PIONEER FACTOR FOCUSED INTRO

**Gene expression important for phenotypic evolution**

Gene regulation is an essential molecular process that guides spatiotemporal gene expression throughout development and organismal life. As such, variation in gene regulatory mechanisms often underlies the evolution of phenotypes such as morphology and behavior (cite). An essential step of gene regulation is transcriptional activation, in which transcription factors bind to cis-regulatory elements to recruit transcriptional machinery. However, the role of other processes, such as chromatin remodeling, is now better understood as an important part of gene regulation, and therefore a place for variation to alter gene expression.

**Chromatin remodeling increasingly well studied and being integrated into our understanding of evolutionary mechanisms**

Chromatin remodeling involves making DNA accessible to allow for subsequent binding of regulatory proteins for transcriptional activation. The molecular mechanisms by which chromatin is made accessible are complex and multifaceted but seem to occur on a chromosome-wide level, in which larger regions are remodeled by X and X, and 2), as well as at a more locus-specific level, in which various transcription factor-based mechanisms determine the accessibility of a regulatory region. More specifically, the molecular mechanisms of chromatin remodeling at regulatory regions is relatively well-understood for *Drosophila* epithelial tissue, in which there are broadly two mechanisms by which regulatory regions are made accessible and activated. Our current model suggests that either 1) the pioneer factor Grainyhead opens up and primes regulatory regions for subsequent activation, or 2) collective TF binding both opens up and activates regulatory regions. The distinction between these various mechanisms is important for understanding the role of chromatin remodeling in gene expression evolution because the gene expression may be altered in different ways, allowing for different types of variation.

Previous work has documented a consistent trend across organisms: chromatin accessibility variation is largely due to *cis* changes and is often not correlated with gene expression variation. These findings shed light on the mechanistic link between chromatin remodeling and gene expression variation, but yet to be studied is how these patterns may vary based on chromatin remodeling mechanism. To address this knowledge gap, we collected chromatin accessibility and gene expression datasets from two *Drosophila* species and their hybrids in the imaginal wing disc tissue. Using the imaginal wing disc allows us to draw on prior work demonstrating regions for which Grainyhead is necessary for accessibility, allowing us to investigate and contrast other regions. We find supporting data for prior work as well as interesting patterns on how Grainyhead regions evolve.

*Experimental schematic*

To interrogate the mode of chromatin accessibility and gene expression evolution, we collected ATAC- and RNA-seq data from the imaginal wing discs of D. melanogaster and D. simulans and the F1 hybrid. These strains were chosen because of 1) their capacity to interbreed and 2) the extensive previous work to catalog the genomic variation between these species. Following alignment and processing, accessibility was measured for four genomic categories: +/- 500bp transcriptions start sites (TSS), +/- 500bp transcriptions end sites (TES), +/- 500bp center of called peaks that fall within coding regions (intragenic), and +/- 500bp center of called peaks that fall between coding regions (intergenic). As expected, variation was greatest for intergenic regions, then intragenic, and then TSS and TES. After filtering regions with low read counts, we had X, X, X, and X numbers per group. We then used an empirical Bayes method to estimate the difference between parental strains, hybrid alleles, and the parental vs hybrid difference. This provided us with divergence estimates as well as statistical metrics to determine cis and trans categories.

FIGURES NEEDED: HEATMAPS OF EACH CATEGORY AND REVISED PHYLOGENY (INCLUDE RNA-SEQ NOW?)

*High divergence due to cis changes is constant across genomic categories*

To quantify the amount of chromatin accessibility divergence due to *cis* changes, we used the **percent *cis*** metric which infers such from parental and hybrid allele differences. With this metric, chromatin accessibility regions are, on average, 75% due to *cis* changes, consistent with prior work in fly embryos, yeast, and stickleback fish. We then contrasted percent cis across the different functional groups to see if there were any differences between groups since the mechanisms giving rise to accessibility between these groups are likely different. Interestingly, percent cis remained within +/- for all groups, with only statistically significant differences between X and Y, with divergence being X and Y.

FIGURES NEEDED: (A) CISxTRANS (B) PERCENT CIS DISTRIBUTION (C) EXTENT AND (D) PERCENT CIS ACROSS CATEGORIES

*Extent and mode of chromatin accessibility divergence differs for regions made accessible by a pioneer factor.*

To contrast the extent and mode of chromatin accessibility divergence by regions made accessible by different molecular mechanisms, we took advantage of previous work that identified regions necessary for accessibility in imaginal disc epithelial tissue. More specifically, regions for which Grainyhead KOs resulted in a loss of accessibility, had Grainyhead binding, and overlapped with accessible regions in the data collected here, and were defined as pioneer factor accessible. Interestingly, these regions are overall less divergent but more often due to cis-regulatory changes. Furthermore, we verified that these differences in chromatin accessibility are associated with Grh motif differences more frequently than expected by chance. Taken together, these results are consistent with our understanding that 1) Grainyhead primes constitutively active regulatory regions in epithelial cells which may be under more constraint, and 2) the Grainyhead protein is essential for genome-wide chromatin remodeling and therefore trans changes may be under more constraint than regions that are remodeled by a multitude of TFs with more specific functionalities. These findings highlight the relevance of understanding the molecular mechanisms to predict evolutionary outcomes.

FIGURES NEEDED: (A) SCHEMATIC FOR CATEGORIZING GRH-BOUND KO NECESSARY VS GRH-BOUND KO NOT NECESSARY (B) PIE CHART FOR PROPORTION OF GRH-KO REGIONS IN EACH CATEGORY (C) ROLE OF VARIATION IN GRH BINDING SITES IN ACCESSIBILITY DIVERGENCE (D) EXTENT AND (E) PERCENT CIS BETWEEN GRH-KO AND NON-KO.

*The relationship between gene expression and chromatin accessibility divergence based on chromatin remodeling mechanism.*

We next sought to contrast the relationship between gene expression and chromatin accessibility divergence between regions remodeled by a pioneer factor or not. Previous work has shown weak to modest correlations between the two mechanistic layers, although for some cases the correlation seems to be non-overlapping with motif variation. Based on our current molecular model for enhancer remodeling, changes to pioneer factor binding could disrupt the binding of all subsequent activators by altering the overall accessibility, whereas alterations to binding for other enhancers may not be as large. Are changes in accessibility for pioneer factor regions associated with greater differences in gene expression?

FIGURES NEEDED: (A) SCHEMATIC FOR LINKING REGIONS (B) SCATTERPLOT OF ACCESSIBLITY VS EXPRESSOIN (C) EXPRESSION DIVERGENCE AND (D) PERCENT CIS FOR ACCESSIBILITY OVER ALL AND (E) GRH KO NON KO.