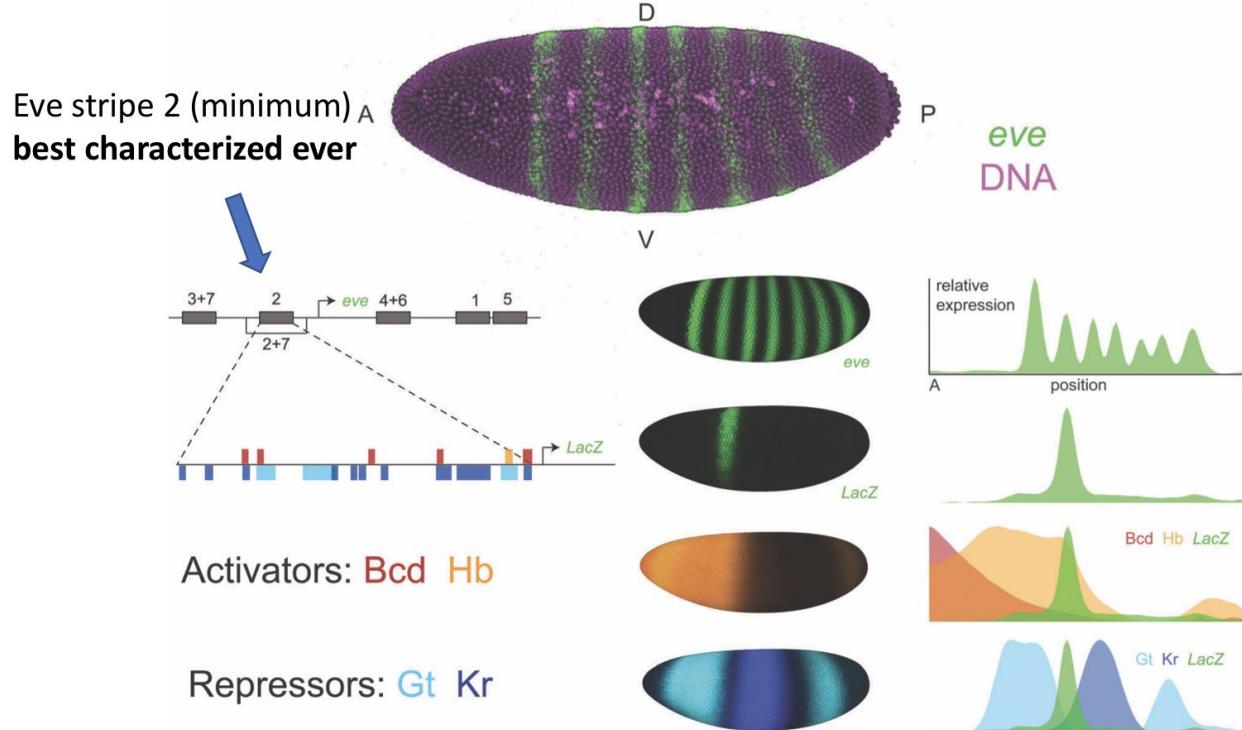


Using evolution to probe enhancer organization

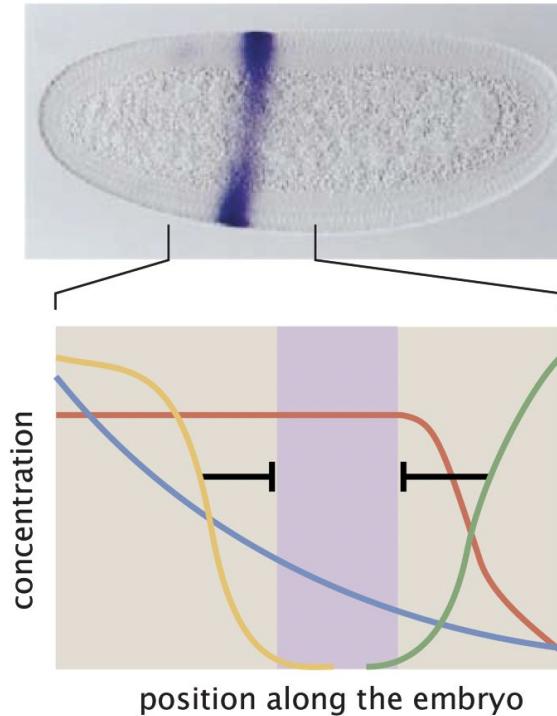
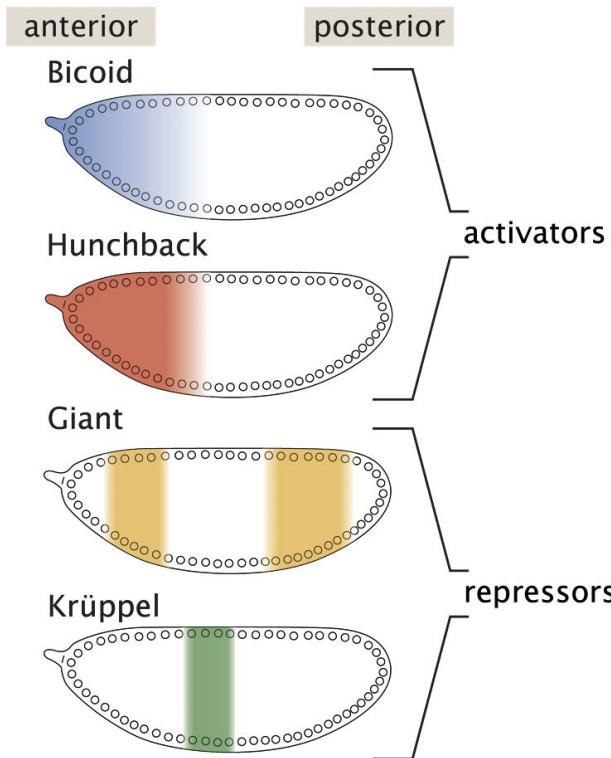
Anmol Desai
Garcia Lab

even-skipped Stripe 2: The MOST studied pattern in development



Adapted from Giana's slides

even-skipped Stripe 2: The Most Studied Pattern in Development



Small *et al.* (1992)

