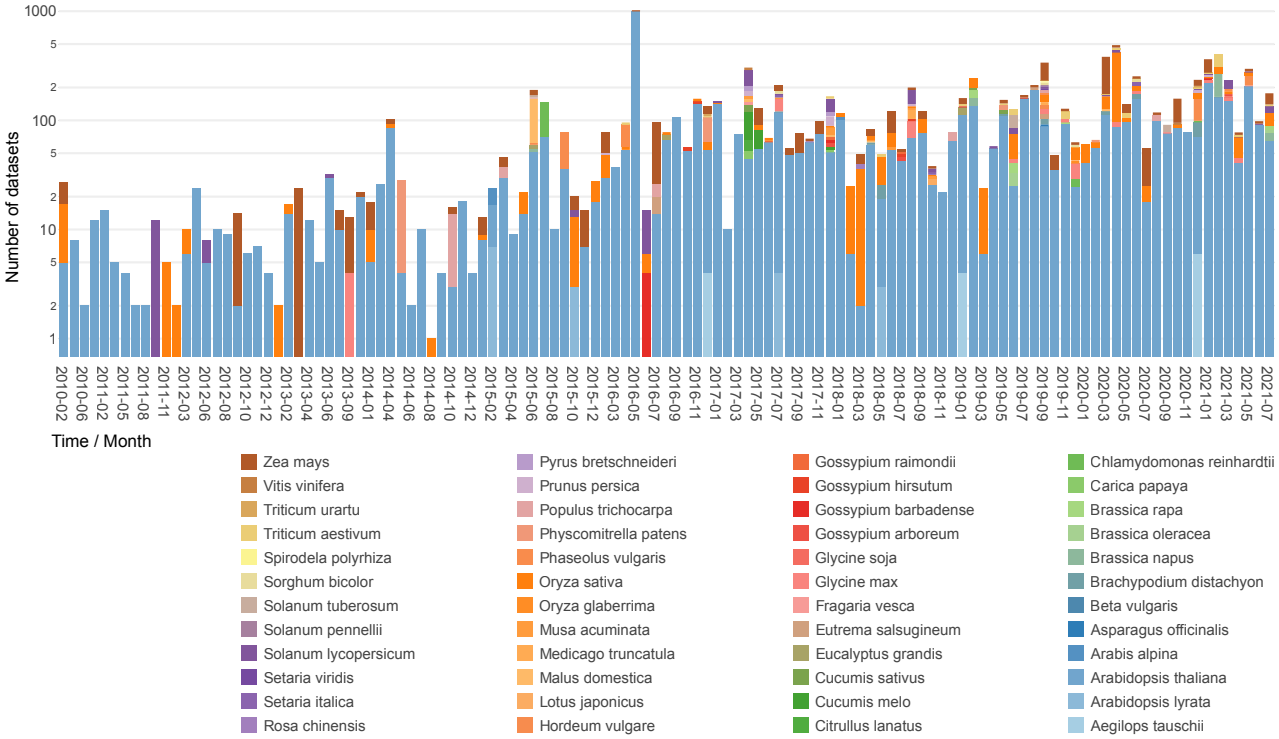
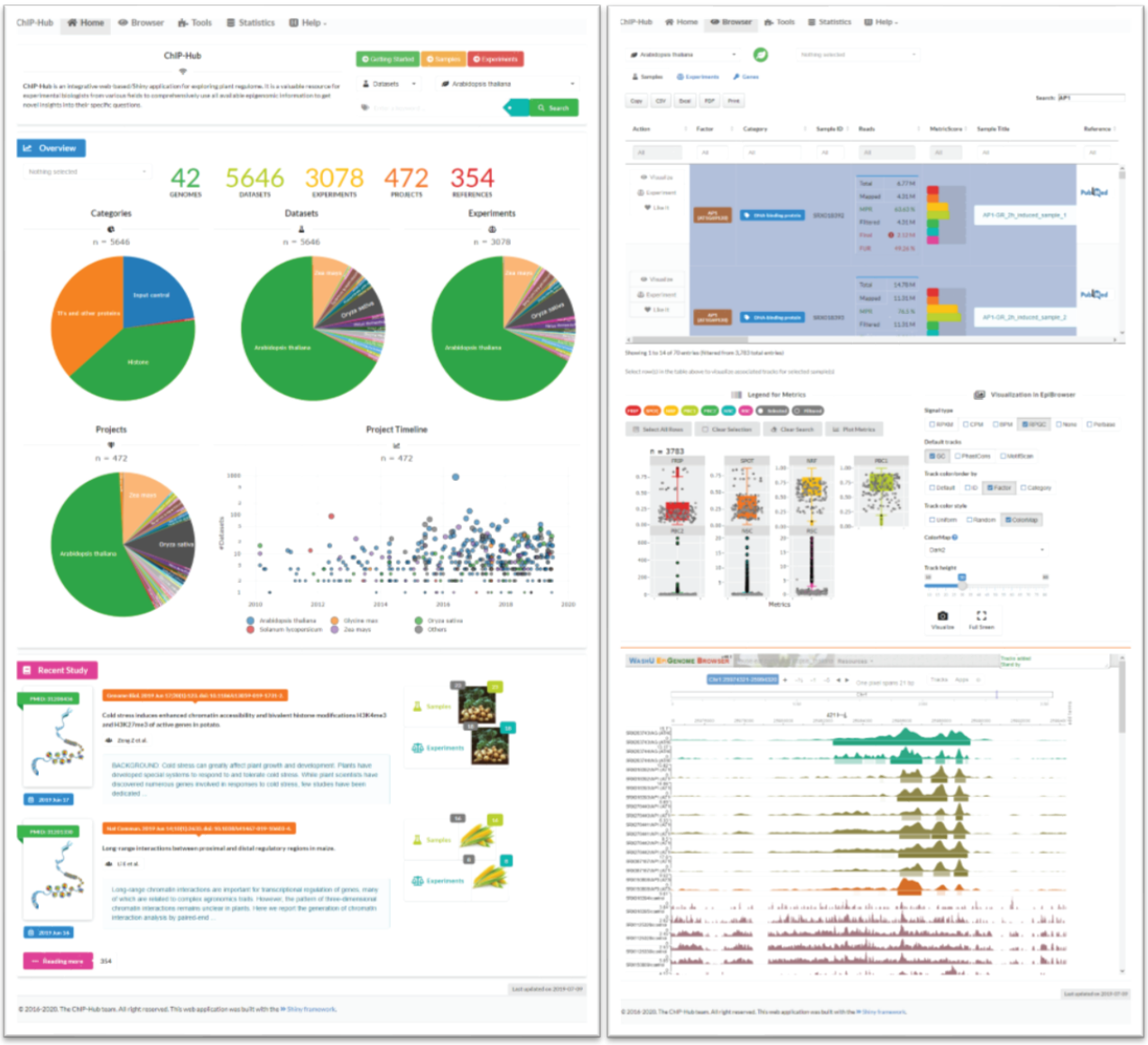
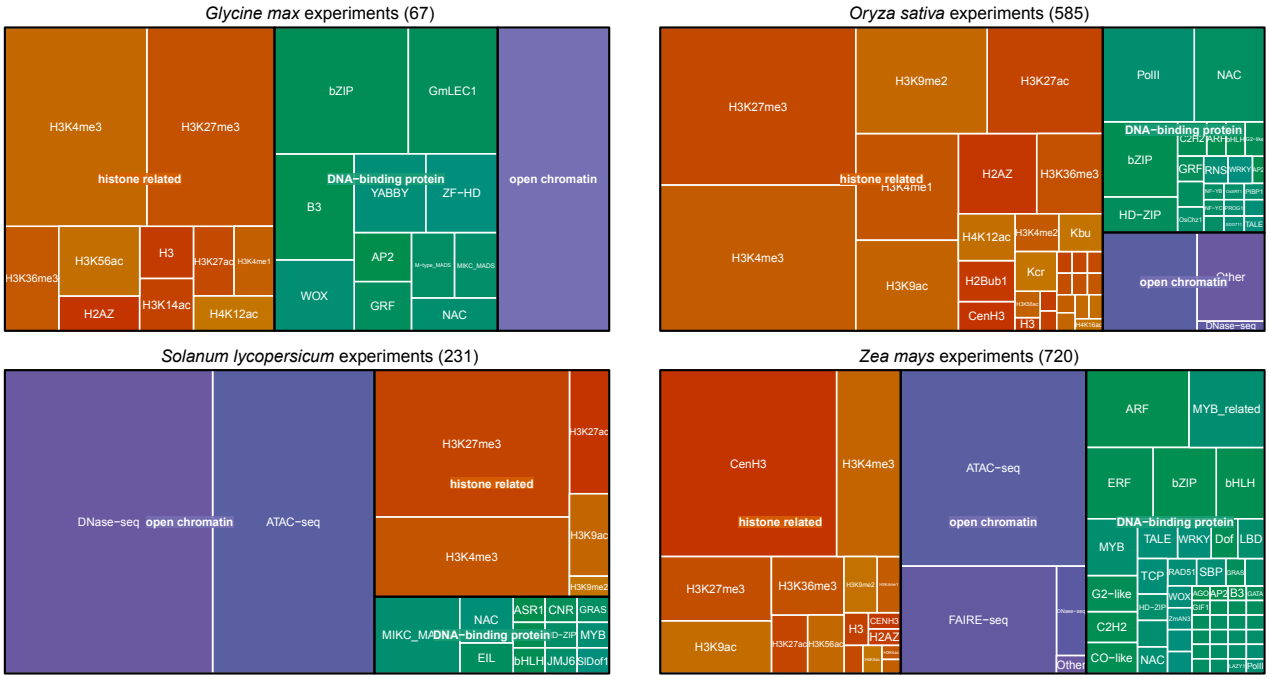
Supplementary Figures



