Principles of Systems Biology, No. 4

What controls gene expression? Perhaps more than we thought reveal the first three responses to this month's Cell Systems Call (Cell Systems *1*, 307). Plus, CRISPR/Cas9 targets RNA.

Prevalent Variation in Transcription Factor DNA Binding Activity

Martha L. Bulyk, Brigham & Women's Hospital and Harvard Medical School

Principles

Recently, we developed a computational pipeline to survey human transcription factor (TF) DNA-binding domains for known Mendelian disease mutations and coding variation found in recent genotyping and exome sequencing of >64,000 individuals from diverse ancestries (Barrera et al., Science *351*, 1450–1454). We identified thousands of variants predicted to alter DNA binding activity. Seventy-seven of 117 tested variants affected DNA binding affinity and/or specificity in protein-binding microarray (PBM) assays. Mutants' altered PBM profiles were consistent with altered occupancies of genomic target sites and dysregulation of the associated target genes. Intriguingly, while individual TF alleles predicted to damage DNA binding activity are rare, in aggregate they are prevalent: our results suggest that most unrelated individuals harbor a unique repertoire of TF alleles with a distinct *trans*-regulatory collective of DNA-binding activities.

> *"…while individual TF alleles predicted to damage DNA binding activity are rare, in aggregate they are prevalent…."*

What's Next?

Many more TF variants likely will be found in rapidly growing exome and whole-genome sequencing studies. What are the consequences of damaging TF DNA binding domain variants on cellular and organismal phenotypes? While the effects of some variants may be buffered genetically by redundancy or epistasis in transcriptional regulatory networks, others may lead to phenotypes that are subclinical, subject to genetic or environmental interactions, or present later in life. Future studies will determine how TF DNA binding domain variants contribute to human phenotypic diversity including risk for various diseases.

From Co-regulated Gene Clusters to Gene-Specific Mechanisms

Ann-Jay Tong and Stephen T. Smale, University of California, Los Angeles

Principles

Systems analyses of transcriptional cascades and networks are usually designed to maximize statistical power by using low-stringency criteria to capture the largest number of differentially expressed genes possible. This approach can reveal molecular features that are statistically enriched among clusters of co-regulated genes, but its ability to uncover precise molecular mechanisms is limited. By using more stringent criteria for the analysis of genome-scale data, we uncovered several mechanistic principles of a macrophage's response to a stimulus that would have been missed in conventional systems analyses (Tong et al., Cell *165*, 165–179). Most importantly, we were surprised to find evidence that key immunoregulatory genes are activated by unique molecular mechanisms, in which common transcription factors employ mechanisms that are tailored to an individual target gene and are not used at any other inducible gene.

> *"…transcription factors employ mechanisms that are tailored to an individual target gene…."*

What's Next?

Our results suggest that the use of more stringent criteria to analyze genomics data will benefit studies of other transcriptional cascades and networks, as a systems-level understanding cannot be gained solely from statistical trends observed at co-regulated clusters. However, it will also be necessary to overcome several remaining technical challenges, including the challenge of distinguishing functional and non-functional protein-DNA interactions identified in ChIP-seq experiments.

A New Histone Mark for Enhancers

Vibhor Kumar and Shyam Prabhakar, Computational and Systems Biology, Genome

Principles

The histone proteins that package eukaryotic DNA can be modified by acetylation at ∼35 different locations. Why does the cell tweak chromatin in so many different ways? Do the acetylation marks have distinct functions? These questions are not easily answered since most existing studies examine only two acetylations: H3K9ac and H3K27ac. In particular, H3K27ac is used as a general marker of active enhancers. However, based on enhancer assays, omics profiling, and analysis of public datasets, we found that another acetylation mark, H2BK20ac, was actually the best enhancer predictor (Kumar et al., Genome Res., published online March 8, 2016. http://dx.doi.org/10.1101/gr.201038.115). Moreover, H2BK20ac had the strongest association with cell-type-specific promoters and cell-type-specific biological functions. Finally, H2BK20ac also marks dynamic regulatory elements that respond to external stimulation. Overall, our results suggest that the functional diversity of histone acetylation marks has been underappreciated.

> *"…we found that another acetylation mark, H2BK20ac, was actually the best enhancer predictor."*

What's Next?

We propose that H2BK20ac be added to the set of marks constituting a minimal reference genome, so as to uncover additional enhancers, cell-type-specific promoters, and dynamic regulatory elements. Our findings also raise fundamental mechanistic questions. Which are the enzymes and transcription factors that decouple H2BK20ac from other acetylation marks? Is H2BK20ac a cause or a consequence of cell-type and cell-state specificity? Are there chromatin "readers" that respond specifically to H2BK20ac?