Adapted from Giana's slides

even-skipped Stripe 2: The Most Studied Pattern in Development



activators: **Bcd, Hb, Zld, Cad, Slp**

repressors: **Kr, Gt**

Adapted from Giana's slides

Highly contiguous assemblies of 101 drosophilid genomes

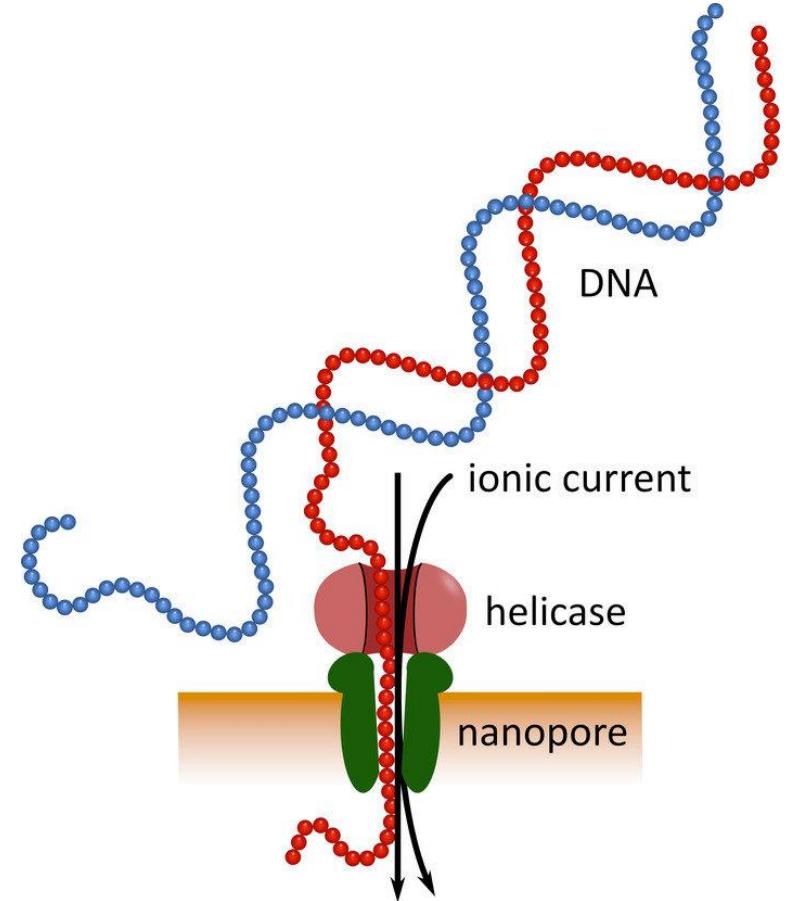


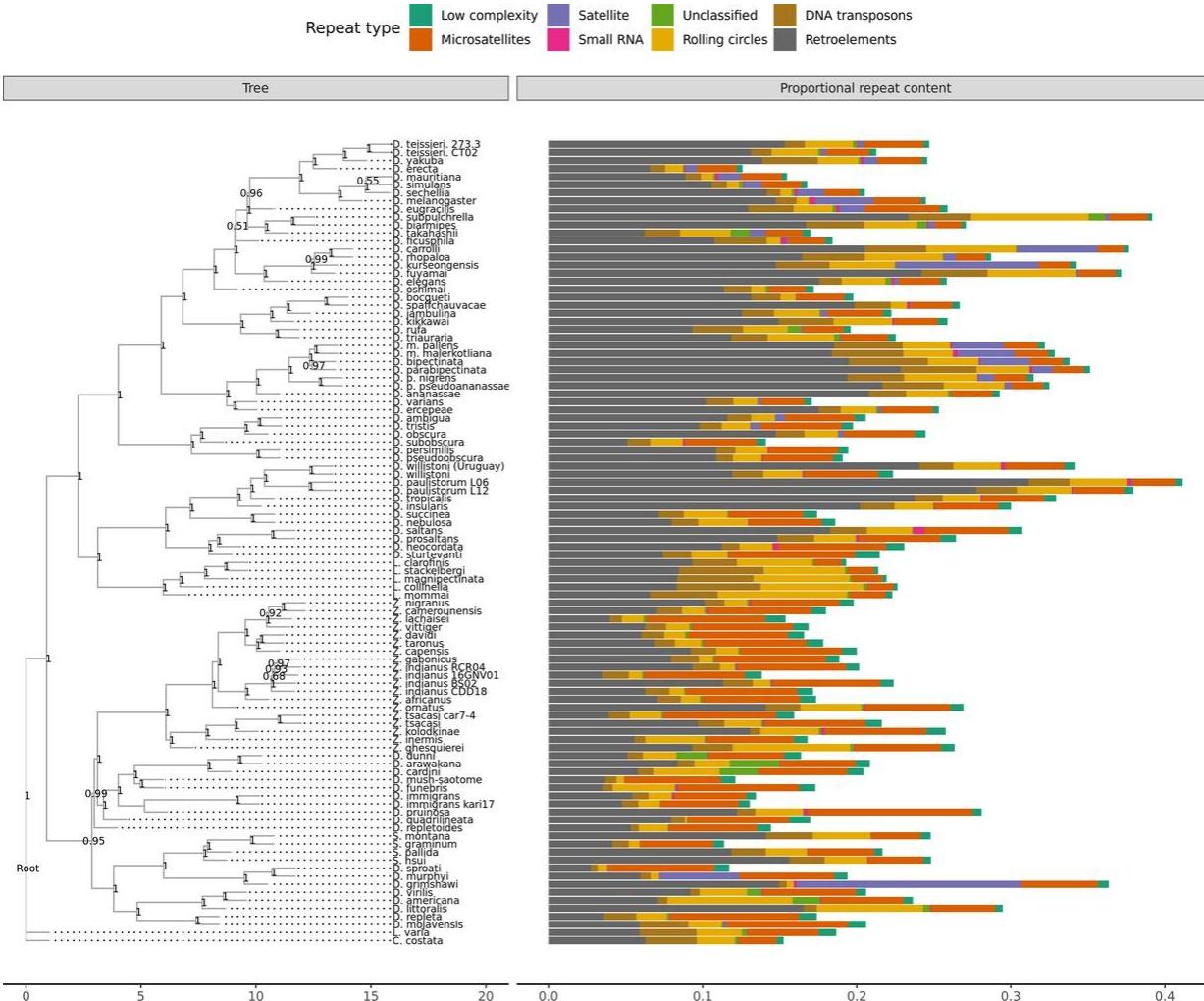
Bernard Y Kim , Jeremy R Wang, Danny E Miller, Olga Barmina, Emily Delaney, Ammon Thompson, Aaron A Comeault, David Peede, Emmanuel RR D'Agostino [see all »](#)

