METHODS FOR AS-ATAC-SEQ PAPER

*Sample collection*

Imaginal wing discs were dissected in 1xPBS on ice at 10 discs per sample. On each day a round of replicates for each genotype was done, and the order of genotypes was randomized across days to equalize degradation. Following dissection, ATAC-seq was performed as described in CITE? 3.5uL of Illumina Tn5 was used per sample. For RNA-seq library preparation, RNA was extracted using carbon bead chemistry provided by Life Magnetics (CITE) followed by libraries construction using Illumina Tru-seq kit (CITE). ATAC- and RNA-seq libraries were multiplexed and sequenced on a shared Novaseq S2 lane (150 paired end).

*Alignment and filtering*

ATAC- and RNA-seq were trimmed and [ ]. ATAC- and RNA-seq reads were aligned to a concatenated genome using Bowtie2 with the following parameters: [ ] Alignment files were then filtered with Samtools for high-quality, allele-specific reads with the following parameters: [ ]. The ATAC- and RNA-seq alignments were converted from ZHR or Z30 genome coordinates to dm6 coordinates using previously constructed liftOver files (CITE).

*Identifying regions for analyses*

ATAC-seq regions were identified in the following way: 1) MACS2 on merged alignments for each sample and then intersected for ZHR Z30, 2) the same process was done for ZHR TSIM and regions that didn’t overlap with ZHR Z30 were added to the list. Reads were counted in each region for the ATAC-seq samples and in each constitutive exon and then summed by gene for RNA-seq samples. Samples for both datatypes were normalized by CPM transformation.

*Estimating accessibility and expression divergence metric for parental strains and hybrid alleles*

[ WRITE WHAT CAN AND THEN MARK ]

*Annotating regions*

Promoter architectures were assigned to regions by downloading the X dataset, which includes promoters annotated as 1) Peaked, 2) Broad, or 3) Unknown. These annotations were intersected with ATAC-seq regions from this dataset and, after removing Unknowns, were used to contrast Peaked vs Narrow regions.

To annotate regions as Pioneer factor-mediated, we downloaded datasets from Jacobs et al. 2018, in which chromatin accessibility was measured with ATAC-seq in wild-type control and grh mutant imaginal eye-antennal discs. We aligned and called peaks with macs2 and then used Diffbind to compute differentially accessible regions between the two genotypes, to identify regions for which Grh activity is necessary to establish accessibility. These regions were then intersected with the ATAC-seq dataset to identify regions that are pioneer-factor dependent.

*ATAC- and RNA-seq dataset integration and statistical analysis*