

Combinatorics and epigenetics of TAD boundaries in embryogenesis

Time series for DNA-DNA interactions



Dmitry Skvortsov, Anna Kalygina, Nikolay Bykov, Sasha Galitsyna

Background information

The three-dimensional structure of the chromatin drastically changes at early stages of embryogenesis by compartmentalizing and forming new spatial interactions, such as Topologically Associated Domains (TADs). TADs “mature” over time and according to their dynamic behaviour we group TADs into different clusters.

Materials and methods

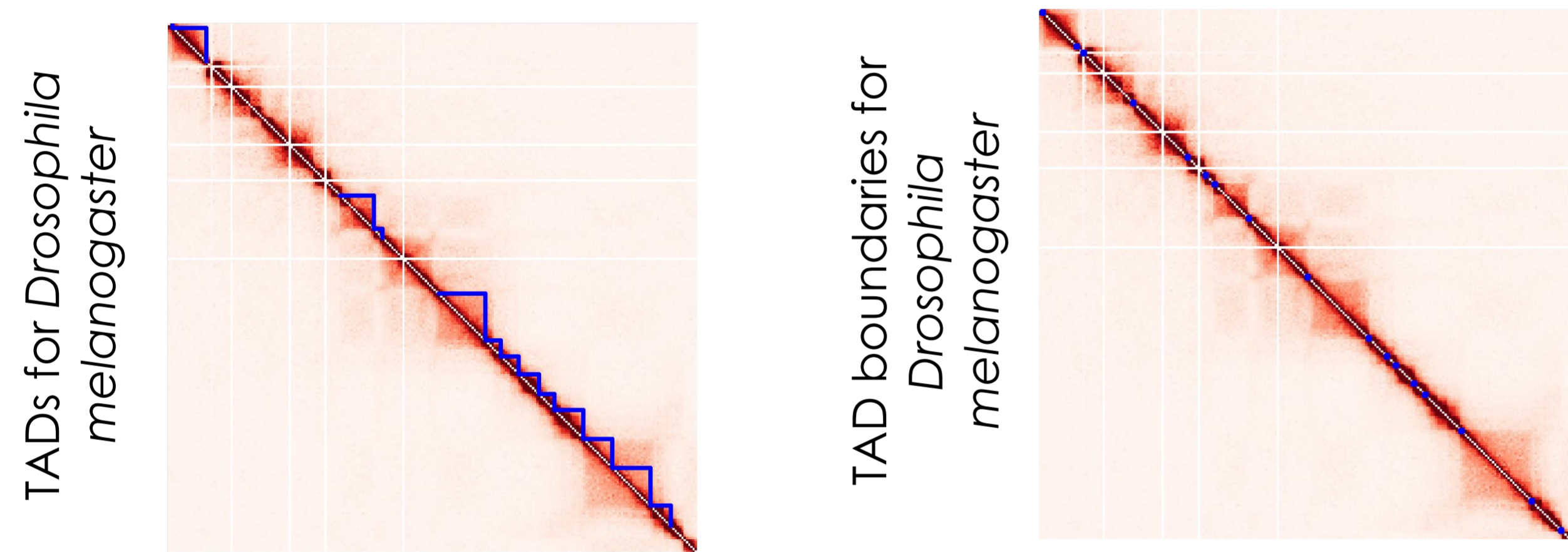
For annotation of TADs we used Hi-C data for *Drosophila melanogaster*, *Danio rerio* and *Homo sapiens* from different papers:

Organism	Paper	Resolution (number of Kb corresponding to one pixel)
<i>Drosophila melanogaster</i>	Hug et al., 2017 (hug_2017)	5 Kb
<i>Drosophila melanogaster</i>	Ogiyama et al., 2018 (og_2018)	5 Kb
<i>Danio rerio</i>	Kaaij et al., 2018 (ka_2018)	20 Kb
<i>Danio rerio</i>	Wike et al., 2021 (wi_2021)	20 Kb
<i>Homo sapiens</i>	Ke et al., 2019	40 Kb
<i>Mus musculus</i>	Du et al., 2020	40 Kb



