Enhancer sharing promotes neighborhoods of transcriptional regulation across eukaryotes

Porfirio Quintero-Cadena and Paul W. Sternberg

Caltech

hhmi | Howard Hughes Medical Institute

Enhancers physically interact with transcriptional promoters, looping over distances that can span multiple regulatory elements.

Given that enhancer-promoter (EP) interactions generally occur via common protein complexes [1], it is unclear whether EP pairing is predominantly deterministic (Figure 1a) or proximity guided (Figure 1b).

To explore the validity of each of these models, we paired every gene of five organisms with its 100 nearest neighbors, and computed the correlation between paired genes across several RNA-seq datasets [2-5]. We found that

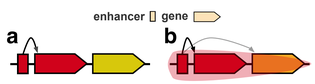

Figure 1. Enhancers have a target gene (a) or a range of action (b).

gene neighbors are transcriptionally correlated genome-wide

in all analyzed organisms (figure 2), and that this correlation decays exponentially over distances that scale with genome size (figure 3).

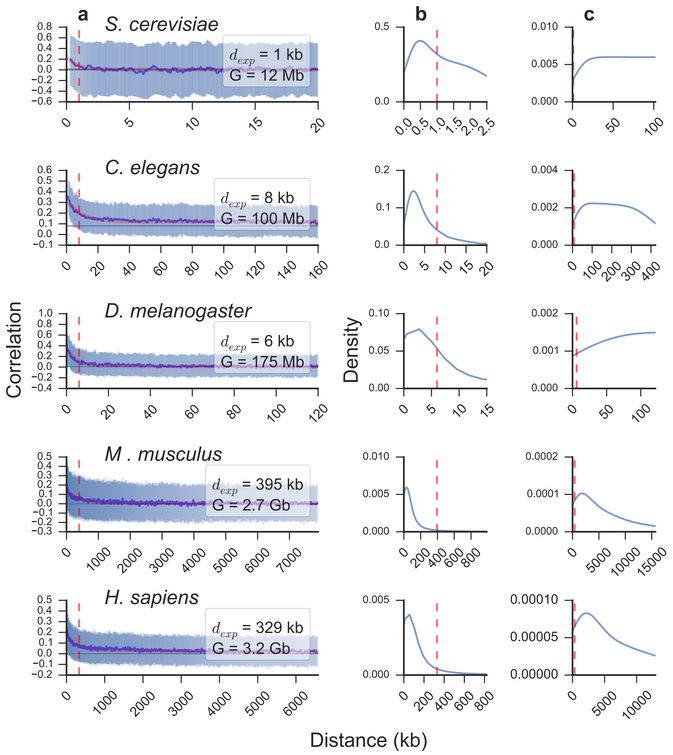

Figure 2. a) Sliding median of correlations between paired neighbors (blue line) and interquartile range (pale blue) with increasing intergenic distance; median of randomly paired genes is shown as a horizontal gray line. Fit to an exponential decay function (red line) was used to compute the mean distance at which gene neighbors remain correlated (d_{exp}, vertical red dashed line). Genome size (G) is shown. Distribution of intergenic distances between each gene and its nearest neighbor (b) and all paired genes (c).

We also computed the percentage overlap in tissue expression in *D. melanogaster*, using RNA *in situ* hybridization data for 6053 genes [6, 7].

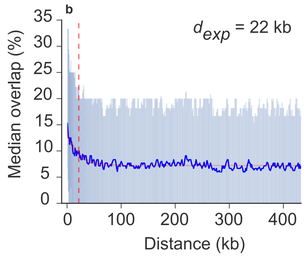

Figure 4. Sliding median of the percentage overlap in tissue specific expression (blue line) and interquartile range (pale blue) with increasing intergenic distance in *D. melanogaster*.

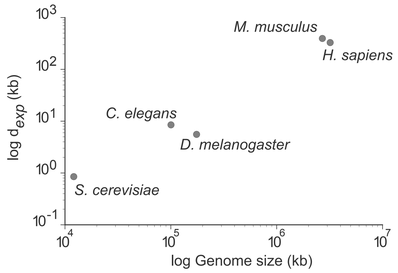

Figure 3. The distance at which a pair of genes remain correlated (d_{exp}) scales with genome size.

This analysis revealed that close neighbors tend to be expressed in the same tissues (Figure 4); gene neighbors thus have a spatio-temporal correlation in expression that is highly dependent upon the spacing between them.

We hypothesized that **enhancer sharing explains the transcriptional correlation of gene neighbors,** and reasoned that transcription of a given gene should also decay exponentially with increasing EP distance.