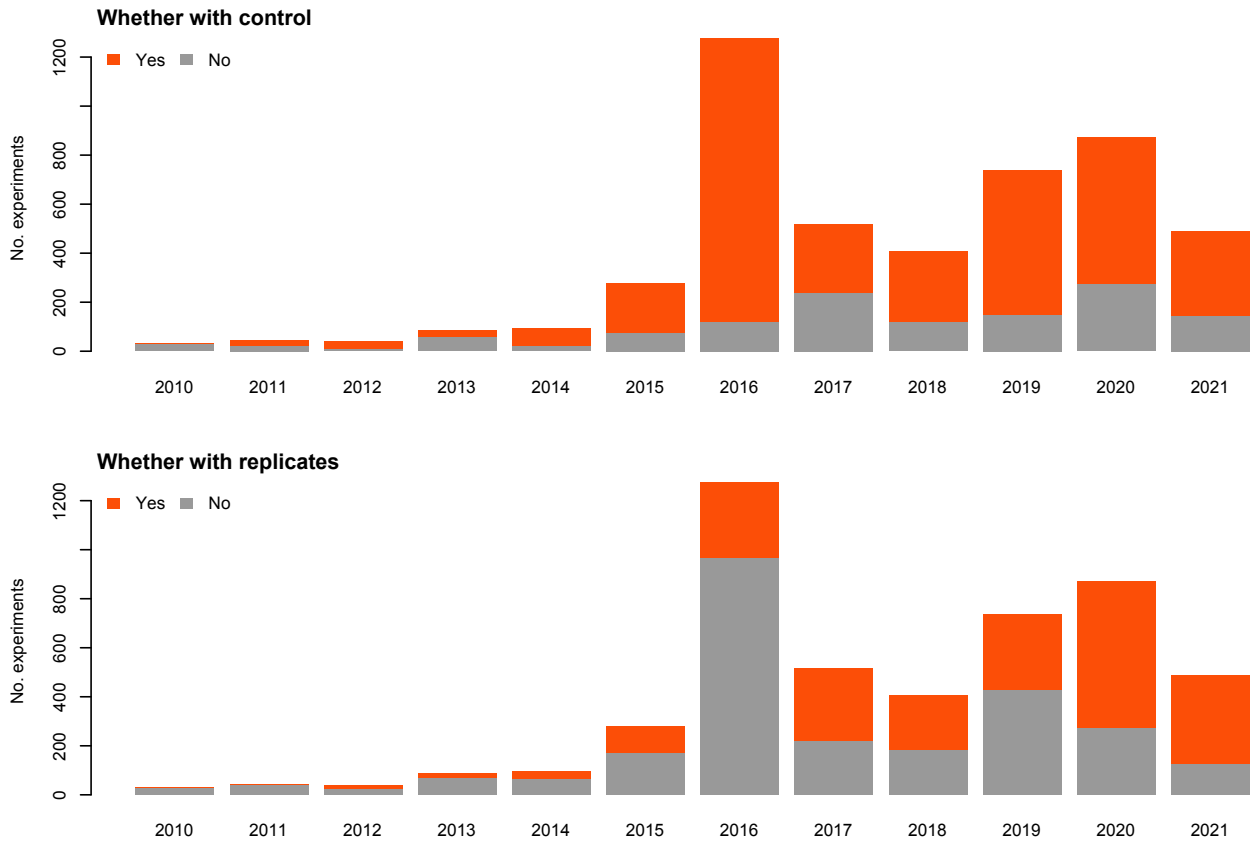
Supplementary Fig. 1: Statistics of publicly available regulome datasets in plants. Barchart showing the release of plant datasets in the past years. Datasets are colored by different plant species.



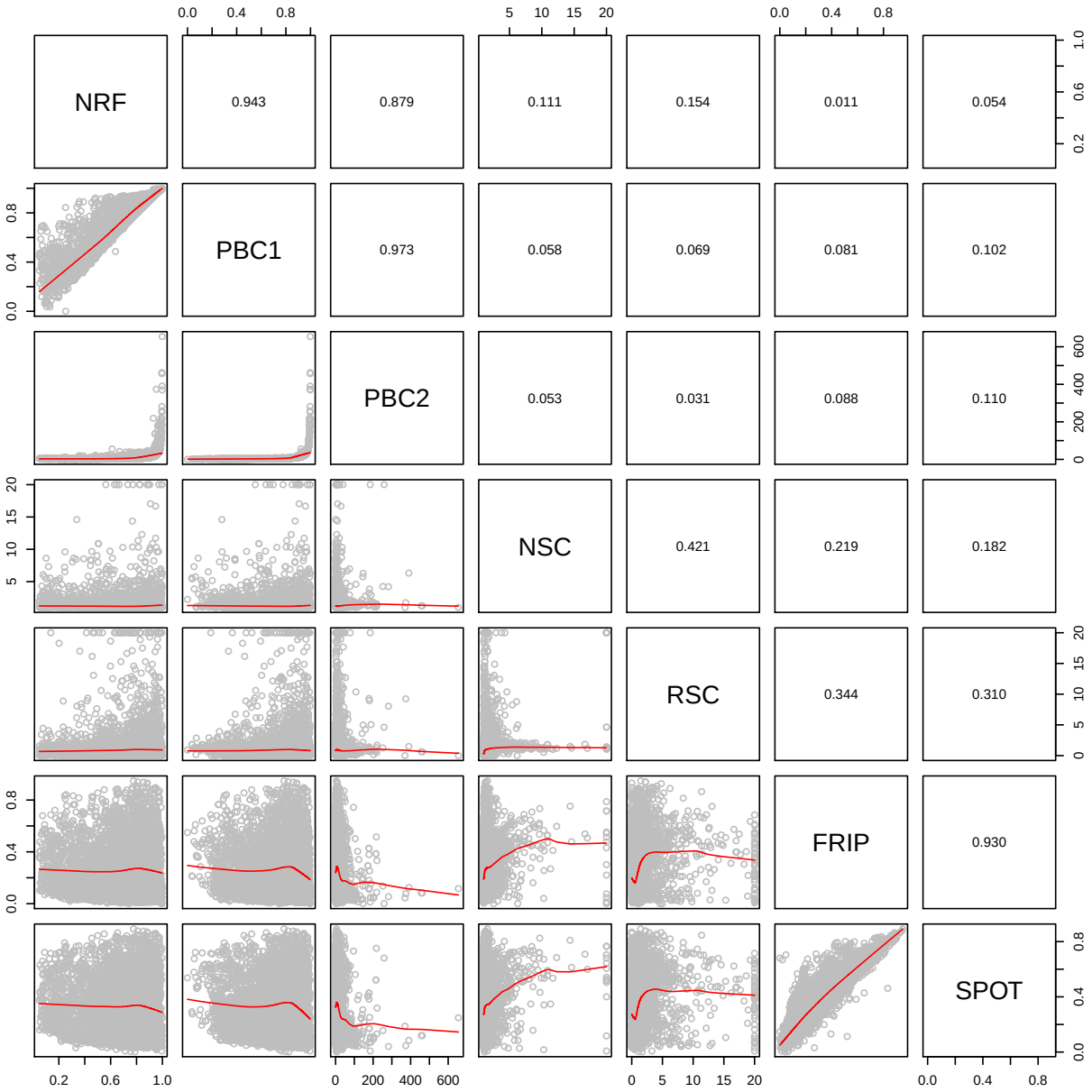
Supplementary Fig. 2: The ChIP-Hub Shiny application. Screenshots of the ChIP-Hub website. The left panel shows the "Home" page of ChIP-Hub. The right panel shows the "Browser" page with example tracks of ChIP-seq datasets shown at the bottom genome browser. Please visit our website (http://www.chip-hub.org) for interactive pages.



