

# 1 Nonlinearity in TF networks

A common model is to use the Hill equation: if  $j$  upregulates  $i$  then

$$\dot{x}_i(t) = \beta \frac{x_j(t)^n}{K_A^n + x_j(t)^n} \quad (1)$$

or if it downregulates,

$$\dot{x}_i(t) = \beta(1 - \frac{x_j(t)^n}{K_A^n + x_j(t)^n}). \quad (2)$$

Citation: Alon’s book? gertz2008analysis?

# 2 Notes on Empirical *Cis*-Regulatory Module Variation and Evolution

Generally, there are three sorts of data on variation:

1. how binding site occupancy varies in a population
2. how promoter SNPs affect expression (or phenotype) by swapping them out
3. how cis-eQTL affect expression by population scans, esp in heterozygotes

*Orphaned, unconvincing paragraph:* In *Drosophila*, [Frankel et al. \[2011\]](#) found that clusters of 2–5 mutations in enhancers reduced expression of what? by 4.6–33.5%, and [Rebeiz et al. \[2009\]](#) found that around five cis-regulatory SNPs cause differences of 8–40% of *ebony* expression differences between light and dark lines; while in human (ENCODE) data, [Shi et al. \[2016\]](#) found that shifts in binding site motif score of 5%–10% are responsible for differential occupancy. These latter observations suggest that changes of several base pairs can shift occupancy and/or expression by tens of percent; if a typical promoter contains 100 bases relevant for binding of a particular transcription factor, in which region 1% of the population contains a SNP, this suggests population expression variance due to such factors of around 0.1–1%.

- [Kasowski et al. \[2010\]](#) use ChIP-seq to survey variation in binding sites between 10 human lymphoblastoid cell lines: RNA PolII (not a TF like we are considering) and NFkB (p65, a regulator of immune responses, which does have binding motifs); individuals differed at 25% and 7.5% of the sites respectively. Also found 32% binding sites having significant difference in PolII occupancy between human/chimp
  - Differences explained by SNPs, structural variants, and other things (like [Shi et al. \[2016\]](#))
  - “Significant differences in binding” have SD of +/- 1 log2 difference
  - Weak connection of binding differences to expression differences: “We examined the effect of binding variation on gene expression by generation of deep RNASeq data from each cell line (9) and comparison with binding data (Fig. 3A, Fig. S9A). A significant correlation was observed (Spearman correlation coefficients of 0.475 and 0.461 for NFB and PolII, respectively) (Fig. 3B, Fig. S9B, Table S9), suggesting an influence of binding differences on mRNA abundance”
  - “Since humans and chimpanzees exhibit 510% differences in gene expression (15)”
- [Schmidt et al. \[2010\]](#) use ChIP-seq to compare TF occupancy (for CEBPA and HNF4A) between five vertebrates (human, mouse, dog, opossum, and chicken), and find (a) conservation of binding motif; (b) little sharing of actual binding sites; and (c) a fair amount of compensatory changes