Department of Biology, Stanford University, United States; Department of Genetics, University of North Carolina, United States; Department of Pediatrics, Division of Genetic Medicine, University of Washington and Seattle Children's Hospital, United States; Department of Evolution and Ecology, University of California Davis, United States; School of Natural Sciences, Bangor University, United Kingdom; Biology Department, University of North Carolina, United States; Department of Integrative Biology, University of California, Berkeley, United States; Molecular and Cellular Biology Program, University of Washington, United States; Department of Biological Sciences, Tokyo Metropolitan University, Japan [see all »](#)

Data Collection

- Nanopore based assembly
- Monitoring changes to electrical current as bases are passed through protein nanopore
- Real time sequencing of ultra-long reads
- Use long reads as a guide to assembling short reads together





Beginning steps (Giana)

- Search for 100 base pair matches to minimal stripe 2 enhancer from *D. mel*
- Extract 102 potential sequences (99 species)
- Create multiple sequence alignment to find conserved regions

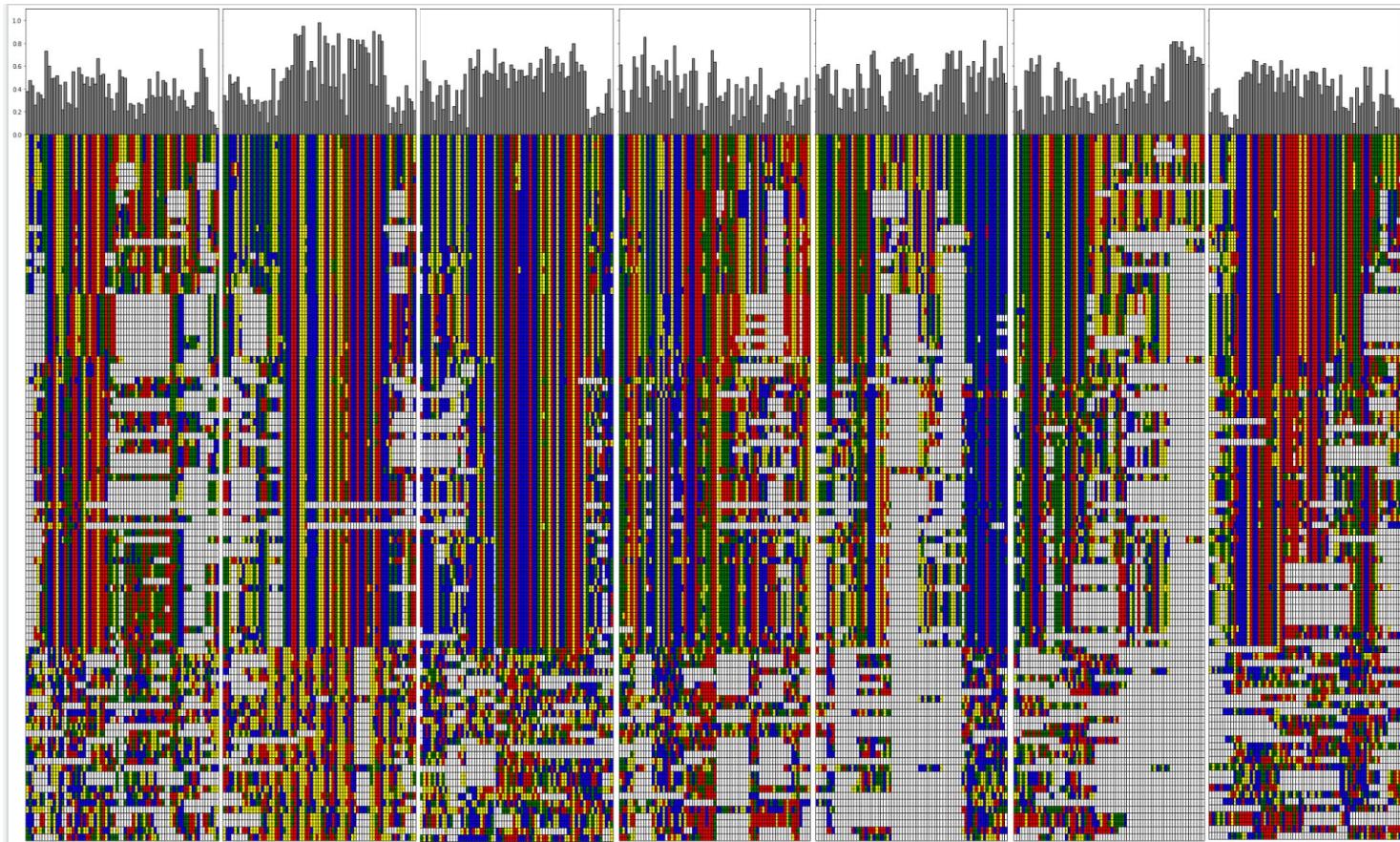
Calculating Conservation Score

- “A comparative study of conservation and variation scores” Fredrik Johansson, Hiroyuki Toh (2010)
- Selected equation

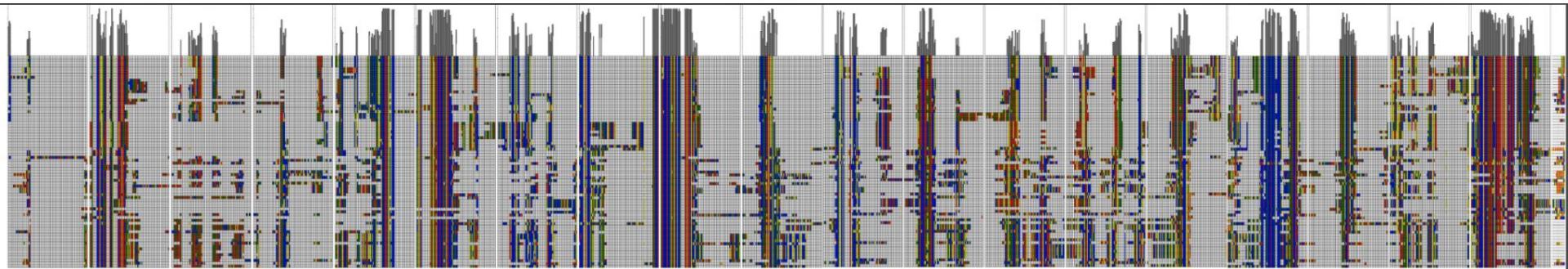
$$-\sum_{\alpha \in AA} p_k(\alpha) \log_b p_k(\alpha)$$

- $b = 4$ for four bases
- Adjusted the score by adding a constant and normalizing so lowest score = 0 and highest score = 1.

