Enhancer networks – Hidden layer of gene regulation

Justin Malin, Mohamed Radhouane Aniba, Sridhar Hannenhalli

Center for Bioinformatics and Computational Biology
Computational Biology, Bioinformatics, and Genomics Program
Department of Cell Biology and Molecular Genetics
University of Maryland, College Park, MD

Background

Eukaryotic transcription is intricately regulated at multiple levels, including epigenomic modifications, chromatin reorganization, and sequence-specific binding of TF to either proximal promoter regions or to distal enhancer/repressor regions of a gene [1]. Distal enhancers, which can regulate their target genes from long distances -- the most extreme case being the Shh gene's enhancer at ~1Mb away -- are especially important in regulating critical developmental and tissue-specific genes[2]. Recent advances in sequencing technologies have revealed putative distal enhancers based on various epigenomic marks, notably P300 binding [3]. Functionally linked genes tend to be co-expressed and are presumed to be co-regulated [4]. Gene networks based on co-expression patterns of gene pairs across multiple conditions and/or cell types reveal intricate organization of genes into pathways and functional groups [5]. Similar to functionally related genes, functionally related enhancers, i.e., those regulating functionally related genes, share TF binding sites and are likely to have spatio-temporal coordinated activity [6]. A network-level analysis of coordinated activities of distal enhancers has not been reported and such an analysis is likely to reveal higher order organization of a global transcriptional regulatory network mediated by distal enhancers. Analogous to using expression level to quantify transcriptional activity of a gene, DHS of an enhancer region has been proposed as a proxy for its condition-specific regulatory activity [7]. The ENCODE project has produced whole genome DHS profiles across numerous human cell types [8]. Analogous to using cross-condition expression correlation to infer gene networks, cross-condition DHS correlation can be used to infer enhancer networks. Using ~100K P300-bound regions as candidate enhancers, we have identified their correlated activity based on their DHS profiles across 72 human cell types, and followed with investigations of mechanisms and functional consequences of the correlated enhancer activity.

Methods Highlight

1. P300 bound regions in 4 cell types – HepG2, GM12878, H1-HESC and SK-N-SH_RA – were used as candidate enhancer regions, yielding 98,353 enhancers with average length of 500 bps;
2. We obtained DHS status (open or closed) for 72 tissue types in ENCODE yielding a 98,353 x 72 binary matrix. In order to minimize dependencies, tissues were clustered into 37 clusters yielding a 98,353 x 37 binary matrix.
3. Correlation between the activity of two enhancers was quantified using Mutual Information.
4. We controlled for cell type-specific DHS autocorrelation to detect significantly correlated enhancer pairs (Figure 1).

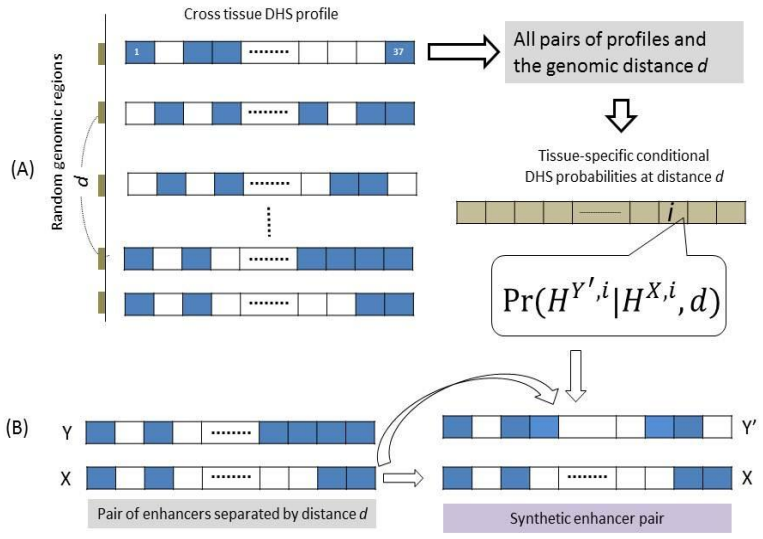

Figure 1. Generating the synthetic enhancer data to account for autocorrelation. (A) Starting with a large set of random genomic regions and their DHS profiles across 37 cell types, we estimate, for each cell type separately, the conditional probability of observing DHS at a location *Y'* given the DHS status at another location *X* at distance *d* from *X*. **(B)** Given a pair of enhancer DHS profiles (*X*,*Y*), we generate a synthetic pair of DHS profiles as (*X,Y'*) where *Y'* is randomly generated from *X* and the conditional probabilities estimated in (A). Blue: DHS=1 (open chromatin); white: DHS = 0 (closed chromatin)

Results Highlight

1. We exhaustively assessed ~35 million intra-chromosomal enhancer pairs separated by less than 12.5 Mb. Despite distance bin-specific FDR control, the fraction of enhancers that are significantly correlated declines with increasing distance (Figure 2). Across all bins, at an FDR of 1% we detect a total of 313,757 significant enhancer pairs, covering 32% of enhancers.