- "Among the binding events lost in one lineage, only half are recovered by another binding event within 10 kilobases"
- "CEBPA and HNF4A were selected as representative transcription factors within the liver-specific regulatory network because both are conserved and constitutively expressed with well-characterized target genes (10, 11). In addition, they represent distinct TF classes, and the DNA binding domains of each factor's orthologs are nearly identical among the study species "
- "About 250 direct functional HNF4A target genes have recently been identified"
- both have consensus motifs of about 5 of 10 bp
- "only between 6-8 % of the genomic regions occupied by CEBPA in opossum liver align with CEBPA binding events also found in mouse, dog, and/or human liver. This divergence was even greater in chicken, which shared only 2% of CEBPA binding with human"
- Shi et al. [2016] looked at variation in TF binding in heterozygotes in ENCODE data and found that differential binding is often explained by SNPs in the TF's motif (19.3%) or potential partner TFs' motifs (9.4%); variation in chromatin accessibility; or histone modifications (NOT just SNPs).
  - shifts of motif score of 0.05–0.1 were responsible for differential occupancy
  - "While altered canonical TFBSs for the ChIP'd TFs were frequently observed (19.3%), most ASB SNVs did not overlap with the primary TF motif."
- Stefflova et al. [2013] In five Mus (sub-)species, found that "Although individual mutations in bound sequence motifs can influence TF binding, most binding differences occur in the absence of nearby sequence variations."
  - "Clustered TF binding appears to result in large part from indirect cooperativity to open chromatin regions, as opposed to direct TF-TF protein interactions (Kaplan et al., 2011; Miller and Widom, 2003; Mirny, 2010). For binding sites within a nucleosome-length distance, each TF contributes partially to a competitive displacement of specific nucleosomes by indirect collaboration with other TFs, mutually aiding each others binding to DNA. TFs within a cluster can have different regulatory roles depending on their motif strength and ability to compete with nucleosomes (Zinzen et al., 2009)."
- Tuğrul et al. [2015] do theory on the gain and loss of TFBS. Conclusion: more than 10bp unlikely; de novo very hard but made possible at the scale of eukaryotic species divergence by pre-existing relics, cooperativity, etc. Euk TFBS typically 6–12bp.
- Gertz et al. [2009] study synthetic yeast promoters. The synthetic promoters contained a random combination of 3-4 TFBSs. Weak MIG1 TFBS has a 6.7-9.0 fold lower affinity than the strong TFBS. Weak sites in concert with a strong site act as a strong site too.
- Frankel et al. [2011] study the loss of trichome phenotype in *D. sechellia* (vs. *melanogaster*). Identify 6 embryonic enhancers (5 kb) that are diverged. Examine in detail the E enhancer – two subunits of it, each 1 kb long, E3 and E6. Say that E3 is conserved, while the E6 region has evolved. Comparing the E6 *sechellia* sequence with five closely related species, all containing a high density of trichomes (in contrast to *sechellia*), they identify 13 substitutions and 1 single nucleotide deletion. There was no polymorphism within *D. sechellia* within the E6 region (9 sequences). Estimates these substitutions occurred at an accelerated rate – 4.8 times the flanking regions.

Introduced all *sechellia* substitutions into *melanogaster* (and the reverse!). Found that expression changed in the directions as expected in the transgenic lines – but did not completely rescue phenotypes – there are likely other regulatory elements involved.

The temporal expression delay in *sechellia* could not be captured by a single mutation – only a combination of the known differences. At least 5 substitutions are necessary. Each cluster of mutations (2-5) reduced trichome expression by 4.6-33.5%.

“none of the *D. melanogaster* to *D. sechellia* mutations led to a significant increase in trichome number, and none of the reciprocal mutations led to a decrease in trichome number”

- [Rebeiz et al. \[2009\]](#) look at adaptive melaninism in *D. melanogaster*. Find a minimum of 5 mutations differentiate regulation of dark from light lines. These account for 8 to 40% of the overall difference in ebony activity. (*But, this is adaptive – different selection in different places.*)

Lots of standing variation preceding phenotype divergence: 3 of the identified causative substitutions in one region were found at high frequencies in both Ugandan light lines and in a Kenyan light population. Two of the substitutions were found in all 5 populations sampled. The third subst was found in 4 of the 5 populations. Dark specific substitutions were not observed in other lines; suggesting they were rare variants or evolved de novo.

Estimate that standing variation in the enhancer happened first, then the fixation of the de novo mutations.

“We have shown that the adaptive evolution of melanism in a Ugandan population of *D. melanogaster* occurred through multiple, stepwise substitutions in one enhancer of the ebony locus. We suggest that this genetic path of enhancer evolution with multiple substitutions of varying effect sizes, which originate from both standing variation and new mutations and combine to create an allele of large effect, may be a general feature of enhancer evolution in populations. This view is consistent with studies that have demonstrated that substitutions at multiple sites within enhancers are responsible for evolutionary changes in gene expression”

- [Verlaan et al. \[2009\]](#) study expression variation within humans. cis-regulatory variation is known to account for 2-6% expression variation in human blood-derived primary cells.
- [Yan et al. \[2002\]](#) studied CEPH families (human). Fraction of individual variation in heterozygous individuals was 3 to 30% (differentially expressed alleles). Expressed transcript ratio varied from 1.3 : 1.0 (FBN1) to 4.3 : 1.0 (p73). Many of these expression differences were Mendelian.
- [\[Lappalainen et al., 2013\]](#) mapped cis-QTLs to transcriptome variation in human samples from CEPH, FIN, GBR, TSI and the 1000 genome project. Population differences explain 3% of the variation. Identified 263-4,379 genes with differential expression (or transcript ratios) between population pairs.

