Integration of high-throughput DNA binding and gene expression data to understand effector B cell fate choice

Mark Maienschein-Cline[1], Kyoko Ochiai[2], Roger Sciammas[2], Harinder Singh[2], and Aaron R. Dinner[1]

Balancing somatic hypermutation and class switch recombination with the differentiation of activated B cells into antibody secreting plasma cells is important for the humoral immune response. These processes require the transcription factor IRF-4, which we examine along with transcription factors IRF-8 and PU.1. We use a combination of structural protein-DNA binding and functional gene expression data, obtained from ChIP-seq and DNA microarray experiments, to examine the regulatory modes of these transcription factors on a genome-wide scale. This allows us both to determine the effect of a protein-DNA binding event on a per-site basis, and to assess the overall effect of the transcription factors over the progression of effector B cell fate choice.

***Keywords* — High-throughput data analysis, ChIP-seq, DNA microarray, effector B Cell fate choice**

I. Outline

The purpose of this study is to obtain a better understanding of how the transcription factors (TFs) IRF-4, IRF-8, and PU.1 interact to affect B cell fate choice. It has previously been shown that IRF-4 is obligate for class switch recombination and plasma cell differentiation [1,2] whereas IRF-8 may play an antagonistic role. Additionally, IRF-4 / PU.1 and IRF-8 / PU.1 are known to bind cooperatively to DNA [3]. The goal then is to understand how these factors function in isolation combinatorially.

Independently, ChIP-seq and DNA microarray data tell very different stories. A ChIP-seq experiment measures protein-DNA binding sites (as chromosomal locations) for a given protein. This can be used to hypothesize gene regulation or interaction between TFs by their locations in the genome, but the data is ultimately structural and not functional. On the other hand, DNA microarray experiments are functional but not structural: they allow us to observe changes in the expression levels of thousands of genes over a variety of conditions. While interactions between genes can be inferred from expression profiles, a direct mapping of regulation to activity is absent. A combination of the two experimental methods allows the assignment of specific regulatory roles to TFs.

II. Specific Goals

Our analysis incorporates data obtained from several conditions: expression data over three time points in the wild type, IRF-4, and IRF-8 knockouts, and binding data for IRF-4, IRF-8, and PU.1 on the second two time points.

Starting with the expression data, we can then determine for each gene (1) how its expression changes over the course of B cell fate choice and (2) how a lack of IRF-4 or IRF-8 affects that gene. We cluster the data to define classes of expression across each conditional dimension, arriving at a summary of several classifications for each gene that describe its behavior.

We then map each putative binding site obtained from ChIP-seq to genes in the expression data. We can distinguish which mapped gene a particular binding site is regulating, and the aggregate effect of the set of binding sites on the regulatory activity (in both cases using the classification summaries to define the regulatory mode). By picking binding site sets appropriately, we can gauge the effect of combinations of TFs on regulation; for example, we can choose all IRF-4 and PU.1 sites that lie near one another to measure the combinatorial effect of these TFs.

References

[1] Sciammas R, Shaffer AL, Schatz JH, Zhao H, Staudt LM, Singh H, *Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation,* Immunity, 2: 225 (2006).

[2] Klein F, *et al, Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination,* Nature Immunology, 7:773 (2006).

[3] Escalante CR, Brass AL, Pongubala HM, Shatova E, Shen L, Singh H, Aggarwal AK, *Crystal Structure of PU.1/IRF-4/DNA ternary complex*, Mol. Cell. 10: 1097 (2002)

[1]

[1]Department of Chemistry, University of Chicago;
presenting author email: mmaiensc@uchicago.edu
[2]Deparment of Molecular Genetics and Cell Biology, University of Chicago