Leveraging multi-omics data to identify genetic-epigenetic interactions and view genome structures

Lauren Kuffler, Daniel A. Skelly, Anne Czechanski, Gary A. Churchill, Steven C. Munger, Christopher L. Baker, Laura G. Reinholdt, Gregory W. Carter

The Jackson Laboratory, Bar Harbor, Maine 04609 USA
Tufts School of Graduate Biomedical Sciences, Boston, Massachusetts 02111 USA

Questions? Contact me at lauren.kuffler@jax.org.

The Jackson Laboratory
Leading the search for tomorrow's cures

Tufts

ABSTRACT

Highly-characterized, genetically diverse mouse models are permitting us to examine how areas of open chromatin and sequence variants interact and influence gene expression, within the context of local genome structure.

INTRO

Our questions: How often do genetic-epigenetic interactions happen? Where? What's the biological basis of a gene's "local regulatory area"?

METHODS

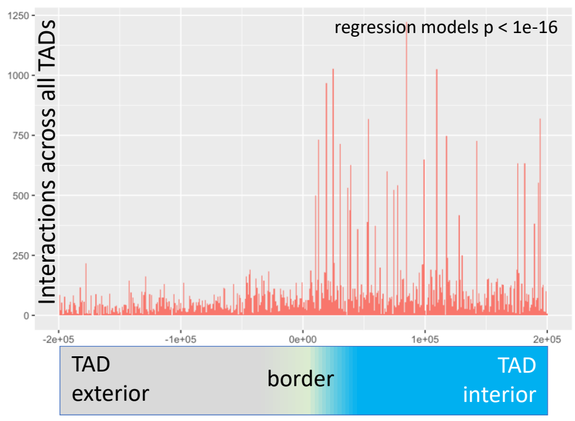
We used 176 samples of Diversity Outbred (DO) mouse embryonic stem cells (mESCs) with GigaMUGA sequencing, ATAC-seq and RNA-seq data. We used a regression model to search for putative interactions.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon_i$$

y_i= expression, x_1= SNP, x_2 = ATAC, $x_1 x_2$ = SNP:ATAC
Equation 1: our regression model. Each combination run from our data was tested to find how many terms were needed to describe its behavior, including evidence of interaction between SNPs and ATAC peak presence.

Interactions in TADs

Interactions inferred with the highest degree of confidence (adj. p < 1e-16) were prevalent throughout our results, and appear preferentially when all interacting elements fall in the same topologically associating domain (TAD). The TAD border can be easily seen in this aggregate view of interaction-involved ATAC peaks:

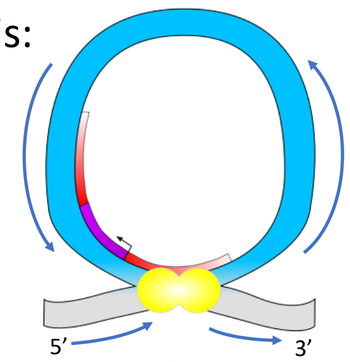


This indicates TADs may contain or constrain the local regulatory area of a gene.

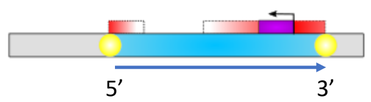
Genome structure informs regulation

Because TADs pin DNA into ~1 Mb loops, distant areas of chromatin can be brought into close proximity. Our question: Gene regulation is known to preferentially occur close to the gene. Do TAD loops shape what counts as "close"?

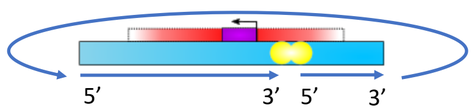
We compared interaction density in TAD loops, between the standard graphing method (linear DNA sequence), and a method that takes TAD structure into account: center on the gene promoter, and terminate halfway around the TAD loop in both directions. If the gene's local area looks like this:



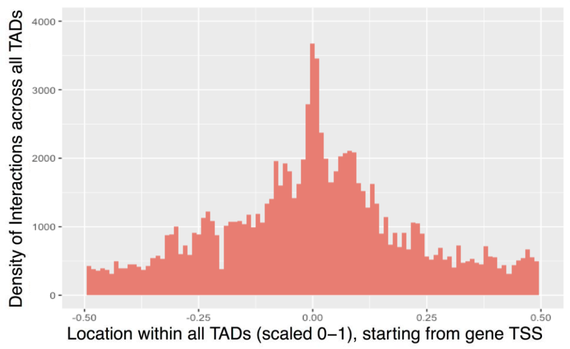
Then it may make more sense to look at it less like this:



…and more like this:



Plotting intra-TAD interacting ATAC peaks by this method gives us a gene-centric peak in interaction density. There is no detectable bias based on TAD start/end site location.



"But what if I don't have TAD data?"

We can provide a revised guideline for how far researchers need to search from a gene TSS for interacting elements if they want to capture 95% of all intra-TAD interactions:

1.29 Mb upstream
1.56 Mb downstream
2.84 Mb total

CONCLUSIONS

- Regression modeling indicates presence of genetic-epigenetic interactions
- Interaction prevalence indicates they may affect reproducibility of variant effects
- Topologically associating domains constrain interactions

Work in progress

- Can TAD boundaries be identified via interaction data?
- How much influence do these local interactions have on downstream gene networks?
- Can we perform an allele swap experiment that demonstrates this?

www.jax.org