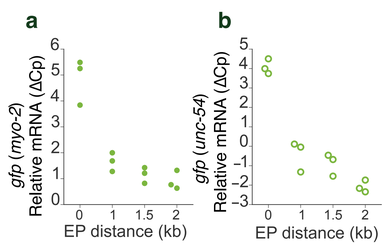

Figure 4. mRNA levels of *gfp* with increasing EP distance for lines with *myo-2* (filled circles, a) and *unc-54* (hollow circles, b) enhancers.

To test this hypothesis, we built a series of genetic constructs with increasing neutral EP distances (0, 1, 1.5 and 2 kb) for two different enhancers, *myo-2* and *unc-54*. We then integrated each construct in single copy into the genome of *C. elegans* and used qPCR to i) measure the influence of EP distance in native chromatin and ii) analyze the impact of the perturbation on the site of transgene insertion.

We found that transcriptional levels of the reporter gene indeed fall rapidly with increasing EP distance (Figure 4).

Expression of the flanking genes *dpy-13* and *col-34* (Figure 5a) was reduced with the introduction of the 2 kb spacer when compared to transgenic lines without it (Figure 5b, c). Spacer-free lines were comparable to wild-type, suggesting the incorporation of ectopic enhancers compensated for the EP distance increase caused by the addition of the genetic construct itself. These observations fit the corollaries of an enhancer-sharing model even amid the complexity of a native regulatory environment.

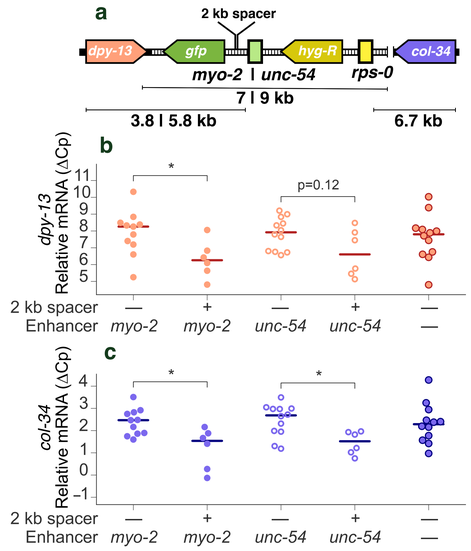

Figure 5. a) Genomic context of the integration site. The inserted construct is shown over a dashed black line. Relative mRNA levels of *dpy-13* (b) and *col-34* (c) in wild-type and lines with and without the 2 kb spacer.

To test whether **EP distance insulates neighboring genes genome-wide,** we looked at the correlation distributions of gene pairs with different orientations, and at those of gene pairs flanked by insulator binding sites determined by ChIP-chip in *D. melanogaster* [8]. We found that each group has distinct distributions of intergenic distances (Figure 6a, d) and distinct transcriptional correlation distributions (Figure 6b, e).

Yet after controlling for the effect of intergenic distance, we found very similar correlation distributions between gene pairs with different orientations (Figure 6c) and insulator and non-insulator flanked gene pairs (Figure 6f), which suggests EP distance is the general source of transcriptional independence for close gene neighbors.

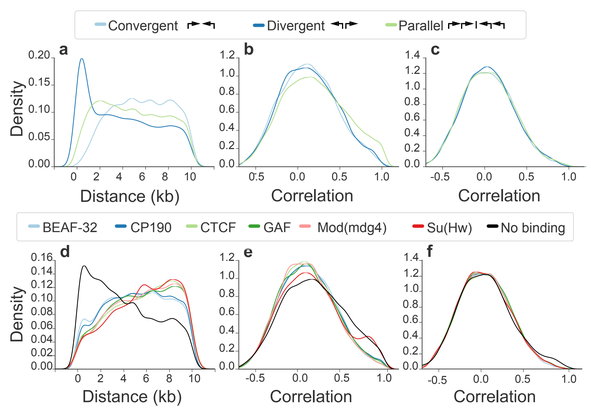

Figure 6. Distribution of intergenic distances below 10 kb of gene pairs in *D. melanogaster* by configuration (a) and flanking insulator binding sites (d). The corresponding distribution of correlations is shown for the same gene pairs (b, e) and pairs with controlled distributions of intergenic distances (c, f).

References • 1. Malik and Roeder (2010) *Nat. Rev. Genet.* • 2. FlyBase consortium (2016) *Nucleic Acids Res.* • 3. Ellahi *et al.* (2015) *Genetics* • 4. Gerstein *et al.* (2010) *Science* • 5. The ENCODE Project Consortium (2012) *Nature* • 6. Hammonds *et al.* (2013) *Genome Biol.* • 7. Tomancak *et al.* (2007) *Genome Biol.* • 8. Negre *et al* (2010) *Plos Genet.*