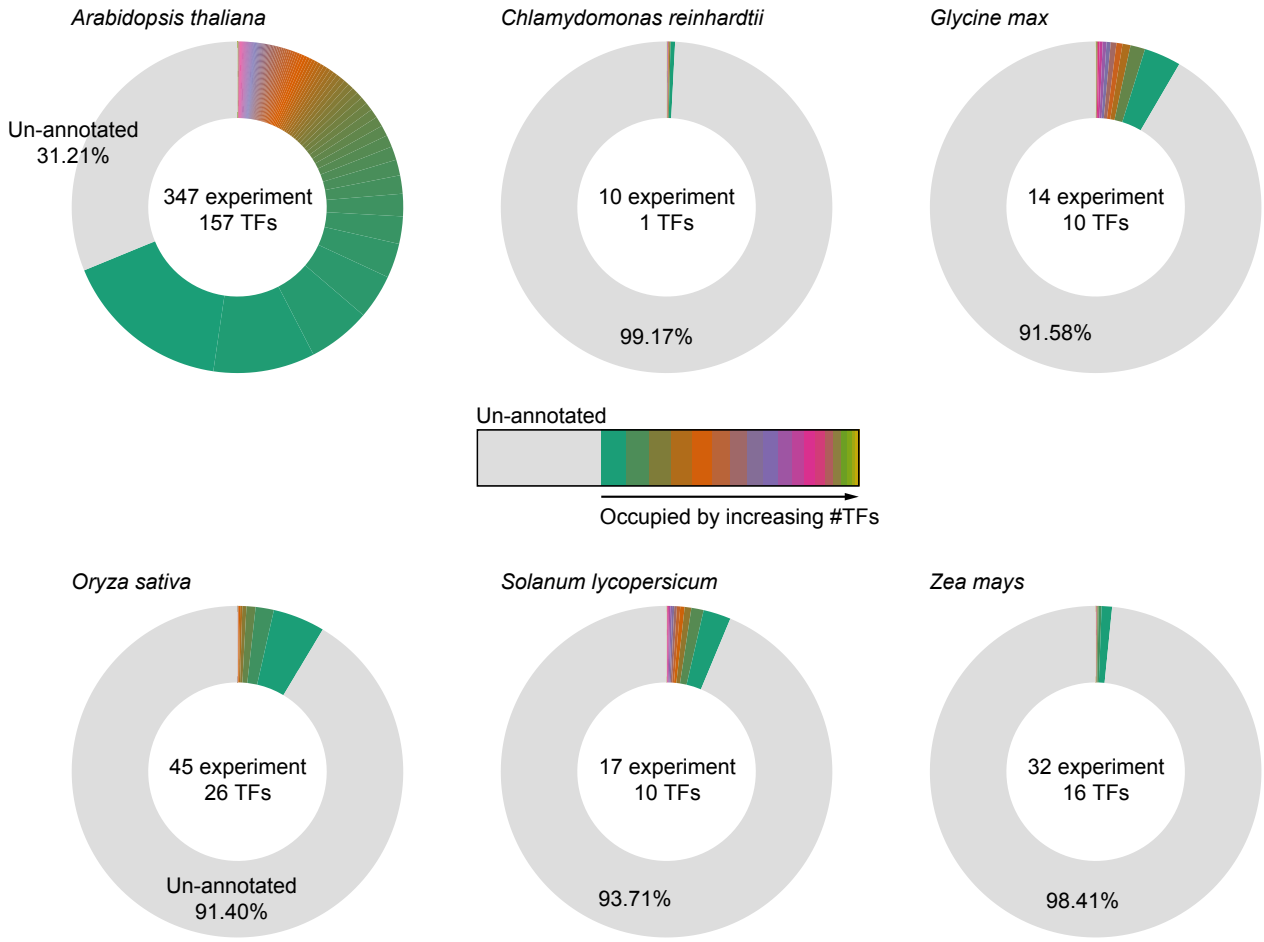
Supplementary Fig. 3: Examples of metadata files. Treemap showing the classification of experiments in Glycine max, (*Oryza sativa*), (*Solanum lycopersicum*) and (*Zea mays*), according to transcription factor (TF) families, the types of histone modifications or open chromatin experiments.



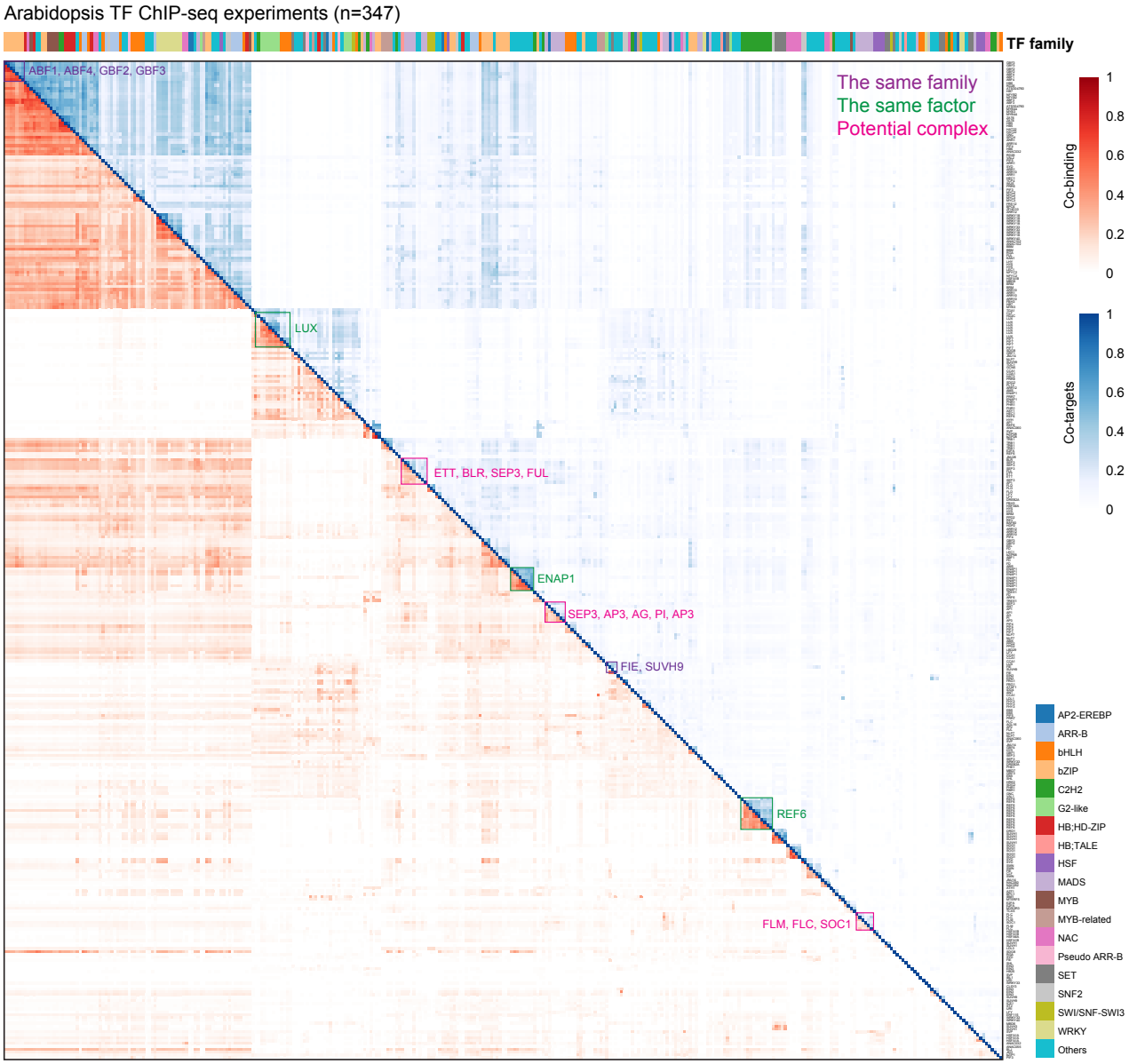
Supplementary Fig. 4: Summary of the quality of experiments according to the time. Barcharts showing whether the experiments with proper input control or with replicates.