Initial Visual

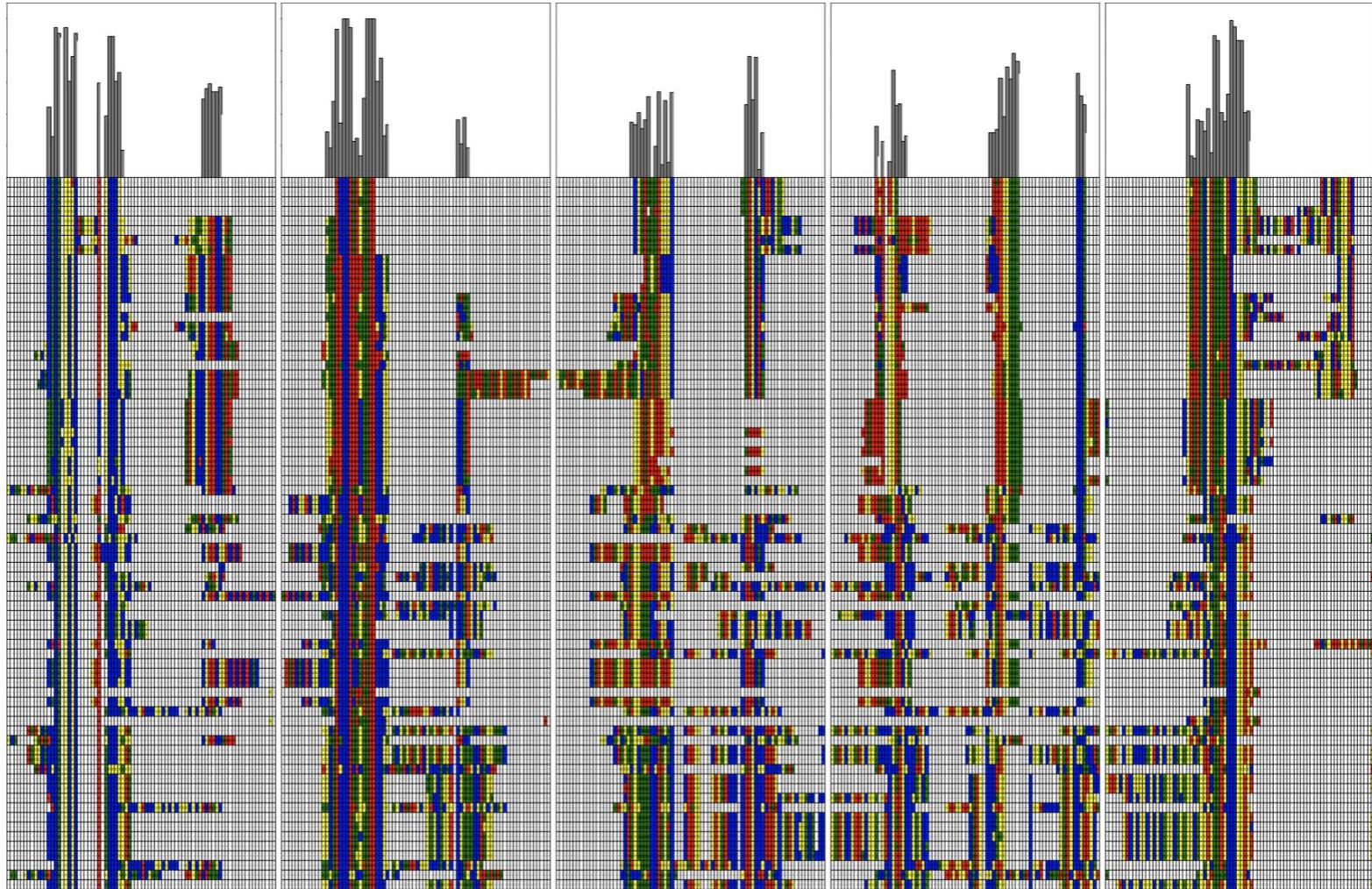


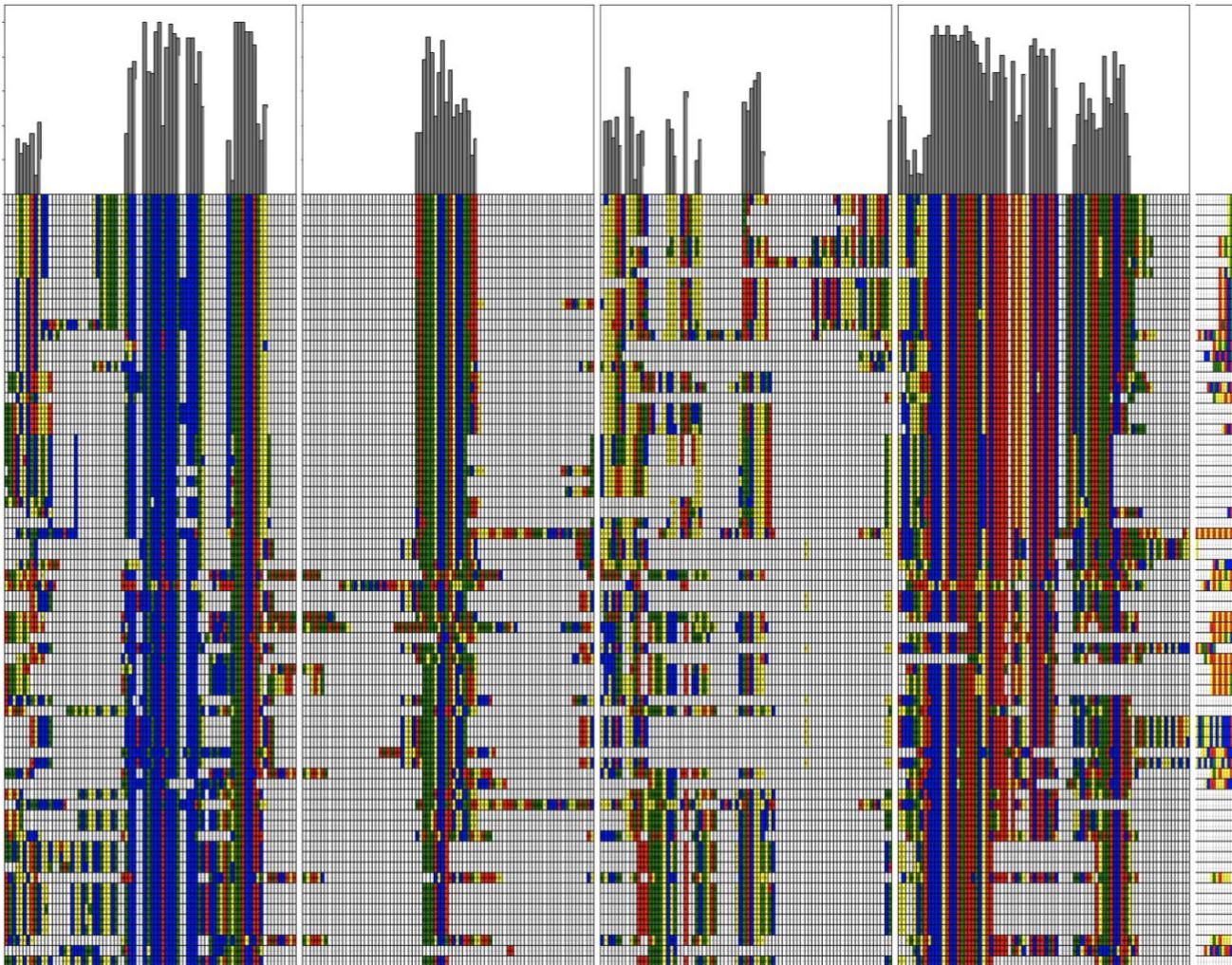
Final Alignment



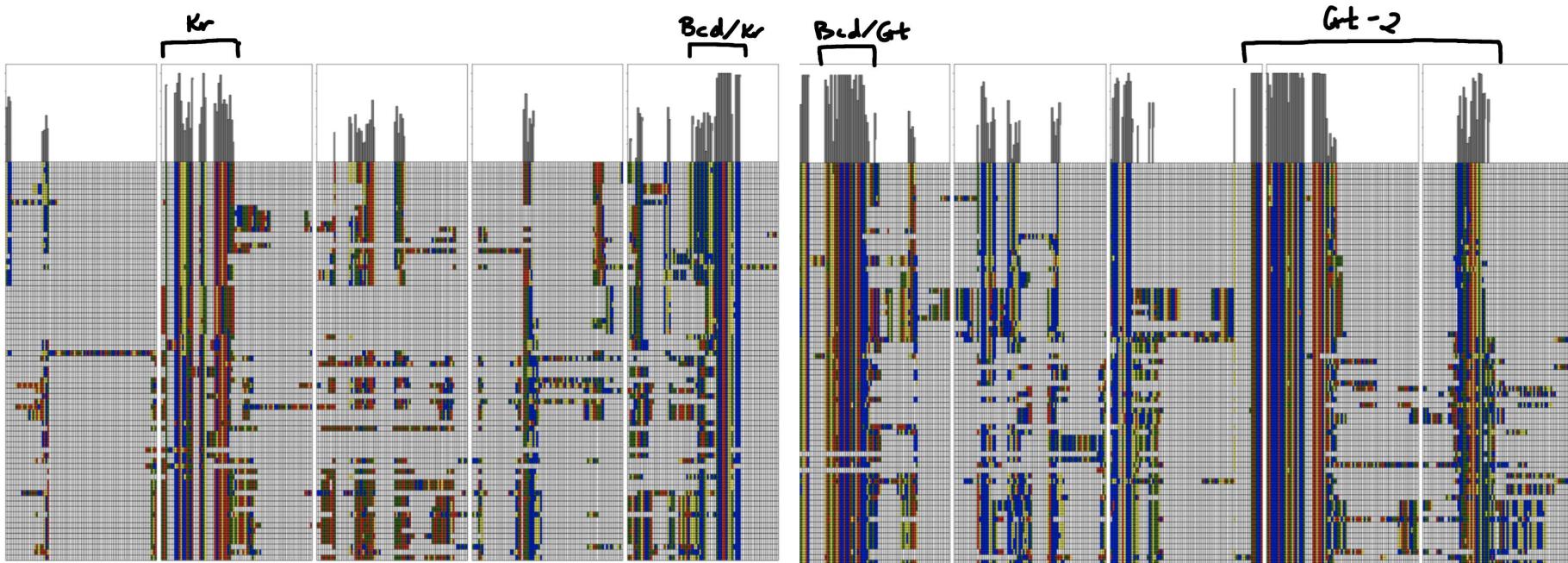








