

# Epigenetically orchestrated chromatin landscapes represent the rules of functional genome organization

Vikas Shukla<sup>1</sup>, Elin Axelsson-Ekker<sup>1</sup> and Frédéric Berger<sup>1</sup>

<sup>1</sup> Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Vienna, Austria.

Nucleosomes, the octameric protein complex consisting of histones H3, H4, H2A, and H2B represent the basic layer of the chromatin organization<sup>1</sup>. Post-translational modifications of histone tails<sup>2</sup> (mainly H3) and sequence variants of histones<sup>3</sup> (mainly H2A, H3, and H2B) constitute epigenetic instructions relevant to enzymes that can remodel the chromatin environment. We identified that certain non-random combinations of epigenetic marks are massively overrepresented in the genome. These combinations are called chromatin states<sup>4</sup>, and they represent the approximate pool of nucleosomal diversity of the cell.

The discovery of a relatively small nucleosomal diversity (26 chromatin states in *Arabidopsis*<sup>5</sup>) opened up the idea that different chromatin environments could be assembled on groups of genes with shared regulatory pathways, using these specific states. To discover recurring patterns in the chromatin state organization, we performed Affinity Propagation (AP) clustering<sup>6</sup> on the emission probabilities of epigenetic marks on bins of 200 bp along the length of all genes in *Arabidopsis*.

We identified 145 distinct clusters of chromatin state assemblies on genes, which we call Chromatin Landscapes. We identified several different variants of chromatin landscapes within the three major domains of chromatin namely- Euchromatin, Facultative Heterochromatin, and Constitutive Heterochromatin. In all cases, Chromatin Landscapes contain genes that, on average, are turned on or off at similar levels. Genes within constitutive heterochromatin, generally inactive areas, display simplicity, usually featuring a uniform chromatin state across the gene. Genes in the Euchromatin are difficult to categorize into smaller groups, likely because they are highly active and their chromatin states might be affected by epigenetic changes that occur co-transcriptionally. Finally, the genes in the chromatin landscapes of facultative heterochromatin show the most diversity in terms of chromatin state organization on genes and their respective expression levels. We suspect that these are the genes that represent the chromatin environment that is preset and can be modulated by their respective regulatory machinery in different tissues, time of development or even in response to external factors.