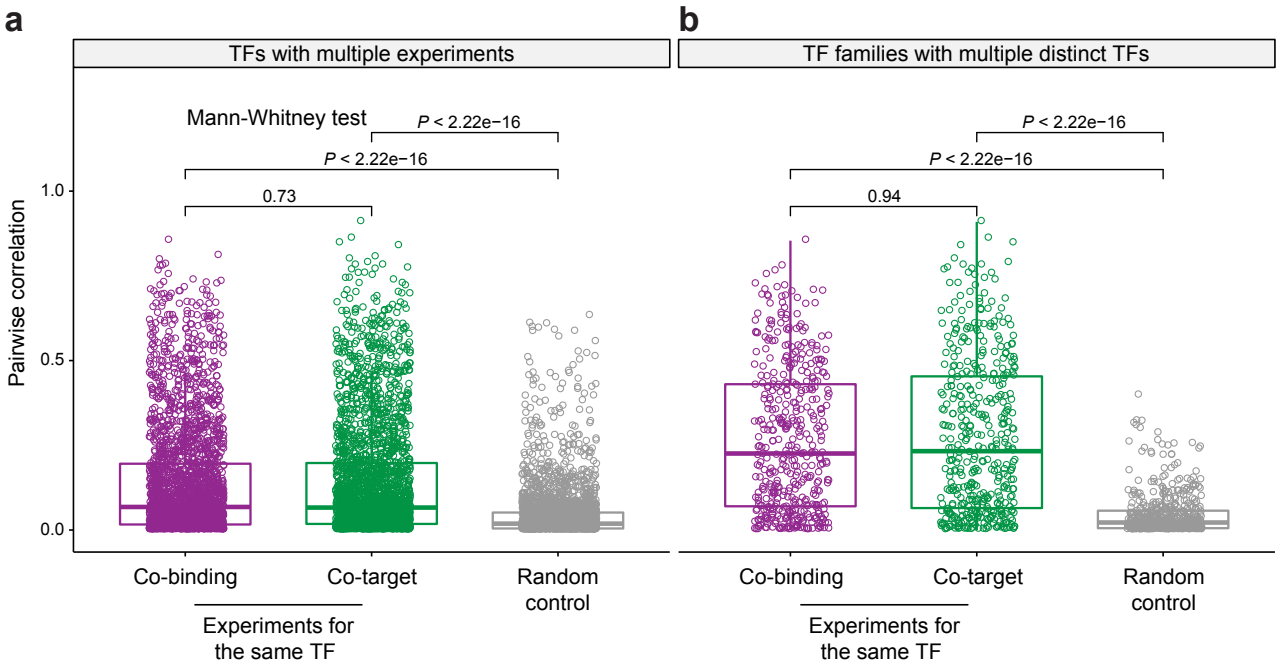
Supplementary Fig. 5: Correlation of metrics scores. Scatter plots (the lower panel) showing the Pearson's correlation coefficients (the upper panel) of different metrics. SPOT: signal portion of tags; FRiP: fraction of reads in peaks; NSC: normalized strand cross-correlation coefficient. RSC: relative Strand cross-correlation coefficient; NRF: non-redundant fraction; PBC1/2: PCR bottlenecking coefficients 1/2.



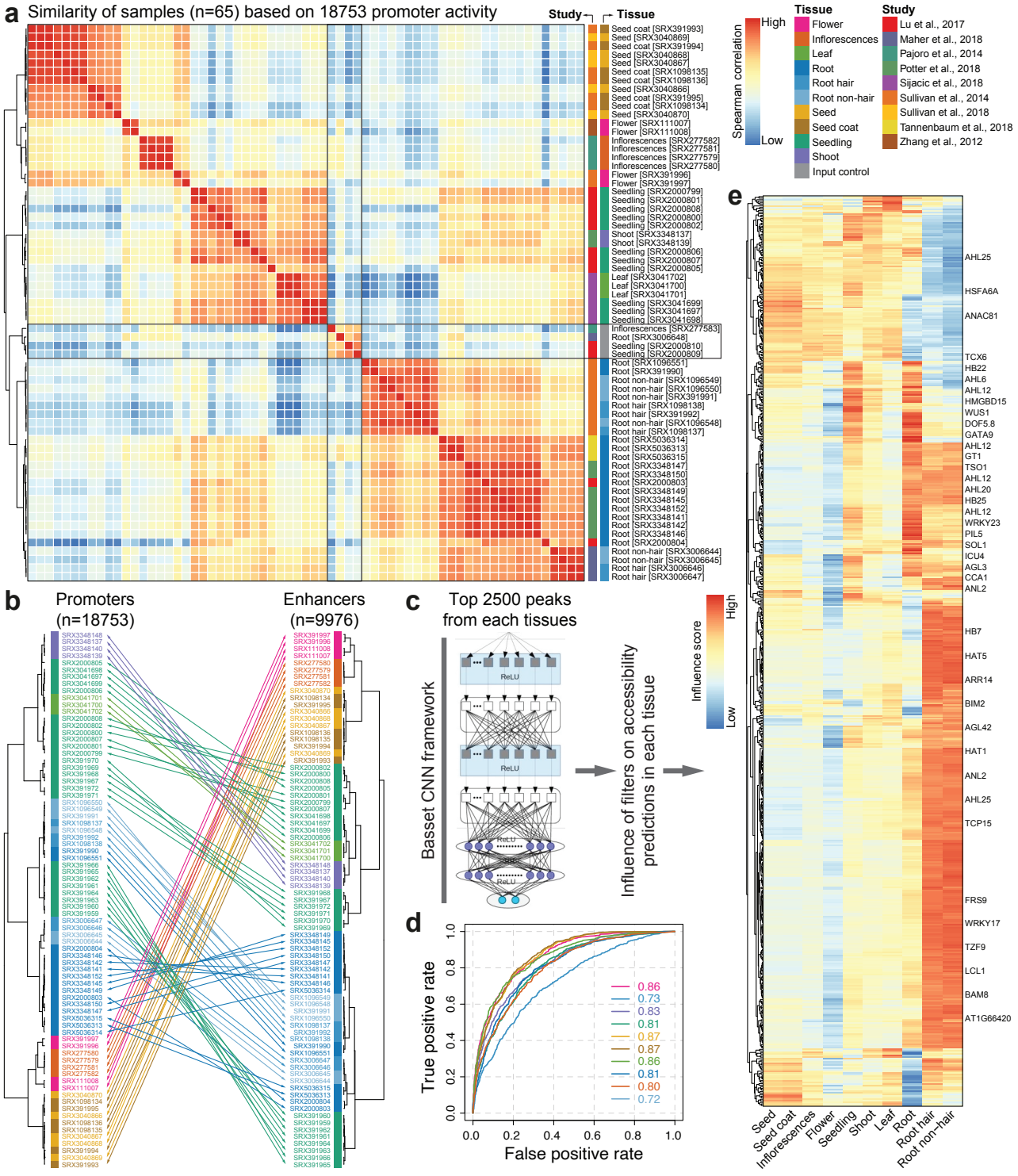
Supplementary Fig. 6: Fraction of plant genomes occupied by transcription factor (TF) binding sites (TFBSs). Donut charts showing the fraction of the entire genome covered by TFBSs based on ChIP-seq data in the selected plant species. The number of ChIP-seq experiments and the associated distinct TFs are indicated. The genome is colored according to the number of occupied TFs.



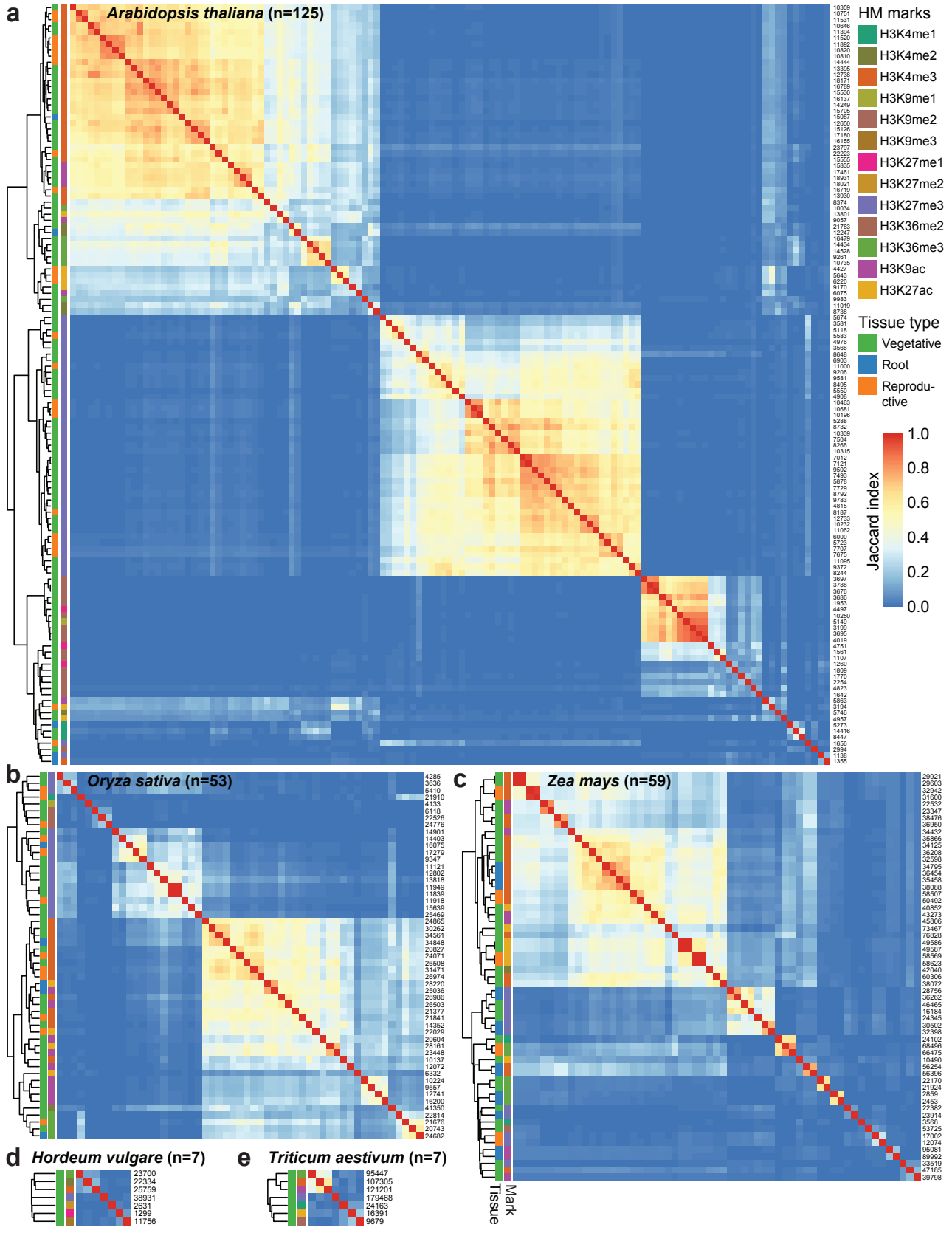
Supplementary Fig. 7: Co-associations of TFs in *Arabidopsis thaliana*. Heatmap showing co-binding relationships (upper triangle) and co-regulated targets (lower) by TFs. Each row or column represent one TF. Co-associations were both calculated based on Jaccard statistics with either peak basepair or target genes. TFs from different families were colored in the left bar. In general, the same TF in different experiments and TFs from the same family or from known protein complexes showed significantly higher associations than random controls (**Supplementary Fig. 8**). For example, MADS-domain TFs that act in a combinatorial manner in specifying floral organ identities (such as SEP3, AP3, AP1, AG and PI)[1] or developmental phase transitions (such as FLC, FLM and SOC1)[2–4] strongly overlap in their DNA-binding sites.