Comparison of methodologies and materials

We annotated boundaries of TADs. For that, we used HiChew in “TAD calling” and “TAD boundaries calling” modes:



Comparison of TAD boundaries from different papers using J-index:

	Fruit Flies	Danio rerio
Jaccard index, delta 1	0.49	0.35
Jaccard index, delta 0	0.24	0.17
Expected Jaccard index	0.03	0.02

Comparison of TAD boundaries obtained by different modes using J-index:

organism	Fruit fly		Danio rerio	
	og_2018	hug_2017	ka_2018	wi_2021
delta 0	0.17	0.18	0.10	0.09
1	0.44	0.51	0.26	0.26

Jaccard index definition:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

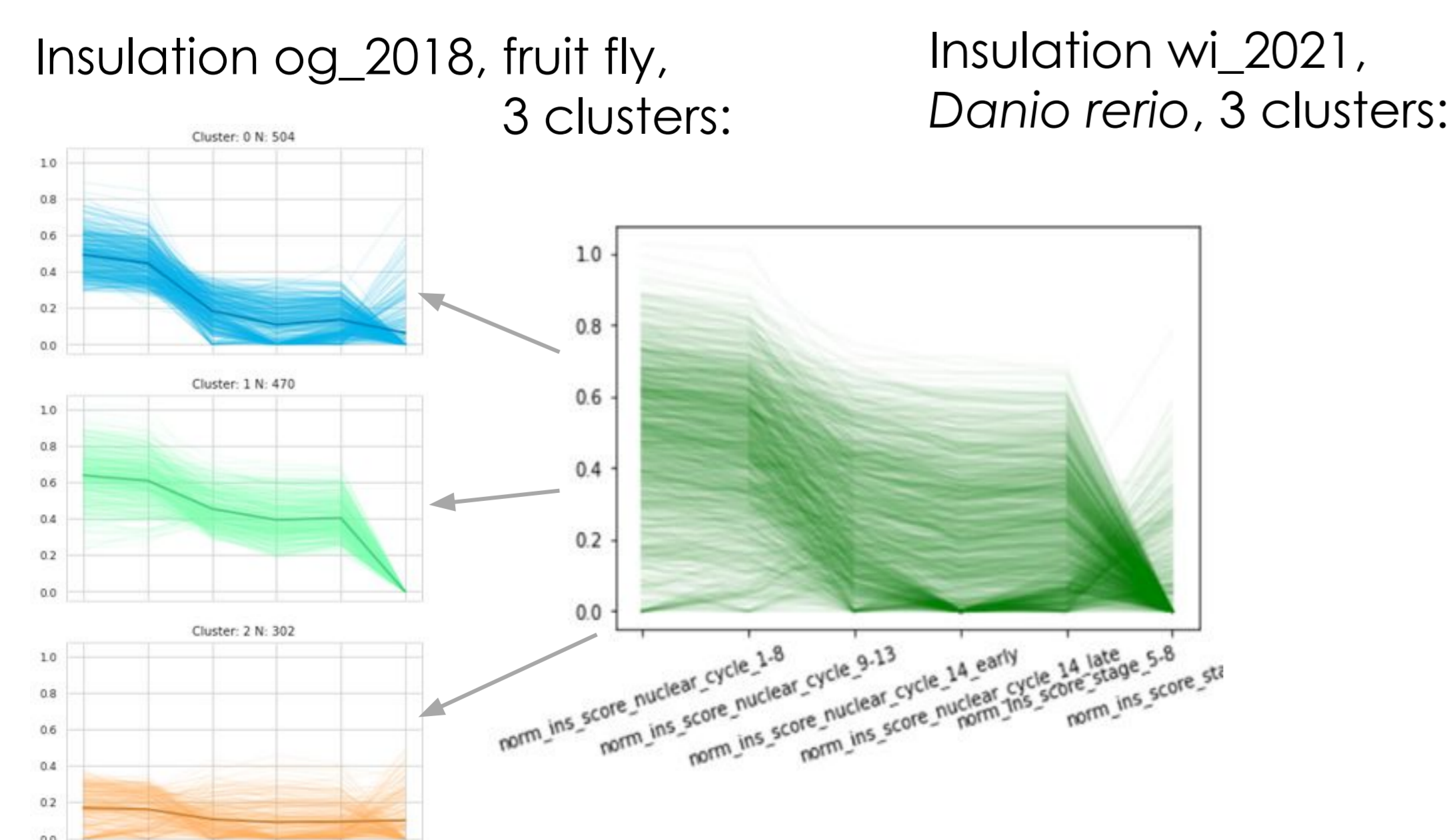
Calculation of the expected Jaccard index for two random clusters in chromatin (EJ):

