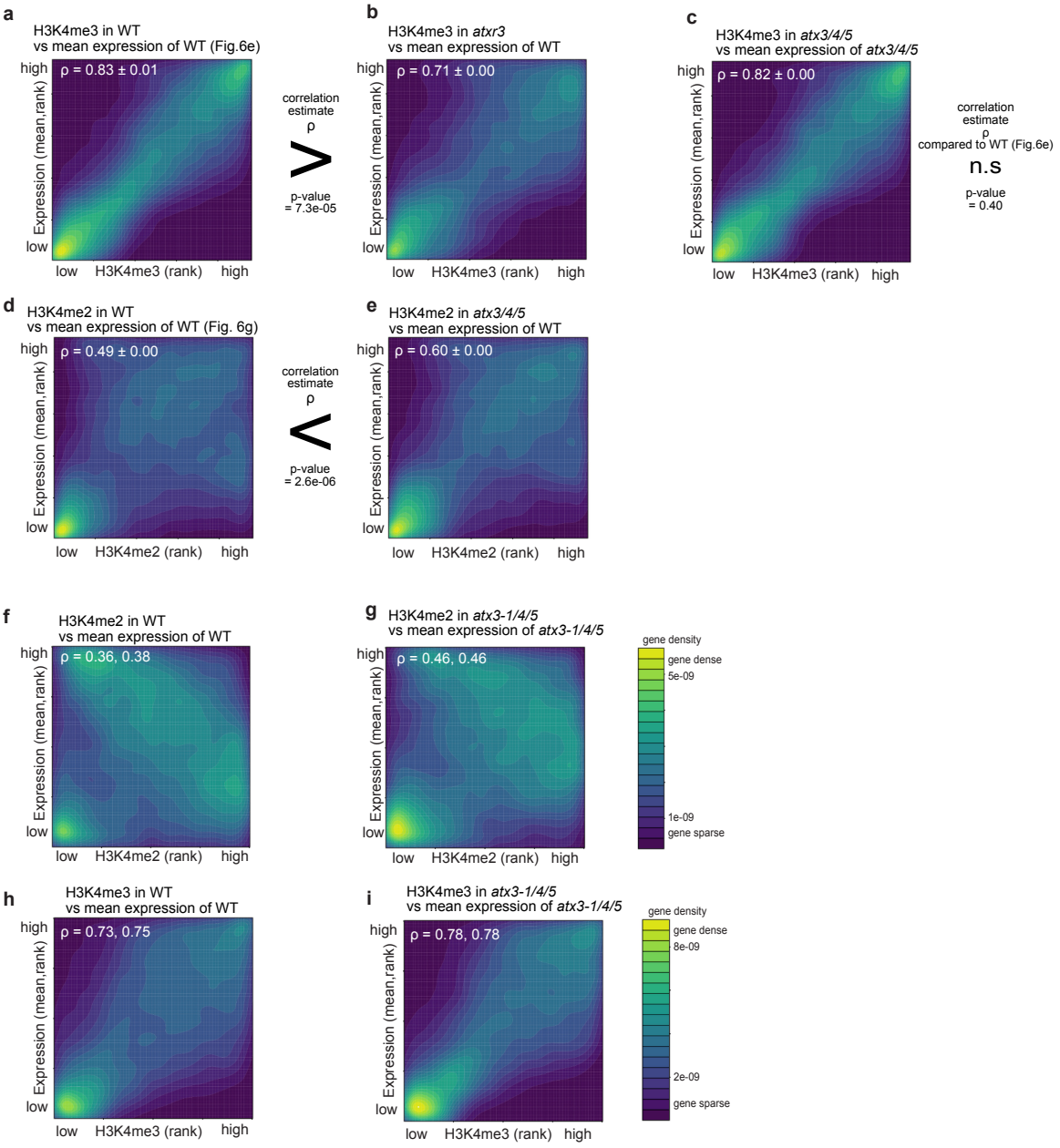
Supplementary Figure 12 An example browser view of ATX1-bound and unbound genes.
Examples of the ATX1-bound and unbound genes. ChIP-seq read coverage tracks of ATX1, negative control (non-transgenic), H2Bub, and H4K16c are shown. AT5G14220 is one of the 'ATX1-bound genes'. Its 'TSS region', highlighted with yellow shade, has higher levels of H2Bub and H4K16ac compared to the other TSS regions in the view. The SVM score of this TSS region in the full model is 0.85 (ranges from 0 to 1. The higher the value is, the more likely it to be ATX1-bound, based on DNA sequence). In the DNA sequence of the TSS region, positively weighted ARGCCCAWT and GAGA-stretch are present, as well as negatively weighted TATA-stretch.



