

## Decoding Noncoding Regulatory DNAs in Metazoan Genomes

Michele Markstein, Angela Stathopoulos, Vicky Markstein, Peter Markstein, Naoe Harafuji, Dave Keys, Byung-in Lee, Paul Richardson, Dan Rokshar, and Mike Levine.  
*Dept. Mol. Cell. Biol., UC Berkeley; Joint Genome Institute, Walnut Creek, CA.*

### Abstract

The recent revelation that the human genome contains only ~30,000 genes underscores the importance of gene regulation in generating organismal diversity. Cis-regulatory DNAs, or enhancers, are short stretches of DNA--300 bp to 1,000 bp in length--that control gene expression. This DNA accounts for a substantial fraction of metazoan genomes, but is largely invisible. It cannot be identified by simple sequence inspection. One of the outstanding issues in the post-genome era is whether there is a "cis-regulatory code" that links primary DNA sequence with gene expression patterns. We have used a combination of bioinformatics methods and functional assays to determine whether coordinately regulated genes share a common "grammar".

The early development of the *Drosophila* embryo is controlled by a regulatory protein called Dorsal. Classical genetic studies identified ~20 target genes that are directly or indirectly regulated by Dorsal. More recent DNA microarray screens have identified another 30-35 target genes. Roughly half, ~25 of the 50-55 target genes, are likely to be directly regulated by Dorsal. Traditional, rather tedious, functional assays

identified regulatory DNAs for 7 of these 25 direct target genes. Whole-genome searches for clusters of Dorsal binding sites led to the rapid identification of at least another 4 authentic enhancers. At least 30% of the optimal Dorsal binding clusters correspond to cis-regulatory DNAs when tested in vivo. Our next major goal is to search the 11 cis-regulatory DNAs for shared sequence features. The combination of Dorsal binding sites and additional conserved elements might identify a common "code" for Dorsal target genes.

We also study regulatory DNAs in a simple chordate, *Ciona intestinalis*. While the adults are primitive filter feeders, the embryos and larvae look like vertebrate tadpoles. A strength of the system is that it is possible to perform high throughput analysis of regulatory DNAs. A pilot screen of random genomic DNAs suggests that the *Ciona* genome contains one tissue-specific enhancer every 15,000 base pairs. Thus, the *Ciona* genome contains something like 10,000 tissue-specific enhancers, and we hope to identify 5-10% of these enhancers during the upcoming year. This large collection of enhancers will be used to search for common features among coordinately regulated genes.