# Chapter 37
# Using Systems Biology Approaches to Predict New Players in the Innate Immune System

**Bin Li**
*Merrimack Pharmaceuticals, USA*

## ABSTRACT

*Toll-like receptors (TLRs) are critical players in the innate immune response to pathogens. However, transcriptional regulatory mechanisms in the TLR activation pathways are still relatively poorly characterized. To address this question, the author of this chapter applied a systematic approach to predict transcription factors that temporally regulate differentially expressed genes under diverse TLR stimuli. Time-course microarray data were selected from mouse bone marrow-derived macrophages stimulated by six TLR agonists. Differentially regulated genes were clustered on the basis of their dynamic behavior. The author then developed a computational method to identify positional overlapping transcription factor (TF) binding sites in each cluster, so as to predict possible TFs that may regulate these genes. A second microarray dataset, on wild-type, Myd88$^{-/-}$ and Trif$^{-/-}$ macrophages stimulated by lipopolysaccharide (LPS), was used to provide supporting evidence on this combined approach. Overall, the author was able to identify known TLR TFs, as well as to predict new TFs that may be involved in TLR signaling.*

## INTRODUCTION

The innate immune system provides the first line of defense against microbial pathogens (Ye et al., 2002; Beutler, 2004; Oda & Kitano, 2006). Toll-like receptors (TLRs) are important components of the innate immune system, which recognize foreign invaders and activate pathogen-specific immune response through the fine-tuned regulation of transcription factors (TFs). In recent years, much progress has been made in discovering new components and understanding their interactions within the Myd88-dependent and Trif-dependent TLR signaling pathways, with NFkBs and IRFs

DOI: 10.4018/978-1-4666-3604-0.ch037

as downstream transcription factors, respectively (Beutler, 2004; Oda & Kitano, 2006). On the contrary, only a limited number of transcription factors have thus far been identified to be involved in the TLR signaling network, namely the NFkB, AP1, IRFs, and CREB families.

Recent technological developments in high-throughput experiments, such as microarrays (Draghici, Khatri, Eklund, & Szallasi, 2006), massively parallel signature sequencing (Stolovitzky et al., 2005), and Chromatin immunoprecipitation coupled to microarray hybridization (Buck & Lieb, 2004), enable the collection of data across an entire genome, making it possible to gain knowledge at the systems level. In addition, databases such as TRANSFAC (Matys et al., 2006) and JASPAR (Sandelin, Alkema, Engstrom, Wasserman, & Lenhard, 2004) have collected TF binding site information in the form of positional weight matrices, enabling computational scanning of known TF binding sites (TFBSs). It is possible to combine these genome-wide data to systematically predict novel TFs in biological systems. Indeed, we recently adopted such an approach to predict and validate ATF3 as a regulator of Lipopolysaccharide (LPS) induced innate immune responses (Gilchrist et al., 2006).

The success of this systematic approach is dependent on finding shared patterns of TFBSs among co-regulated genes. One strategy for defining shared TFBS patterns requires that the predicted TFBSs form a spatial cluster on the DNA (Frith, Li, & Weng, 2003). Additional constraints may be added so that predicted TFBSs must occur in the same 5'-3' order on the DNA, or that distances among the predicted TFBSs be conserved. Here, we developed a novel computational method for identifying positional overlapping TF binding sites to search TFBSs that are at similar distances from transcription start sites among putative co-regulated genes.

To better identify putative co-regulated genes, we selected a dataset from a series of microarray experiments with six different stimuli (LPS, PAM2, PAM3, Poly I:C, R848, and CpG), each at six time points (0 min, 20 min, 40 min, 60 min, 80 min and 120 min) (Ramsey et al., 2008). We utilized this fine-grained and relatively short time-course in order to circumvent positive and negative feedback loops and thus to define direct transcriptional targets. Using this set of time-course data, we identified differentially regulated genes for each stimulus and then clustered genes based on their dynamic behavior. We then applied a stringent computational method to predict possible regulators for genes in each cluster. As a proof-of-principle, we were able to recapitulate roles for Nfkb and Irf which are well-known in the TLR signaling system. More interestingly, we identified a novel regulatory role for Egr1 in regulating an early transient gene cluster. Overall, the novel TF predictions and the high-quality microarray dataset represent a useful resource for the research community.

## METHODS

### Microarray Expression Measurements

Femurs from the C57BL/6 (Jackson Laboratories) mice were flushed with complete RPMI (RPMI 1640 supplemented with 10% FBS, 2mM L-glutamine, 100 IU/mL penicillin and 100 mg/mL streptomycin. All of these reagents were from Cellgro, Mediatech, except that the FBS was from Hyclone). Bone marrow cells were plated on non-tissue culture treated plastic in complete RPMI supplemented with recombinant human M-CSF (rhM-CSF) at 50 ng/mL (gift from Chiron). On day 4, the cells were washed two times with RPMI with no additions and then allowed to grow for 2 more days in complete RPMI supplemented with 50 ng/mL of rhM-CSF. On day 6, the cells were lifted from the non-tissue culture treated plastic, counted and plated at a density of $1\times10^5$ cells/cm2 ($1\times10^6$ cells per well in a 6-well dish)