Chromatin-based Temporal Clustering of Enhancers in Developing Fly Embryos

Arbel Moshe **Tommy Kaplan**

School of Computer Science and Engineering

The Hebrew University of Jerusalem

Abstract

We describe a chromatin-based clustering method, used to find novel transcription factors associated with the establishment of transcriptional enhancers during the maternal-to-zygotic transition (MZT) in the fruitfly (*D. melanogaster*) embryos.

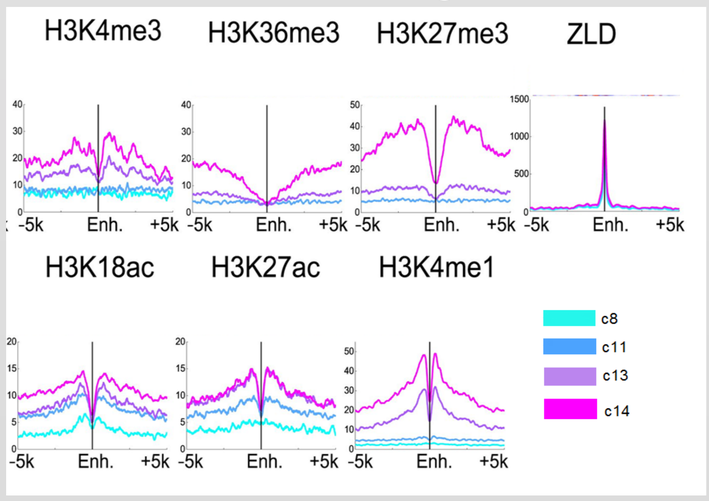
The protein Zelda was shown to have a key role during early embryonic development in Drosophila. Prior to MZT, Zelda binds thousands of promoters and enhancers, and marks them for transcriptional activation [Harrison *et al*, 2011]. Recently, we showed that Zelda acts through specific chromatin patterns of histone modifications to mark developmental enhancers [Li *et al*, 2014]. Intriguingly, some Zelda sites still maintain these chromatin patterns, even in *Drosophila* embryos lacking maternal Zelda. This suggests that additional Zelda-like pioneer factors may act in early fly embryos.

To test this hypothesis, we analyzed ChIP data near early (mitotic cycle 8) Zelda peaks, to characterize their temporal dynamic through MZT (mitotic cycles 8, 11, 13, and 14). Specifically, we focused on H3K4me1 and me3, H3K18ac, H3K27ac, and H3K27me3 and identified three main categories, matching "Enhancers", "Promoters", and "Transiently-bound" peaks. We then used these unique chromatin patterns to further scan the genome and identify novel loci showing similar characteristics with no Zelda binding.

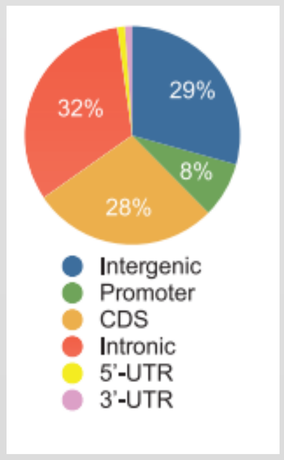
Introduction - Zelda as pioneer factor

- Early Zelda binding (at cycle 8) "marks" enhancers of early developmental genes [1,2].

Enhancers modifications through MZT:



Early (cycle 8) in vivo ZLD binding loci:



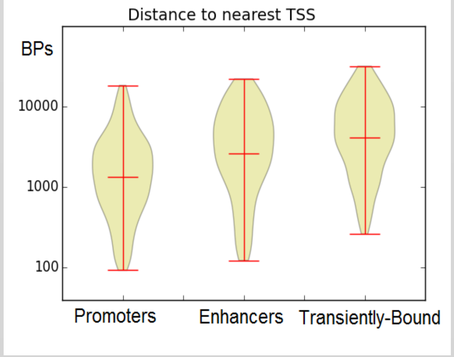
- **However,** some enhancers are still established, even in maternal ZLD mutants [2,3,4].

Zelda Peaks Clustering

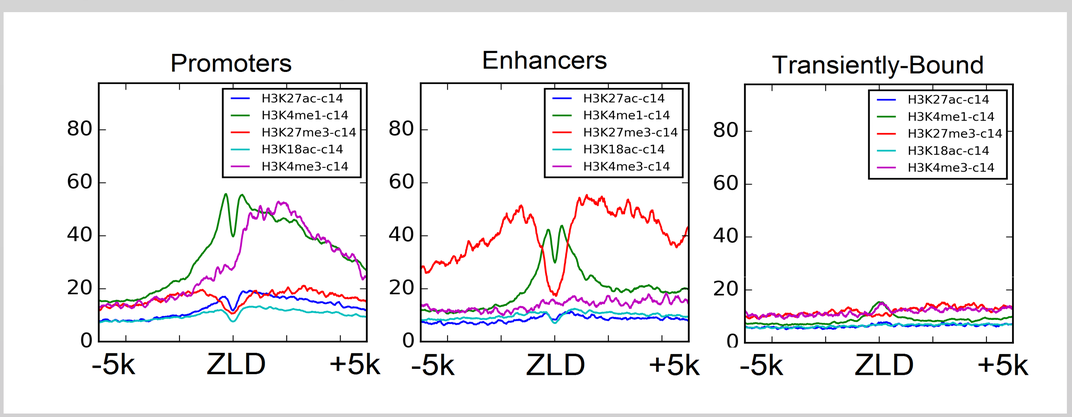
- We want to identify what other factors could "backup" Zelda.