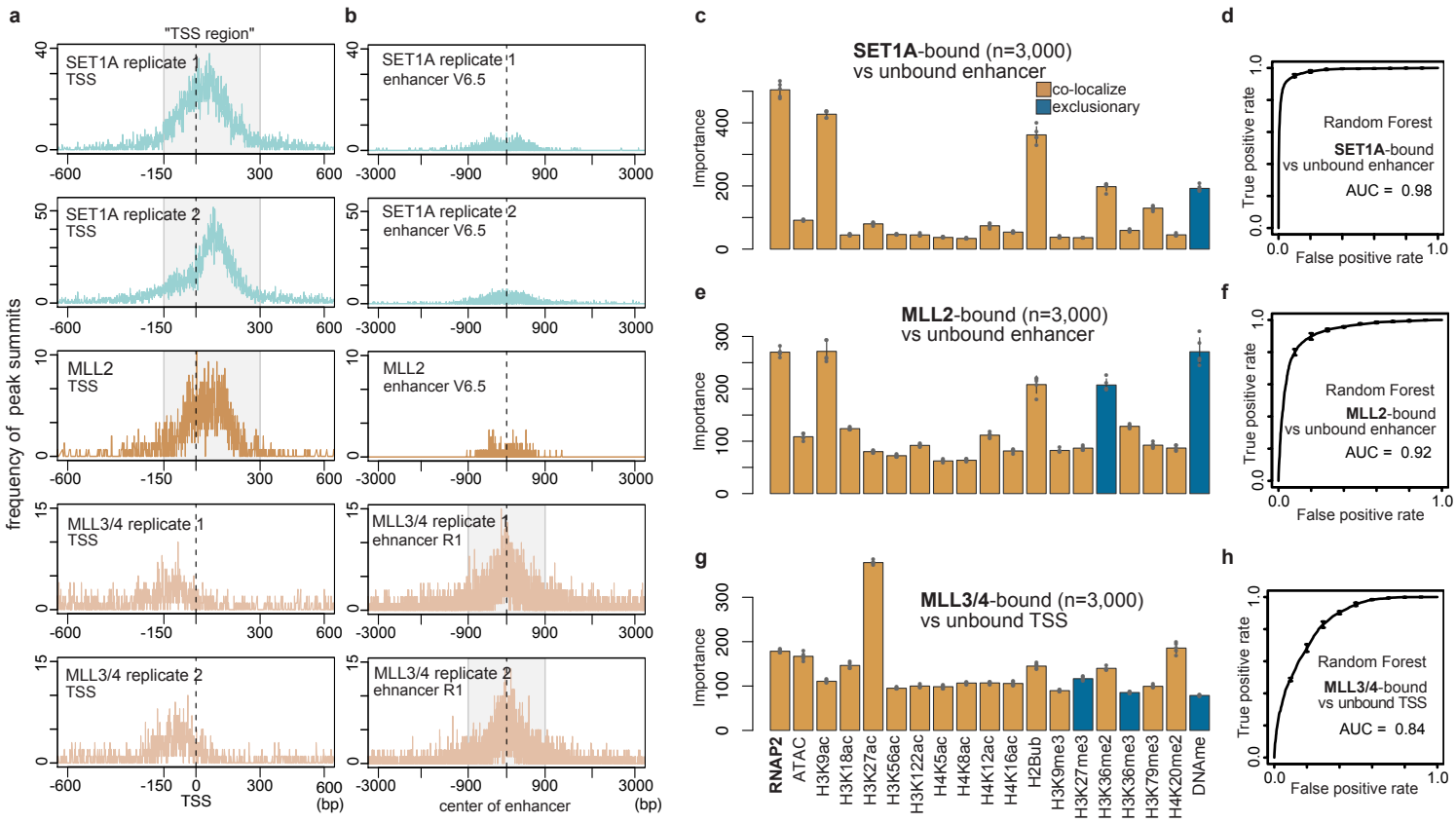
Supplementary Figure 13 Conservation of the predictive motifs in angiosperm genomes.
The conservation score PhastCon[99] score calculated by ref. 101 based on genome sequence alignment of 63 angiosperm species. Higher PhastCons score reflects higher conservation within the TSS region. **(a,b)** SVM weights and PhastCons lack overall correlation **(c,d)** RGCCCAW motif is under evolutionary constraint. Leftmost violin plots represent PhastCon distribution of all 6-mers (n= 4^6 = 4096) in the TSS region. The center point of violin plot, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. Each circle represents highly weighted 6-mers with the same color and numbers as Fig. 4 and Supplementary Fig. 10. 6-mers connected with lines in Fig. 4 and Supplementary Fig. 10 are clustered together.