## Thinking gene-by-gene: from chromatin states to chromatin landscapes

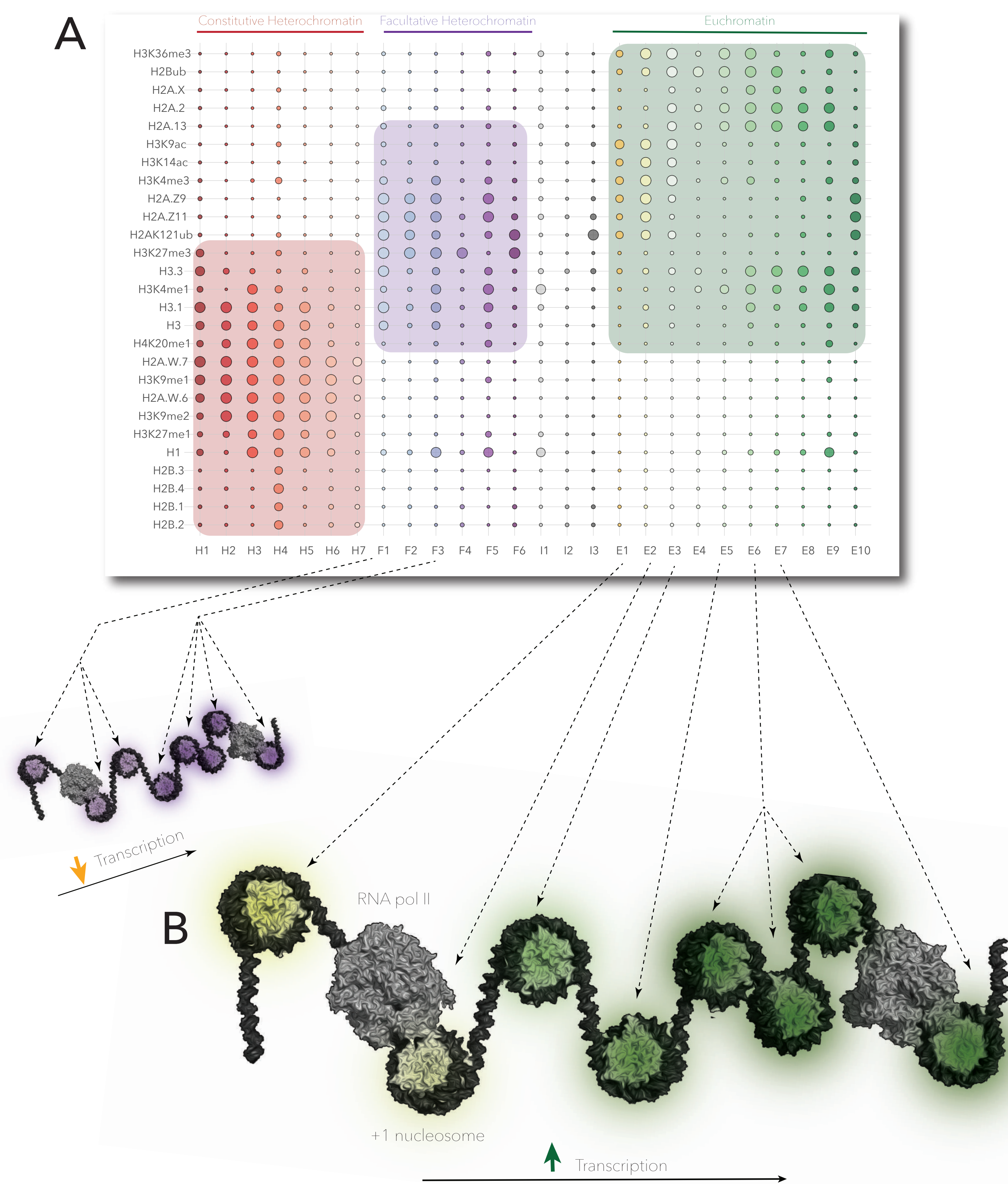
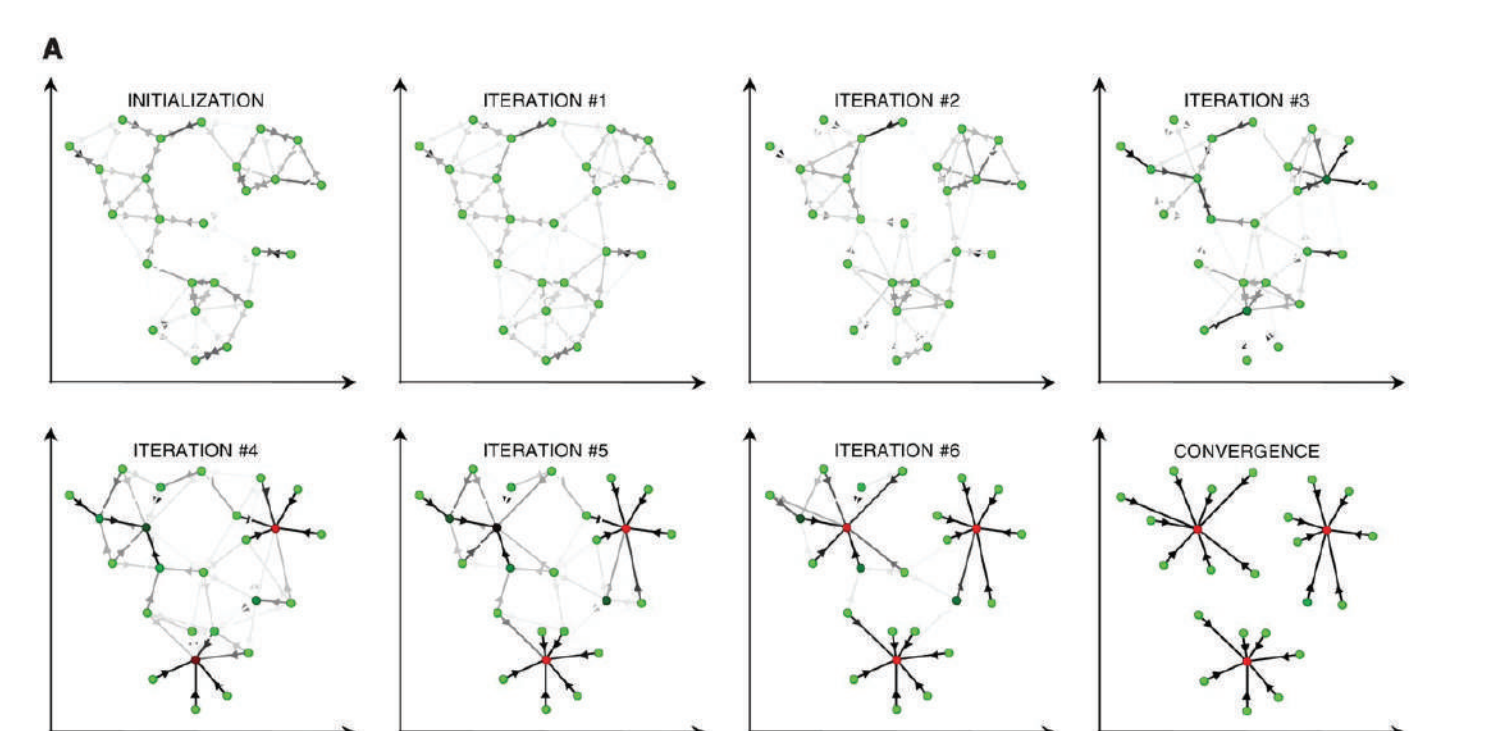