Spectral Clustering followed by K-Means:

- Data: log-ChIP of H3k27ac, H3k4me1, H3k27me3, H3k18ac, H3k4me3 around top 2000 early Zelda peaks (10K bps window for each HM).
- Metric: RMSE.
- Gaussian Kernel: $e^{-\frac{dist^2}{2\sigma^2}}$. σ fitted to each HM separately.

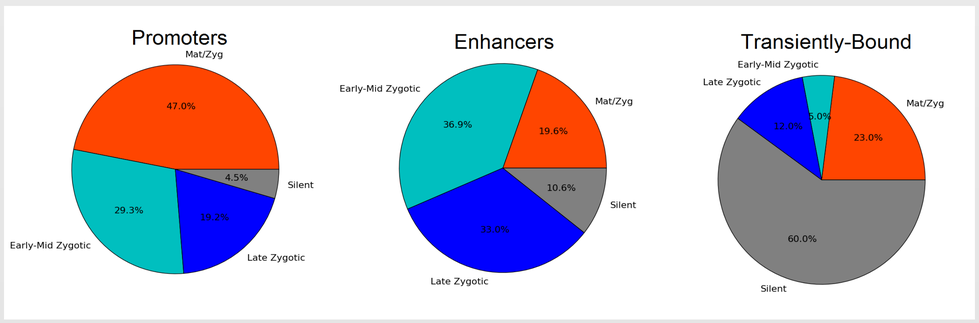


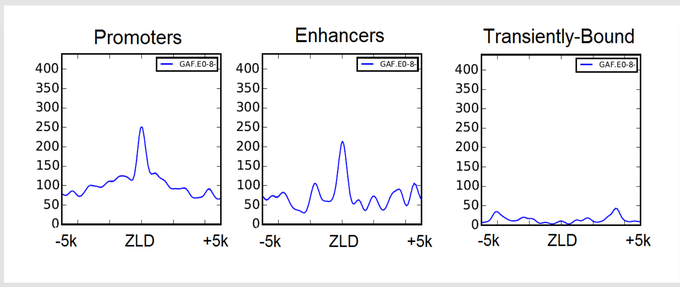
Unique modification patterns for clusters:



Clusters Behaviour

- *Transiently-bound* peaks correspond to mostly silent genes.
- *Promoters* of early zygotic genes, as well as *Enhancers* of early/late zygotic genes.



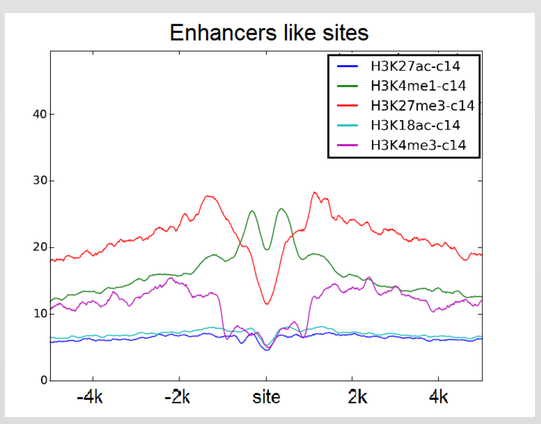


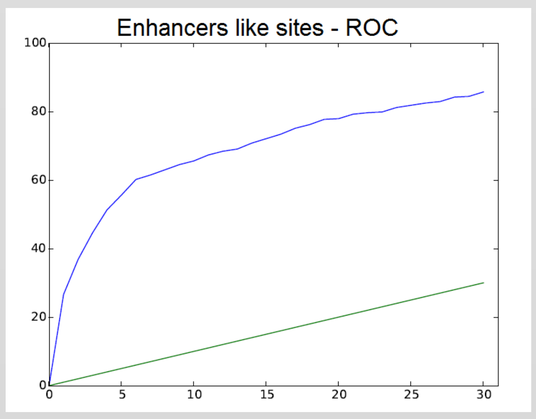
- GAF ChIP in WT embryos, at 0-8hr [5].
- *Promoters* and *Enhancers* peaks, are enriched in GAF.
- May GAF be the "Other factor"?

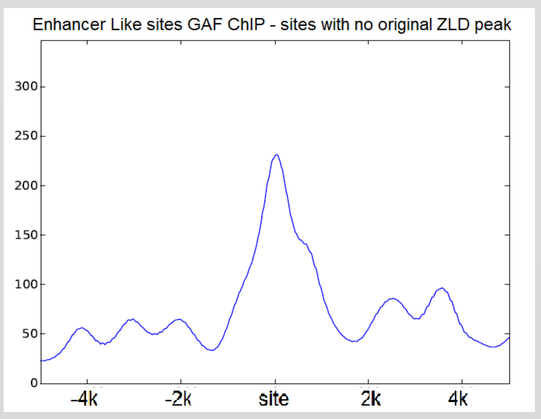
Identification of ZLD-independent enhancers

Scanning the genome for the patterns:

- cross-correlation of the patterns with HMs in the entire genome.
- We concatenated the HM patterns to preserve the ratio between them.
- (Correlation X Scaling Probability) as score.
- Search for "Enhancers" produced ~2000 similar sites.







- Sites are enriched in GAF, even when we look at sites which didn't have an early Zelda peak.

Discussion

- Does GAF affect chromatin like Zelda?
- Are there any other factors?
- How does Zelda know to "run away" in *Transiently-bound* peaks?

References

1. Harrison et al, "Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition", PLoS Genetics, 2011.
2. Li et al, "Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition", eLife, 2014.
3. Xu, Small et al, "Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila." Genes Dev, 2014.
4. Schulz et al, "Zelda is differentially required for chromatin accessibility, transcription-factor binding and gene expression in the early Drosophila embryo", *under review.*
5. Negre et al, "A cis-regulatory map of the Drosophila genome". Nature, 2011.