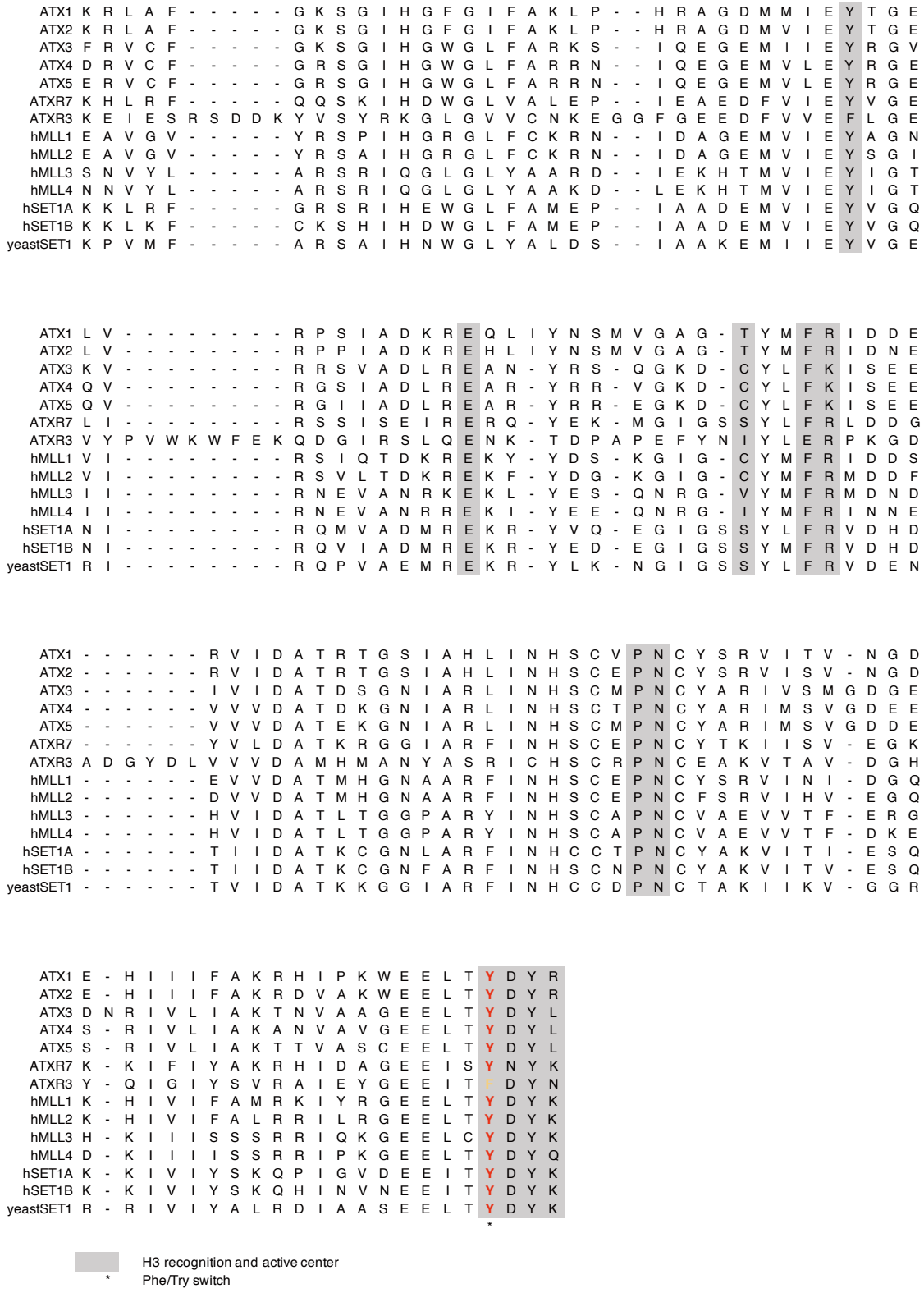
Supplementary Figure 14 Change in H3K4me1-transcription correlation is due to change in H3K4me1, not in transcription. **(a,b)** Correlation between H3K4me1 and expression is shown in the same way as Fig. 5, except that in this figure the expression (mRNA-seq) data are of WT. The *atx1/2* mutant has a higher, while *atxr7* has a lower correlation than WT, demonstrating that the change in H3K4me1 rather than transcription causes the change in the correlation. ρ= Spearman's correlation coefficient. The color scale is the same as Fig. 5. **(c-h)** Another ChIP-seq datasets replicated the trend. ChIP-seq data used for **(c-e)** *atx1/2* -WT comparison is the data sets for double mutants shown in Supplementary Fig. 1c (ChIP-seq A), and for **(f-h)** *atxr7*-WT comparison is the datasets shown in Supplementary Fig. 1a (ChIP-seq B). P-values indicate ρ are significantly different between genotypes (Welch's two sample t-test).