$$E = \sum_{i=1}^{M_2} i \cdot \frac{C_{N-M_1}^{M_2-i} \cdot C_{M_1}^i}{C_N^{M_2}}$$

$$EJ = \frac{E}{M_1 + M_2 - E}$$

Delta is a distance measured in bins, by which the boundary can deviate from its position in another dataset - it is accounted in J-index. It reflects uncertainties in Hi-C data and boundaries calling

We then focused on boundaries calling mode and annotated temporal changes of boundaries. For this, we found insulation score for the different stages of embryogenesis and then divided all boundaries into groups according to their temporal dynamics.



Insulation score measures depletion of the contacts between TADs

Results of the data comparison

We compared the results of the methods using J-index and found that they give similar results. Then we compared data from different sources.

Data comparison by clusters using J-index:

hug_2017	cluster 0, 383	cluster 1, 557	cluster 2, 554	wi_2021	cluster 0, 845	cluster 1, 466	cluster 2, 436
og_2018				ka_2018			
cluster 0, 504	0.130	0.200	0.112	0.198	0.053	0.093	
cluster 1, 470	0.102	0.201	0.075	0.145	0.042	0.081	
cluster 2, 302	0.004	0.011	0.026	0.054	0.041	0.084	
				cluster 0, 845	0.040	0.079	0.001
				cluster 1, 59	0.064	0.109	0.006
				cluster 2, 1275	0.075	0.118	0.010