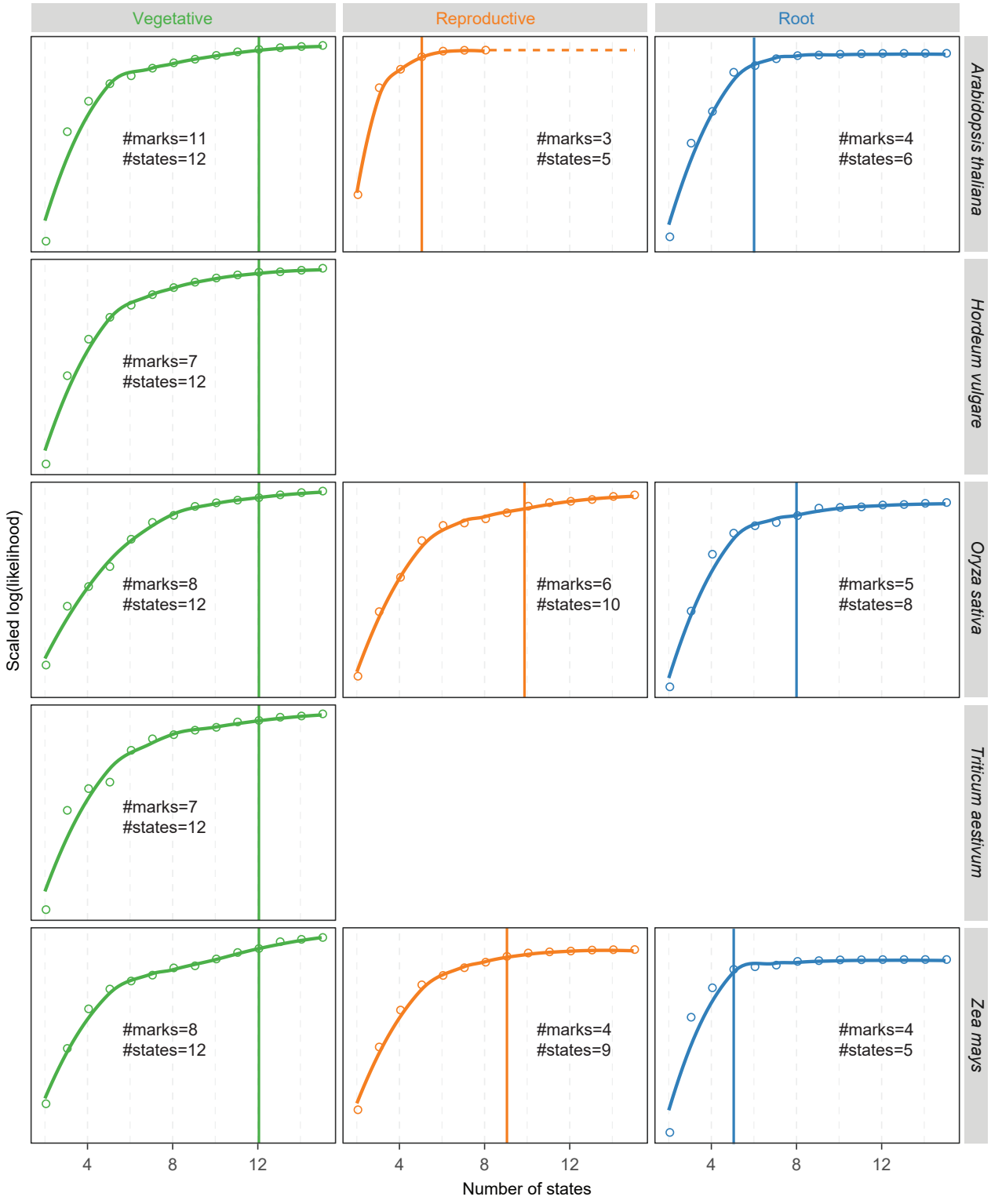
Supplementary Fig. 8: Co-associations of TFs, related to Supplementary Fig. 7. Boxplots (individual data points as overlays) showing that the same TF with multiple experiments (**a**) or TFs from the same family (**b**) have significantly higher associations than random controls. As random control, the same number of associations were sampled from the rest of comparisons.



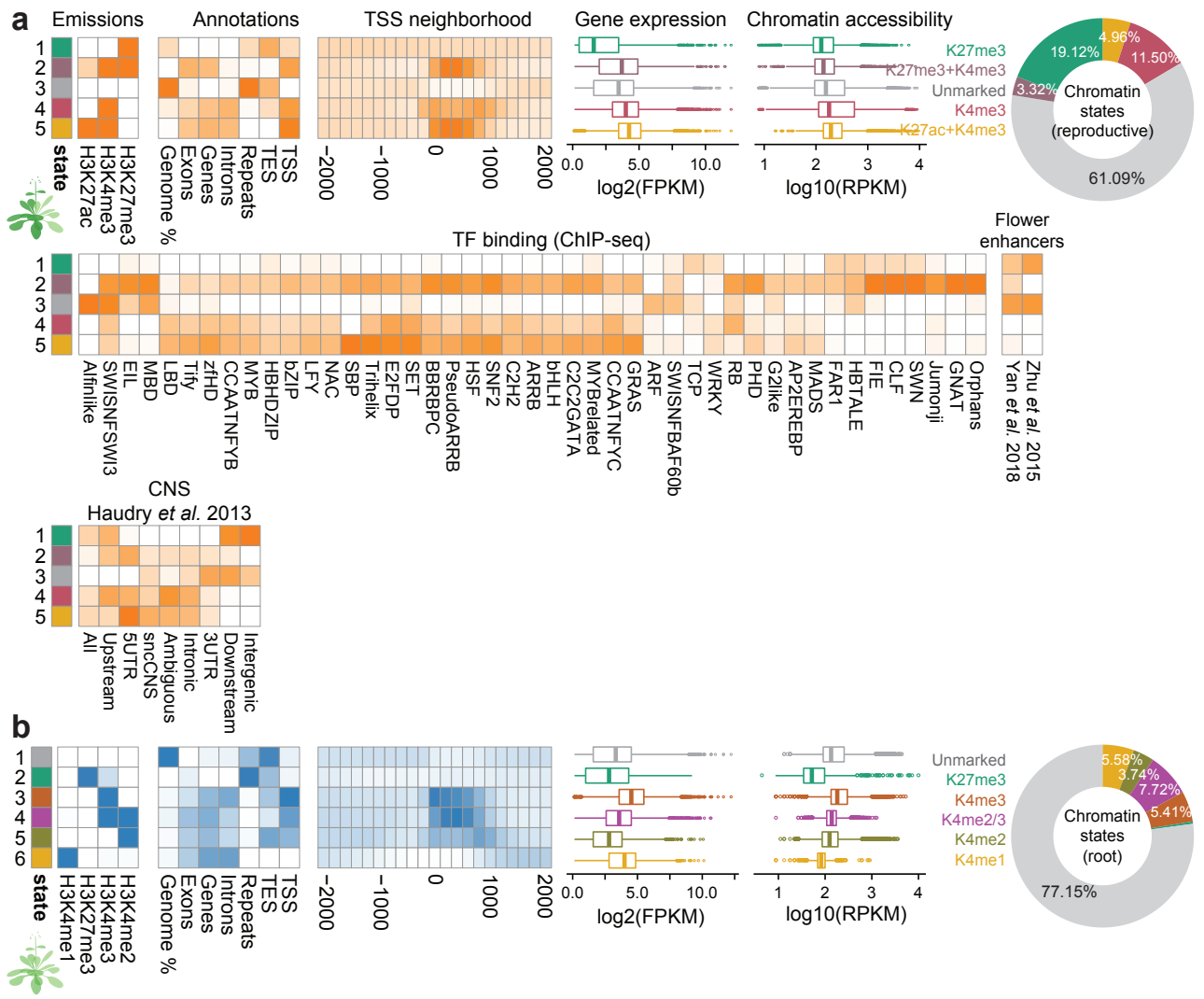
Supplementary Fig. 9: Dynamics of tissue-specific chromatin accessbility. (**a**) Sample similarity based on promoter activity. Open chromatin samples (with IDs labeled in square brackets) were collected from nine different studies. The input DNA samples (in grey; n=4) are used for control. (**b**) Comparing the clustering analysis based on enhancers (as in **Fig. 4a**) and promoters (as in **a**). (**c**) Predicting the sequence grammar underlying the chromatin dynamics of tissue-specific regulatory elements using the Basset[5] convolutional neural network (CNN) framework. (**d**) The ROC curves display the Basset false-positive rate versus true-positive rate for different tissues. (**e**) Heatmap showing normalized influence of motif-annotated filters on classification of enhancers in different tissues. Filters matched to known motifs are labeled.