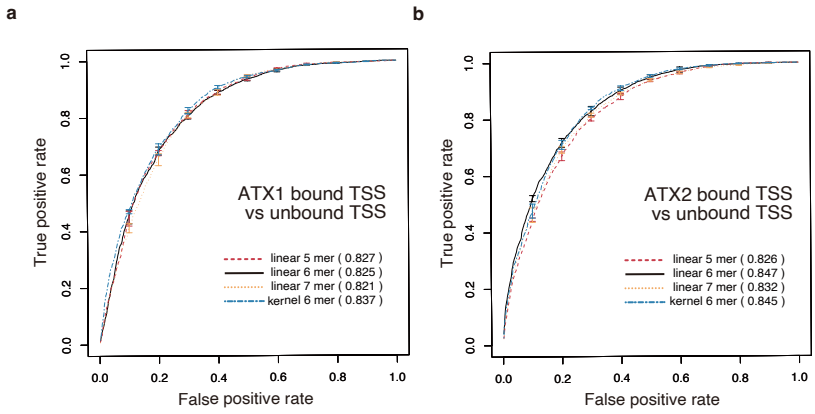
Supplementary Figure 15 Correlational trend between H3K4me and transcription of *atxr3* and *atx3/4/5.*
(a) Identical to Fig. 6e and shown for reference. **(b)** Correlation between H3K4me3 level in *atxr3* and expression level in WT. Compared to WT-WT comparison in **(a)**, the correlation is diminished and almost as weak as a mutant-mutant comparison in Fig. 6f, demonstrating that the change in H3K4me3 rather than transcription causes the change in the correlation. **(c)** *atx3/4/5* mutant does not show a large correlational change between H3K4me3 and transcription. **(a-c)** and Fig. 6e,f are shown in the same color scale. **(d)** Identical to Fig. 6g. **(e)** Correlation between H3K4me2 level in *atx3/4/5* and expression level in WT. Compared to a WT-WT comparison in **(d)**, the correlation is stronger and almost as strong as a mutant-mutant comparison in Fig. 6h, demonstrating that the change in H3K4me2 rather than transcription causes the change in the correlation. **(f-i)** Change in the correlation between H3K4me2 **(f, g)** or H3K4me3 **(h, i)** and transcription in the *atx3-1/4/5* mutant, from the datasets reported by Chen et al., 2017 (ref. 20). In the *atx3-1/4/5* mutant, which has stronger effects on H3K4me2 and H3K4me3 compared with *atx3/4/5* (Fig. 1e, Supplementary Fig. 2), there is a stronger correlation compared to WT not only in transcription-H3K4me2 but also in transcription-H3K4me3. P values indicate ρ are significantly different between genotypes (Welch's two sample t-test).