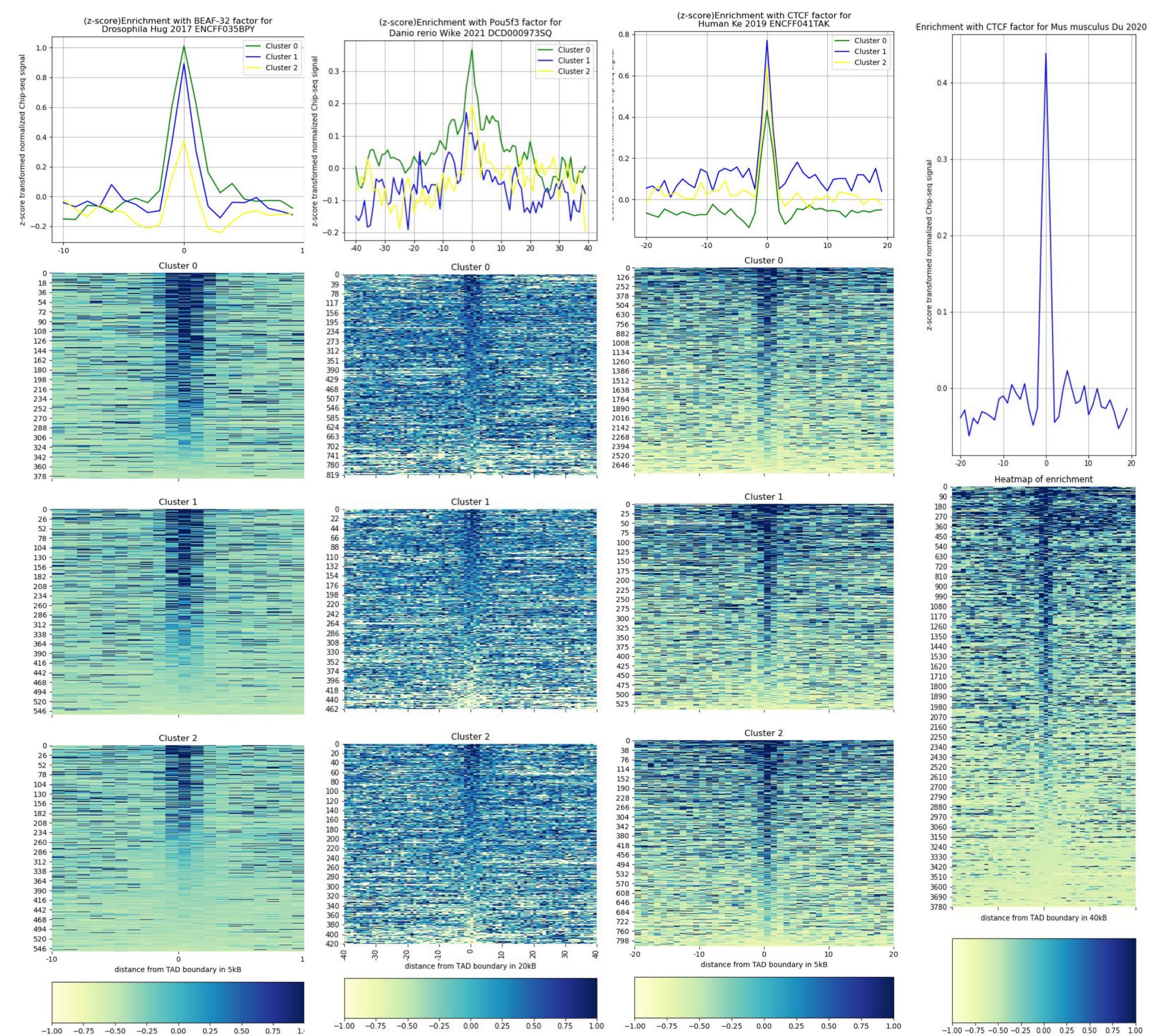
Data comparison using logarithm of observed over expected J-index:

hug_2017	cluster 0, 383	cluster 1, 557	cluster 2, 554	wi_2021	cluster 0, 845	cluster 1, 466	cluster 2, 436
og_2018				ka_2018			
cluster 0, 504	4.0	3.5	2.4	cluster 0, 845	2.9	-1.8	4.5
cluster 0, 470	3.7	3.0	2.1	cluster 1, 59	6.2	2.8	6.1
cluster 0, 302	-0.5	1.8	3.1	cluster 2, 1275	3.2	0.9	3.0