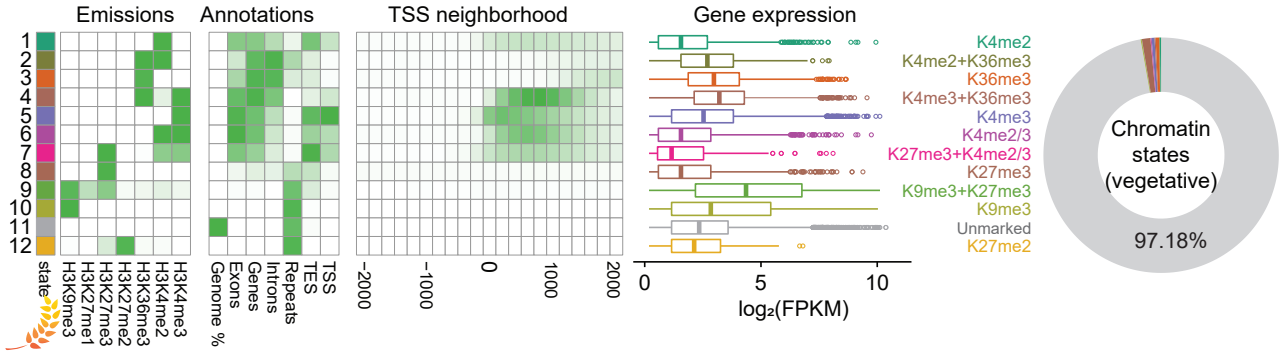
Supplementary Fig. 10: Manually curated histone modification ChIP-seq experiments used for ChromHMM. Heatmaps showing the similarity of histone modification ChIP-seq experiments in *Arabidopsis thaliana* (**a**), rice (*Oryza sativa*; **b**), maize (*Zea mays*; **c**), barley (*Hordeum vulgare*; **d**) and wheat (*Triticum aestivum*; **e**). The pairwise correlation of any two experiments were calculated based on Jaccard index. Each experiment was assigned to one of the reference tissues (vegetative-, reproductive- and root-related tissues). Experiments are colored according to the type of tissues or marks (as indicated in the bar on the left of heatmaps). The number of peaks in each experiments is show on the right of heatmaps. The full list of experiments can be found in Supplementary **Table S7**.



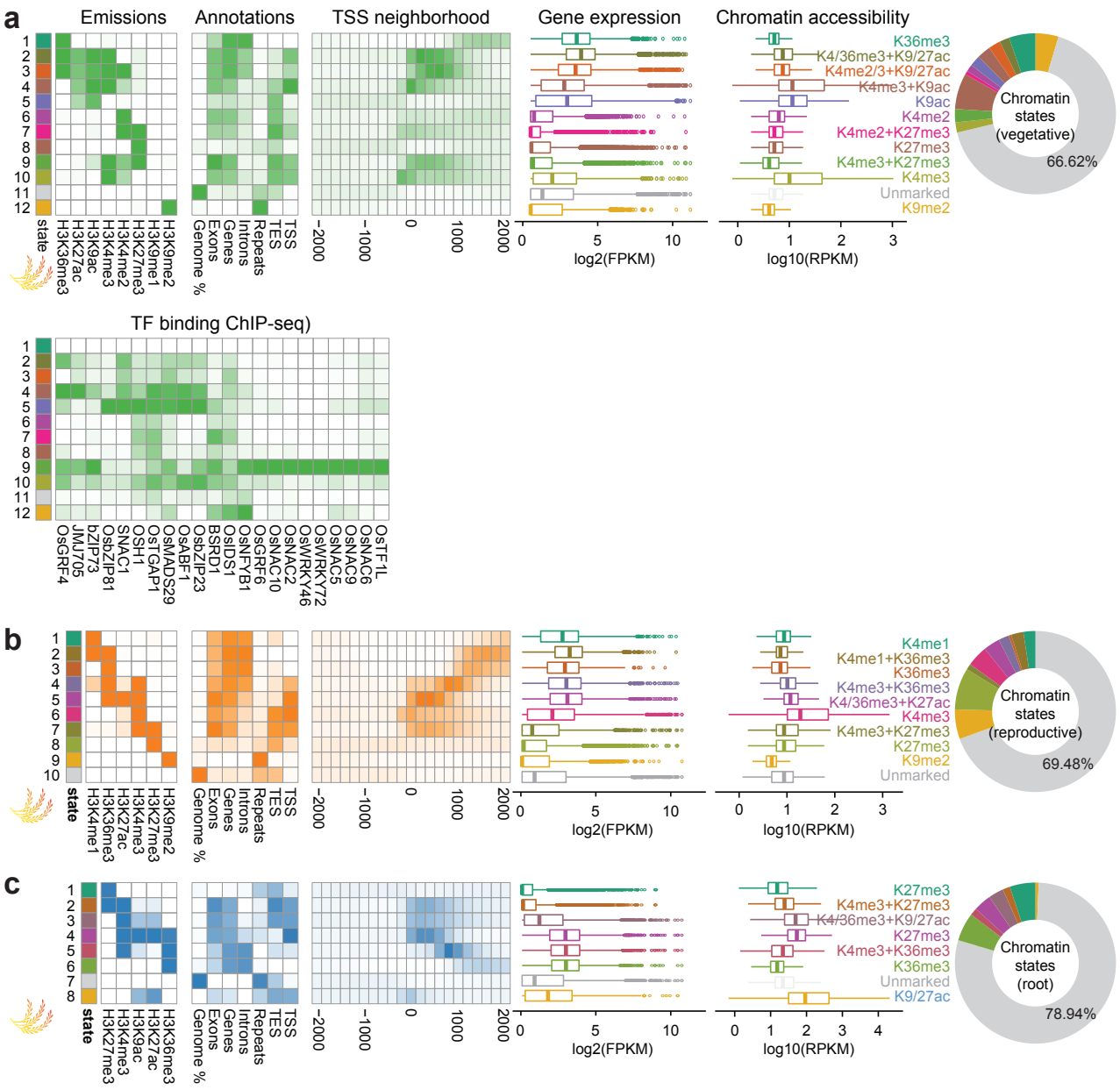
Supplementary Fig. 11: Chromatin states determined by ChromHMM. The ChromHMM[6] model was learnt with up to 15 states for each reference tissue type (column) in each genome (row). The log(likelihood) of the model output by the program increased as number of states increased, while the extent of increment declined after a specific number of states (dependent on tissue types and genomes) and the model was considered as "optimal" when reaching this number of states. The number of histone modification marks and the optimal numer of chromatin states were indicated.