Supplementary Figure 16 Localization of mammalian H3K4 methyltransferases in TSSs and enhancers.
(a,b) Position of mammalian H3K4 methyltransferases' peaks relative to TSS (**a**) and enhancer (**b**), visualized as a frequency of ChIP-seq peak summits. x-axis, distance from TSS (**a**) or center of enhancer (**b**); y-axis, number of peak summits. Most of SET1A and MLL2's peaks belong to the region spanning from 150 bp upstream to 300 bp downstream of TSS, where we hereafter refer to as the 'TSS region'. Most of the MLL3/4 peaks belong to 1800 bp around the center of the enhancer, where we hereafter refer to as 'enhancer'. **(c,e,g)** 'Importance' of features derived from the Random Forest models trained to predict enhancer occupied by SET1A (**c**) and MLL2 (**e**), or to predict TSS region occupied by MLL3/4 (**g**). Graphs are shown like Fig. 7b,e,h. **(d,f,h)** ROC plot showing the prediction accuracy of the random forest models corresponding to (**c**,**e**,**g**). All ROC and AUC are calculated with test data (25% of the original data). Average and standard deviation of the 5 repeats of training are plotted.



Supplementary Figure 17 Amino acid alignment of catalytic domain SET in ATXs and their homologs.
Y at 'Y/F-switch' is typical for SET domains of mono/di-KMTs [50]. Annotations are from ref. 105.



Supplementary Figure 18 Parameter tuning of SVM models.
(a,b) Prediction accuracy of linear SVM models trained with different sizes of k-mers, and kernel SVM with 6-mers. Error bars represent the standard deviation of the 5 repeats of training. The numbers in parenthesis are AUC calculated with test data

References

102. Liu, Y., Koornneef, M. & Soppe, W. J. J. The absence of histone H2B monoubiquitination in the Arabidopsis hub1 (rdo4) mutant reveals a role for chromatin remodeling in seed dormancy. *Plant Cell* **19**, 433–444 (2007).

103. Fleury, D. et al. The Arabidopsis thaliana homolog of yeast BRE1 has a function in cell cycle regulation during early leaf and root growth. *Plant Cell* **19**, 417–432 (2007).

104. Bourbousse, C. et al. Histone H2B monoubiquitination facilitates the rapid modulation of gene expression during Arabidopsis photomorphogenesis. *PLoS Genet.* **8**, e1002825 (2012).

105. Zhang, Y. et al. Evolving catalytic properties of the MLL Family SET Domain. *Struct./Fold. Des*. **23**, 1921–1933 (2015)