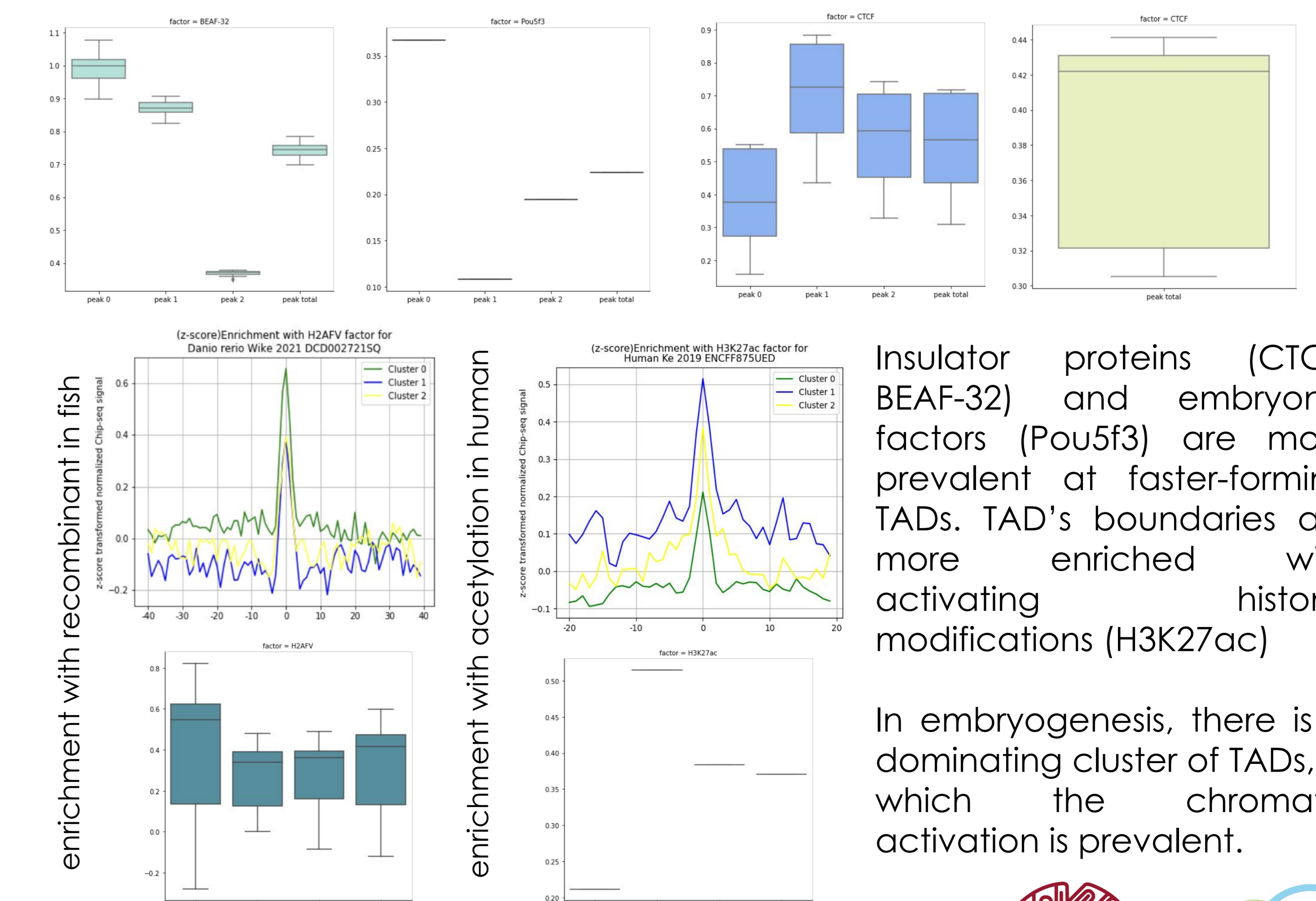
Interpretation: amber color reflects strong correspondence of clusters while blue suggests that there is rather no similarity at all. For example, based on JI, cluster 1 in ka_2018 contains boundaries from both cluster 0 and 2 in wi_2021. However, log of observed over expected is more reliable since it takes into account that Cluster 1 is very small. For fruit fly, cluster 0 corresponds to cluster 0.

Epigenetic analysis

Dynamics of TAD formation is associated with different epigenetic factors, such as insulator proteins, embryonic transcription factors and histone modifications.



Stackup and profile plots show the enrichment of the region with the factor



Insulator proteins (CTCF, BEAF-32) and embryonic factors (Pou5f3) are more prevalent at faster-forming TADs. TAD's boundaries are more enriched with activating histone modifications (H3K27ac)

In embryogenesis, there is a dominating cluster of TADs, in which the chromatin activation is prevalent.