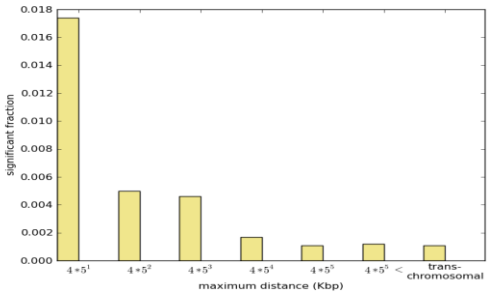

Figure 2. Chromatin states of a large number of enhancer pairs are significantly correlated. The plot shows the fraction of pairs with significant mutual information (*MI*) as a function of inter-enhancer distance. The plot is based on significant pairs after greedily removing pairs inducing transitive relationships.

2. Strong enhancers, those with higher expression levels of the nearest gene, tend to be correlated with fewer enhancers than weak enhancers but preferentially correlate with other strong enhancers, while weak enhancers are correlated with a greater number of enhancers and preferentially correlate with other weak enhancers.
3. Correlated enhancers tend to share common TF binding motifs. We identified 52 TF motifs significantly co-occurring in correlated enhancer pairs relative to uncorrelated enhancers. Using presence of shared motifs as features, correlated enhancers can be distinguished from uncorrelated ones with 73% accuracy. Several chromatin modification enzymes preferentially interact with these 52 TFs.
4. Using the gene closest to an enhancer as putative target, we found that the targets of correlated enhancers have correlated expression and are involved in common biological processes.

5. Correlated enhancers tend to be spatially close (although not highly significantly so) based on Hi-C data.
6. We constructed enhancer networks based on correlated activity and shared TF motifs, and found significant enrichment of specific biological processes among the putative gene targets of the enhancer modules (Figure 3).

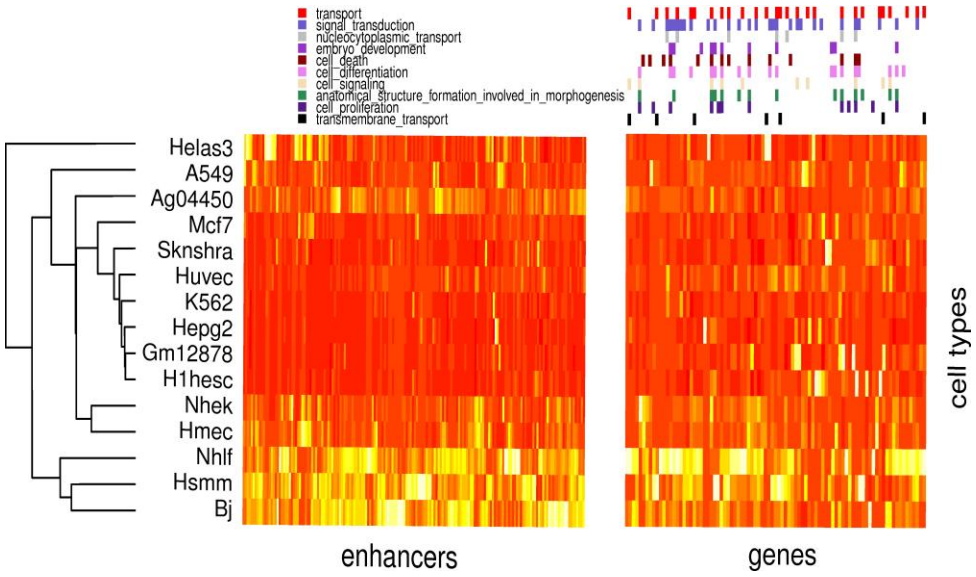

Figure 3: Tissue activity profile of an enhancer cluster and the corresponding target genes. Left: The tissue-specific DHS activity for 179 coordinately activated enhancers across 15 cell types. **Right:** Corresponding expression of the 98 target genes. The GO memberships for enriched terms each gene are shown above the heat plots.

Conclusions

Overall, our analysis suggests that functionally linked genes may be co-regulated by distal enhancers whose activities are regulated by common sets of TFs and mediated by both 3D chromatin structure as well as chromatin modification enzymes.

References

1. White, R.J., *Transcription by RNA polymerase III: more complex than we thought.* Nat Rev Genet, 2011. **12**(7): p. 459-63.
2. Lettice, L.A., et al., *A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.* Hum Mol Genet, 2003. **12**(14): p. 1725-35.
3. Visel, A., et al., *ChIP-seq accurately predicts tissue-specific activity of enhancers.* Nature, 2009. **457**(7231): p. 854-8.
4. Stuart, J.M., et al., *A gene-coexpression network for global discovery of conserved genetic modules.* Science, 2003. **302**(5643): p. 249-55.
5. Dewey, F.E., et al., *Gene coexpression network topology of cardiac development, hypertrophy, and failure.* Circ Cardiovasc Genet, 2011. **4**(1): p. 26-35.
6. Narlikar, L., et al., *Genome-wide discovery of human heart enhancers.* Genome Res, 2010. **20**(3): p. 381-92.
7. Pique-Regi, R., et al., *Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data.* Genome Res, 2011. **21**(3): p. 447-55.
8. Dunham, I., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.