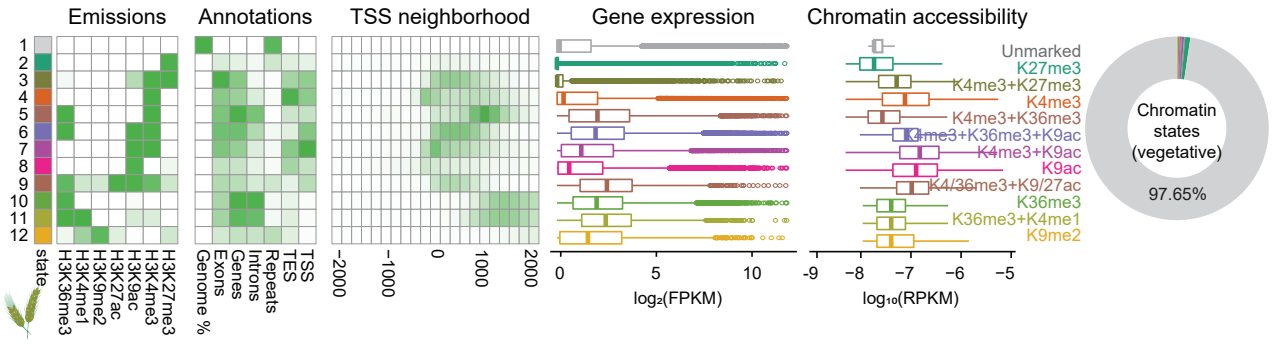
Supplementary Fig. 12: Chromatin states in Arabidopsis reproductive-related and root-related tissues. Prediction of chromatin states by ChromHMM[6] in Arabidopsis reproductive-related (**a**) and root-related tissues (**b**). In each figure, the left panel displays a heatmap of the emission parameters in which each row corresponds to a different state and each column corresponds to a histone modification mark; the second heatmap displays the overlap fold enrichment for different genomic annotations; the third heatmap shows the fold enrichment for each state for each 200-bp bin position within 2 kb around a set of transcription start sites (TSSs); Expression patterns for neighboring genes in each state are shown in the first boxplot; chromatin accessibility is shown in the second boxplot; the right donut chart shows the fraction of the entire genome covered by each of the states as shown on the left heatmaps. "Unmarked" state is always shown in grey. TF binding and flower-related enhancers[1,7] are shown for each state in reproductive-related tissues. Gene expression and chromatin accessibility data in matched tissues from ref.[8].



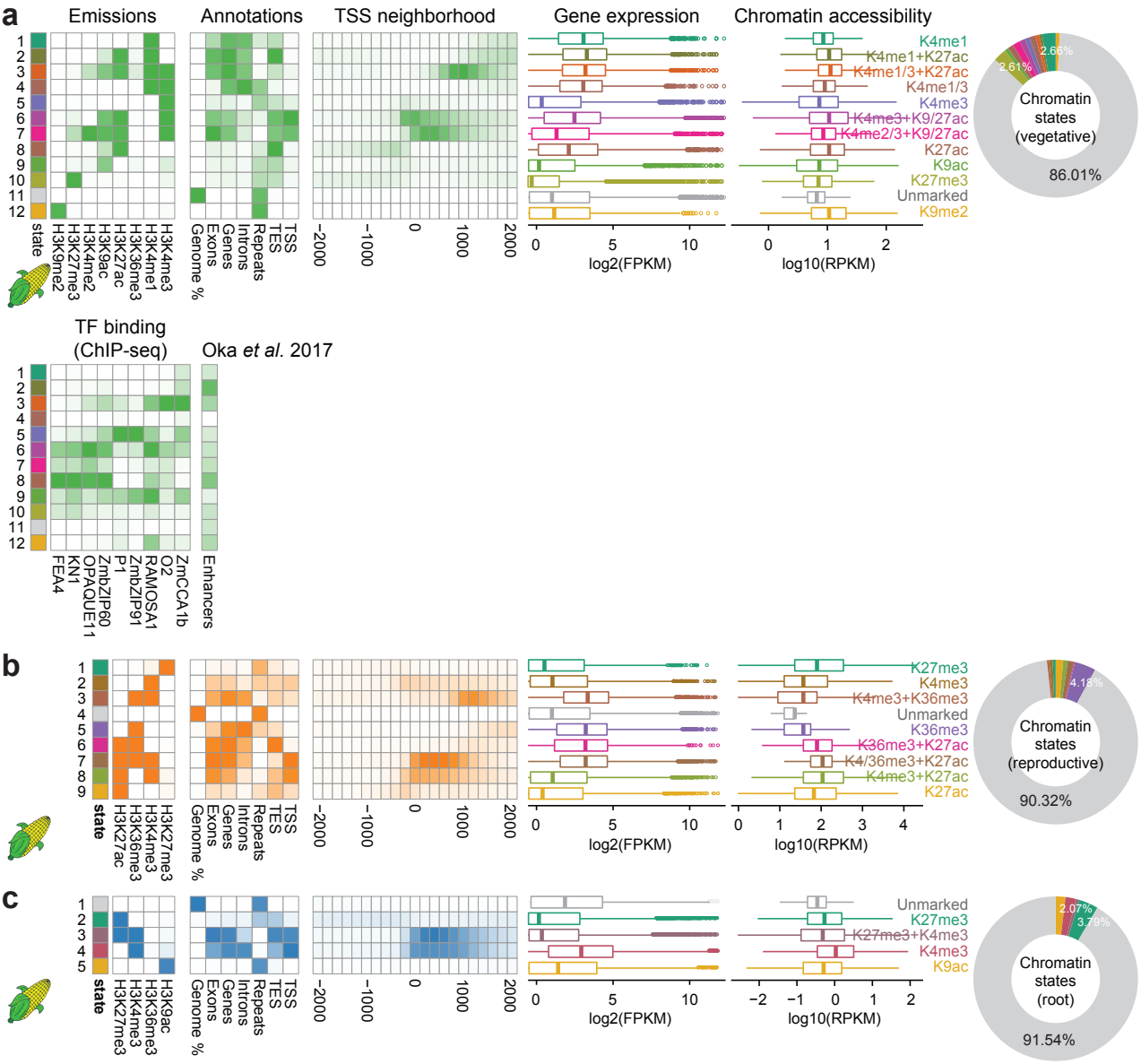
Supplementary Fig. 13: Chromatin states in barley (*Hordeum vulgare*). Refer to **Supplementary Fig. 12** for figure legend description.



Supplementary Fig. 14: Chromatin states in rice (*Oryza sativa*). Refer to **Supplementary Fig. 12** for legend description.



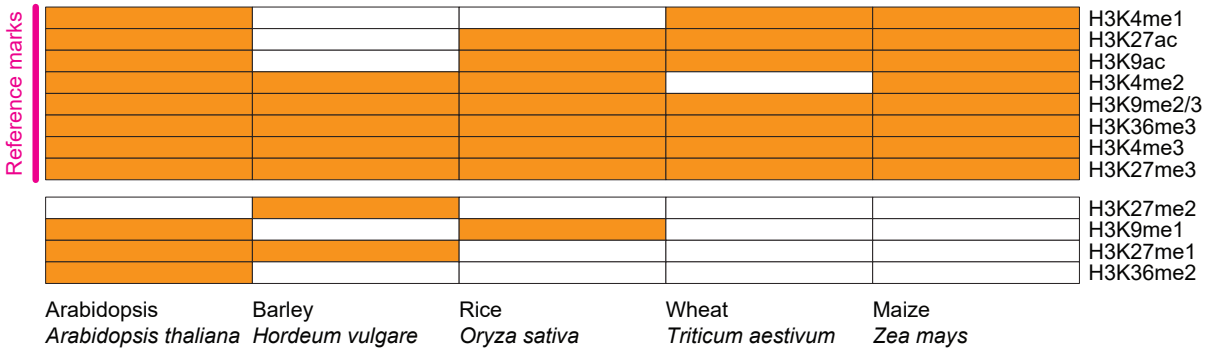
Supplementary Fig. 15: Chromatin states in wheat (*Triticum aestivum*). Refer to **Supplementary Fig. 12** for legend description.