See figure 1b.

## References

- Nicolás Frankel, Deniz F Erezylmaz, Alistair P McGregor, Shu Wang, François Payre, and David L Stern. Morphological evolution caused by many subtle-effect substitutions in regulatory dna. *Nature*, 474(7353): 598–603, 2011. [1](#), [2](#)
- Jason Gertz, Eric D. Siggia, and Barak A. Cohen. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, 457(7226):215–218, January 2009. ISSN 00280836. URL <http://dx.doi.org/10.1038/nature07521>. [2](#)
- M. Kasowski, F. Grubert, C. Heffelfinger, M. Hariharan, A. Asabere, S. M. Waszak, L. Habegger, J. Rozowsky, M. Shi, A. E. Urban, M. Y. Hong, K. J. Karczewski, W. Huber, S. M. Weissman, M. B. Gerstein, J. O. Korbel, and M. Snyder. Variation in transcription factor binding among humans. *Science*, 328(5975): 232–235, April 2010. [1](#)
- Tuuli Lappalainen, Michael Sammeth, Marc R Friedländer, Peter ACt Hoen, Jean Monlong, Manuel A Rivas, Mar Gonzalez-Porta, Natalja Kurbatova, Thasso Griebel, Pedro G Ferreira, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468):506–511, 2013. [3](#)

- Mark Rebeiz, John E Pool, Victoria A Kassner, Charles F Aquadro, and Sean B Carroll. Stepwise modification of a modular enhancer underlies adaptation in a drosophila population. *Science*, 326(5960):1663–1667, 2009. [1](#), [3](#)
- D Schmidt, M D Wilson, B Ballester, P C Schwalie, G D Brown, A Marshall, C Kutter, S Watt, C P Martinez-Jimenez, S Mackay, I Talianidis, P Flicek, and D T Odom. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 328(5981):1036–1040, May 2010. doi: 10.1126/science.1186176. URL <https://www.ncbi.nlm.nih.gov/pubmed/20378774>. [1](#)
- W. Shi, O. Fornes, A. Mathelier, and W. W. Wasserman. Evaluating the impact of single nucleotide variants on transcription factor binding. *Nucleic Acids Res.*, 44(21):10106–10116, December 2016. [1](#), [2](#)
- K. Stefflova, D. Thybert, M. D. Wilson, I. Streeter, J. Aleksic, P. Karagianni, A. Brazma, D. J. Adams, I. Talianidis, J. C. Marioni, P. Flicek, and D. T. Odom. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell*, 154(3):530–540, August 2013. [2](#)
- Murat Tuğrul, Tiago Paixão, Nicholas H. Barton, and Gašper Tkačik. Dynamics of transcription factor binding site evolution. *PLOS Genetics*, 11(11):1–28, 11 2015. doi: 10.1371/journal.pgen.1005639. URL <https://doi.org/10.1371/journal.pgen.1005639>. [2](#)
- Dominique J Verlaan, Bing Ge, Elin Grundberg, Rose Hoberman, Kevin CL Lam, Vonda Koka, Joana Dias, Scott Gurd, Nicolas W Martin, Hans Mallmin, et al. Targeted screening of cis-regulatory variation in human haplotypes. *Genome research*, 19(1):118–127, 2009. [3](#)
- Hai Yan, Weishi Yuan, Victor E Velculescu, Bert Vogelstein, and Kenneth W Kinzler. Allelic variation in human gene expression. *Science*, 297(5584):1143–1143, 2002. [3](#)