Figure 1 (A): Chromatin states of model plant *Arabidopsis thaliana*. The bubble plot shows the Emission probabilities of histone modifications and variants. The size of the bubble represent values of emission probabilities ranging between 0 and 1. (B) The concept of chromatin landscapes: The stereotypical assemblies of chromatin states on genes. Displayed are two schematic examples of chromatin landscapes, one showcasing high expression levels and the other demonstrating lower expression levels.

## Clustering groups of genes with similar chromatin state patterns on genes

Dynamic time wrapping of the emission probabilities followed by Affinity Propagation (AP) clustering on protein-coding genes predicts 145 chromatin landscapes in *Arabidopsis*



Brendan J. Frey and Delbert Dueck; Science (2007)

Figure 2: Illustration of Affinity propagation clustering by passing messages between data points. Adapted from

Brendan J. Frey and Delbert Dueck; Science (2007).

## Many ways to build a gene: towards functional genome organization



Figure 3: Arrangement of chromatin states on genes in 145 chromatin landscapes identified by Affinity Propagation clustering.

## Making sense of the complexity: merging similar groups

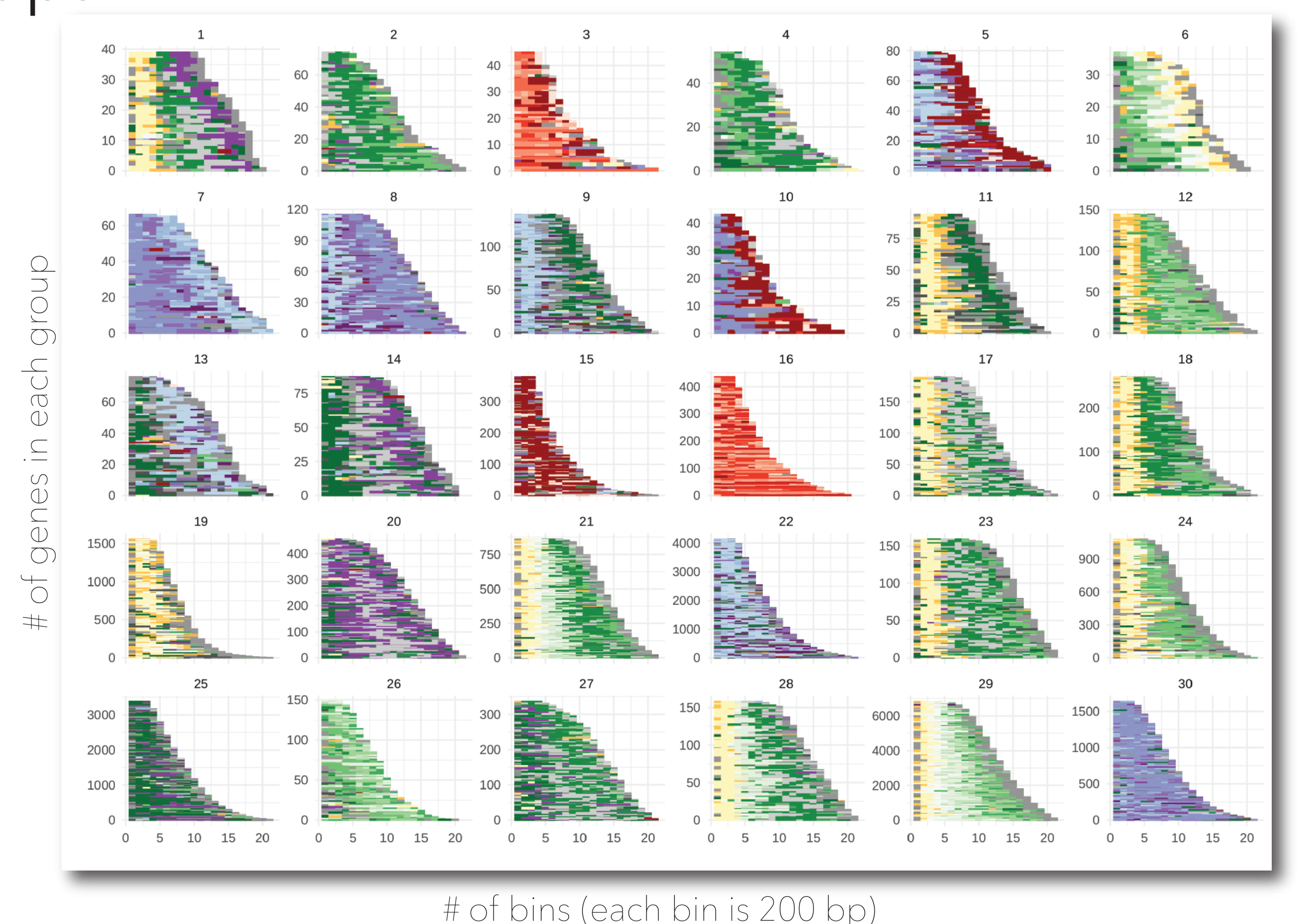


Figure 4: Agglomerative clustering of the exemplars help in merging clusters with similar patterns into fewer groups

## Summary and Outlook

1. Chromatin states represent the nucleosomal diversity of cellular chromatin.
2. Groups of genes share similar chromatin state assemblies.
3. Chromatin landscapes can be identified using Affinity Propagation clustering on the emission probabilities of chromatin states.
4. Genes within facultative heterochromatin landscapes exhibit significant diversity in chromatin state organization.
5. Chromatin landscapes may represent the presets of chromatin environment ready to be interpreted by the regulation machinery in different contexts.

### References:

1. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997 Sep 18;389(6648):251-60. DOI: 10.1038/38444. PMID: 9305837.
2. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011 Mar;21(3):381-95. doi: 10.1038/cr.2011.22. Epub 2011 Feb 15. PMID: 21321607; PMCID: PMC3193420.
3. Borg M, Jiang D, Berger F. Histone variants take center stage in shaping the epigenome. *Curr Opin Plant Biol*. 2021 Jun;61:101991. doi: 10.1016/j.cpb.2020.101991. Epub 2021 Jan 9. PMID: 33434757.
4. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*. 2012 Feb 28;9(3):215-6. DOI: 10.1038/nmeth.1906. PMID: 22373907; PMCID: PMC3577932.
5. Bhagysree Jamge, Zdravko J Lorkovic, Elin Axelsson, Akihisa Osakabe, Vikas Shukla, Ramesh Yelagandula, Svetlana Akimcheva, Annika Luisa Kuehn, Frédéric Berger (2023) Histone variants shape chromatin states in *Arabidopsis* eLife 12:RP87714 <https://doi.org/10.7554/eLife.87714.3>
6. Brendan J. Frey, Delbert Dueck, Clustering by Passing Messages Between Data Points. *Science* 315, 972-976(2007). DOI: 10.1126/science.1136800