Matching to binding sites



?

?

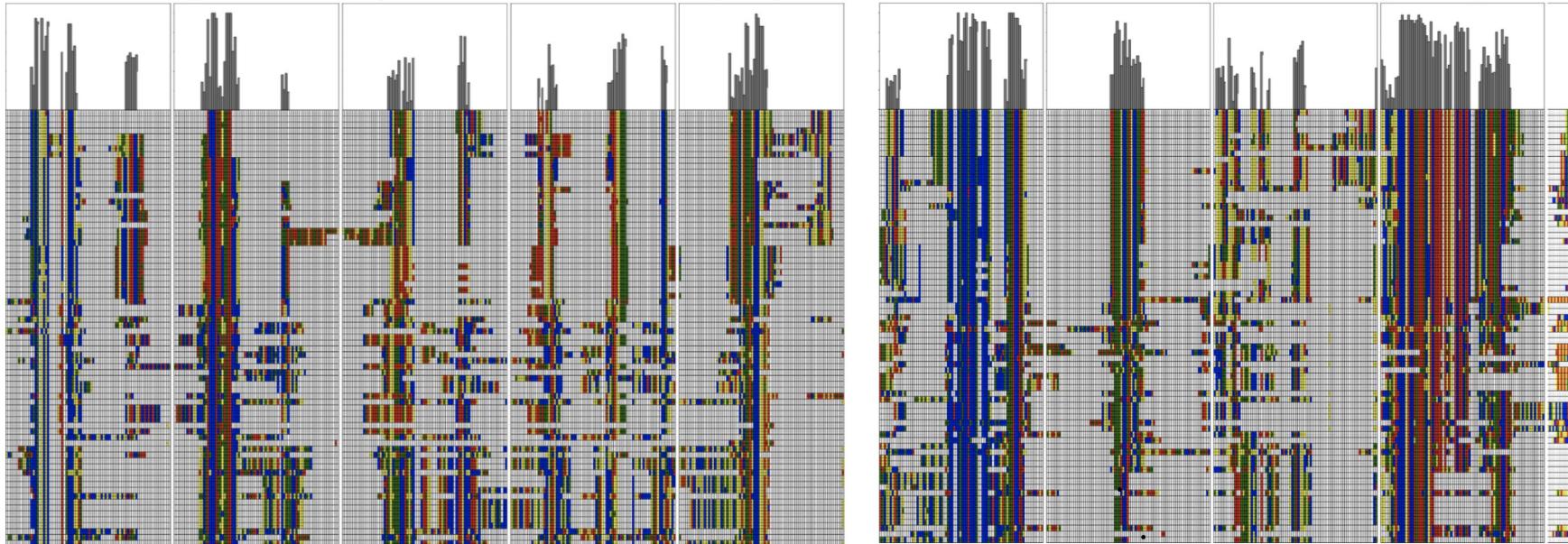
Kr

?

sfp1 bed

?

Hb cad bdy/Kr



Final alignment



Future steps:

- Low conservation between binding sites/clusters suggests the distance between them is not as important
- Look into conserved regions not associated with binding site