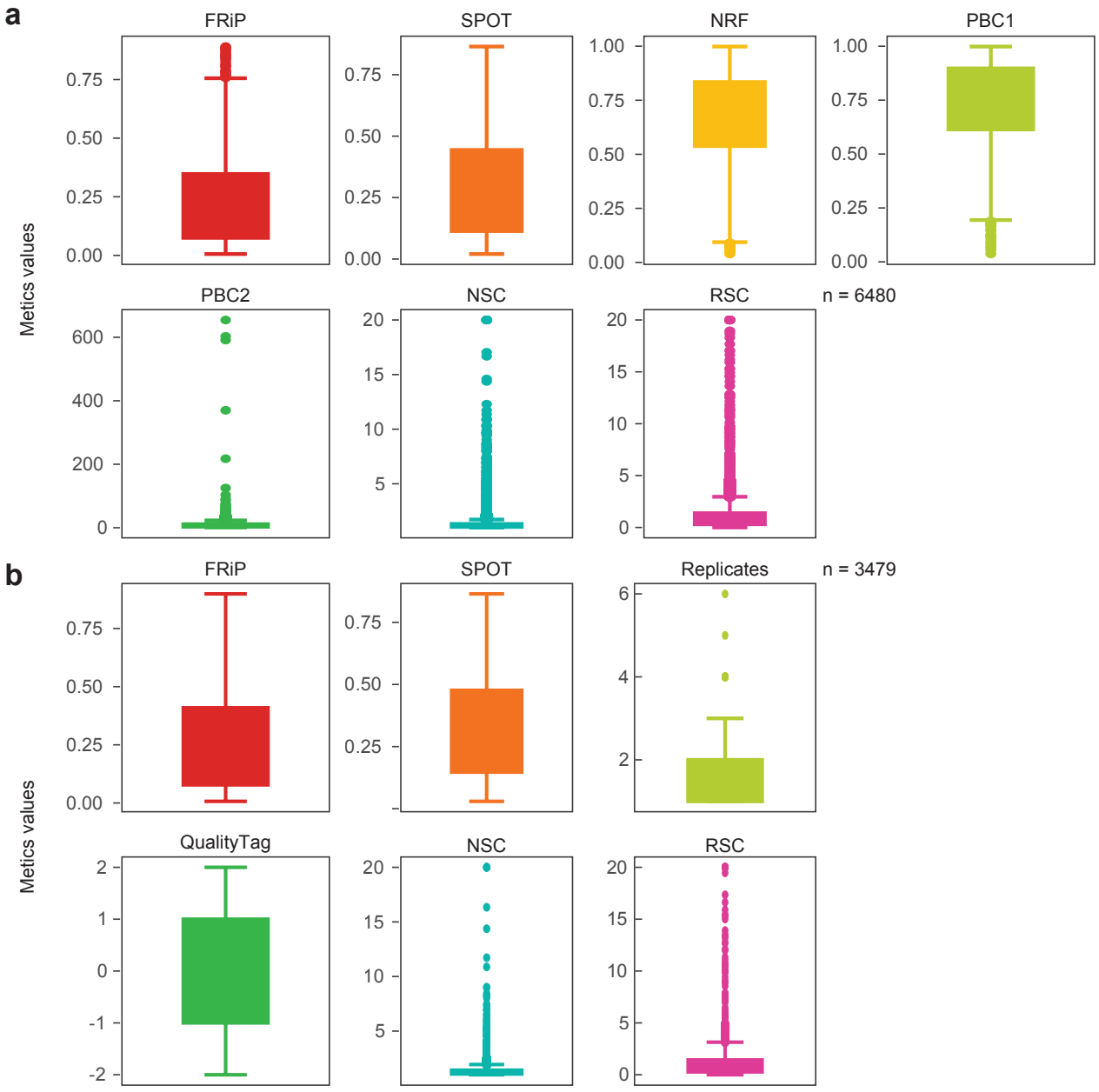
Supplementary Fig. 16: Chromatin states in maize (*Zea mays*). Refer to **Supplementary Fig. 12** for legend description. Predicted enhancers from ref.[9].

	Factor	Experiment	Replicate	ChIP	Input/control	
Li *et al.*, Nature 2018	OsGRF4	OsGRF4	Rep1	SRR7142413	SRR7142412	Use "Run" accession (SRR*) as sample accessions
	OsGRF4	OsGRF4	Rep2	SRR7142415	SRR7142414	
	OsGRF4	OsGRF4	Rep3	SRR7142417	SRR7142416	
Simonini *et al.*, Plant Cell 2017	ETT	ETT	Rep1	ERX1938605	ERX1938611	Different replicates with identical control samples
	ETT	ETT	Rep2	ERX1938606	ERX1938611	
	ETT	ETT	Rep3	ERX1938607	ERX1938611	
	ETT	ETT_IAA	Rep1	ERX1938608	ERX1938611	
	ETT	ETT_IAA	Rep2	ERX1938609	ERX1938611	
	ETT	ETT_IAA	Rep3	ERX1938610	ERX1938611	
Bencivenga *et al.*, Dev Cell 2016	BLR	BLR	Rep1	SRX1603698	SRX1603701	Different replicates with different control samples
	BLR	BLR	Rep2	SRX1603699	SRX1603702	
	BLR	BLR	Rep3	SRX1603700	SRX1603703	
Wang *et al.*, Plant Cell 2019	H3K4me3	H3K4me3_shoot	Rep1	SRX012382	NA	No control sample available and no replicate
	H3K9ac	H3K9ac_shoot	Rep1	SRX012383	NA	
	H3K27me3	H3K27me3_shoot	Rep1	SRX012384	NA	
	H3K36me3	H3K36me3_shoot	Rep1	SRX012385	NA	
	H3K4me3	H3K4me3_root	Rep1	SRX012387	NA	
	H3K9ac	H3K9ac_root	Rep1	SRX012388	NA	
	H3K27me3	H3K27me3_root	Rep1	SRX012389	NA	
	H3K36me3	H3K36me3_root	Rep1	SRX012390	NA	

Supplementary Fig. 17: Examples of metadata files. Different cases of experiment metadata. Examples are based on studies from refs.[10–13].



Supplementary Fig. 18: Types of histone modification marks used for chromatin state annotation in different plant species. Overview of histone modifications with avaible ChIP-seq data for chromatin state annotation in different plant species. A common set of marks (called "reference marks") avaible in most (if not all) species is used for comparison of chromatin states annotated in different species.



c

Metrics	Very poor	Poor	Moderate	Good	Very good
PBC1	<0.5	≥0.5 & <0.7	≥0.7 & <0.8	≥0.8 & <0.9	≥0.9
PBC2	<1	≥1 & <3	≥3 & <5	≥5 & <10	≥10
NRF	<0.5	≥0.5 & <0.7	≥0.7 & <0.8	≥0.8 & <0.9	≥0.9
QualityTag	-2	-1	0	1	2
NSC	<1.05	≥1.05 & <1.1	≥1.1 & <1.2	≥1.2 & <1.5	≥1.5
RSC	<0.3	≥0.3 & <0.8	≥0.8 & <0.9	≥0.9 & <1.1	≥1.1
SPOT	<0.05	≥0.05 & <0.1	≥0.1 & <0.3	≥0.3 & <0.5	≥0.5
FRIP	<0.05	≥0.05 & <0.1	≥0.1 & <0.3	≥0.3 & <0.5	≥0.5
Reads	<5M	≥5M & <10M	≥10M & <20M	≥20M & <30M	≥30M

Supplementary Fig. 19: Quality metrics for plant regulome data. Various quality metrics for datasets/samples (**a**) and experiments (**b**) in *Arabidopsis thaliana*. Similar plots for other plant species can be drawn via the ChIP-Hub online tool. (**c**) Quality categories for different metrics. Taken reference from the ENOCDE ChIP-seq standards (https://www.encodeproject.org/data-standards/terms/), each metircs were divided into five different levels based on their values. SPOT: signal portion of tags; FRiP: fraction of reads in peaks; NSC: normalized strand cross-correlation coefficient. RSC: relative Strand cross-correlation coefficient; NRF: non-redundant fraction; PBC1/2: PCR bottlenecking coefficients 1/2.

Supplementary References

1. Yan, W., Chen, D. & Kaufmann, K. Molecular mechanisms of floral organ specification by MADS domain proteins. *Current opinion in plant biology* **29,** 154–162 (2016).
2. Mateos, J. L. *et al.* Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. *Genome biology* **16,** 31 (2015).
3. Posé, D. *et al.* Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* **503,** 414 (2013).
4. Lee, J. H. *et al.* Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. *Science* **342,** 628–632 (2013).
5. Kelley, D. R., Snoek, J. & Rinn, J. L. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome research* **26,** 990–999 (2016).
6. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9,** 215 (2012).
7. Zhu, B., Zhang, W., Zhang, T., Liu, B. & Jiang, J. Genome-wide prediction and validation of intergenic enhancers in Arabidopsis using open chromatin signatures. *The Plant Cell* **27,** 2415–2426 (2015).
8. Sullivan, A. M. *et al.* Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. *Cell reports* **8,** 2015–2030 (2014).
9. Oka, R. *et al.* Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome biology* **18,** 137 (2017).
10. Li, S. *et al.* Modulating plant growth–metabolism coordination for sustainable agriculture. *Nature* **560,** 595 (2018).
11. Bencivenga, S., Serrano-Mislata, A., Bush, M., Fox, S. & Sablowski, R. Control of oriented tissue growth through repression of organ boundary genes promotes stem morphogenesis. *Developmental cell* **39,** 198–208 (2016).
12. Yang, Z. *et al.* EBS is a bivalent histone reader that regulates floral phase transition in Arabidopsis. *Nature genetics* **50,** 1247 (2018).
13. Wang, X. *et al.* Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *The Plant Cell* **21,** 1053–1069 (2009).