Title
Transcriptional regulation in the immune system: a status report

Journal
Trends in Immunology, 35(5)

ISSN
1471-4906

Author
Smale, Stephen T

Publication Date
2014-05-01

DOI
10.1016/j.it.2014.03.003

Peer reviewed

Published in final edited form as:
Trends Immunol. 2014 May ; 35(5): 190–194. doi:10.1016/j.it.2014.03.003.

Transcriptional Regulation in the Immune System: A Status Report

Stephen T. Smale
Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA 90095

Abstract

Regulated changes in transcription play a central role in virtually all events that accompany the development of the immune system and its response to microbial and environmental cues. Over the past 30 years, a large number of proteins that regulate transcription in the immune system have been discovered and much has been learned about their mechanisms of action. However, the field remains in its infancy, with technical challenges and the complexity of gene regulation circuitry limiting our current knowledge and providing formidable barriers to further advancement. Despite these barriers, the development of new and increasingly sophisticated technologies is speeding progress toward an understanding of the gene-specific and global logic through which transcription is regulated in key immunological settings.

Benchmarks in molecular biology

The immune system has held a prominent place in the eukaryotic gene regulation field since the emergence of the field soon after the molecular biology revolution of the 1970s. The immune system initially was of special interest to molecular biologists because of the high abundance of immunoglobulin (Ig) molecules, which made them relatively easy to isolate and study, thereby allowing Ig heavy chain and light chain genes to be among the first cell type-specific genes isolated. The discovery of VDJ recombination as a mechanism for generating antigen receptor diversity led to the discovery of the RAG1 and RAG2 recombinase proteins and to continuing efforts to elucidate recombination mechanisms. However, Ig genes were also the focus of studies that led to key early advances in our understanding of transcriptional regulation in eukaryotic cells.

One key discovery was the identification of a DNA sequence element, called the octamer motif, that is present in the promoters of most Ig V segment genes in diverse species [1]. The octamer motif discovery provided early support for the hypothesis that eukaryotic promoters contain sequences that contribute to cell type-specific transcription via their recognition by sequence-specific DNA-binding proteins. These studies soon led to the

discovery of the OCT transcription factors [2, 3]. A second major discovery was that mammalian genes contain DNA regions at a distance from their promoters, referred to as enhancers, which contribute to transcriptional activation and cell type-specificity. Indeed, the Ig μ intronic enhancer was the first enhancer reported for a cellular (as opposed to viral) gene [4, 5]. A third early discovery was the first eukaryotic transcription factor whose activity can be rapidly induced by a post-translational mechanism. This factor, NF-κB, was originally identified during an analysis of the Ig κ light-chain enhancer [6] and opened the door to extensive studies of stimulus-induced transcription in the immune system and in other physiological settings.

Since these initial discoveries approximately 30 years ago, our knowledge of transcriptional control of immune development and immune responses has increased dramatically. In fact, it could be argued that our knowledge is now quite sophisticated, with the successful discovery and characterization of dozens of transcription factors that contribute to the development of many immune cell types and their response to a broad range of stimuli. In addition to the discovery of transcription factors that act by binding DNA in a sequence-specific manner, dozens of co-regulatory proteins and chromatin proteins that contribute to transcriptional control have been reported. Furthermore, progress has been made toward an understanding of transcriptional and post-transcriptional mechanisms regulating transcription factor activity, the role of nuclear architecture in transcriptional control, single-cell dynamics of gene regulation, and many contributors to the post-transcriptional control of gene expression.

Despite these advances, another view is that the field remains in its infancy, with most of our progress largely providing a portion of the groundwork needed to ultimately understand both the global and gene-specific logic through which transcription factors, co-regulatory proteins, chromatin structure, nuclear organization, and signaling pathways act in concert to coordinate cell-fate decisions, the development and maintenance of individual immune cell types, and their highly specific responses to diverse stimuli and combinations of stimuli. An understanding of this logic is, in turn, needed to fully appreciate immune development and immune responses in the context of both normal physiology and disease. In this commentary, I present a brief overview of the current state of the field, with an emphasis on the limitations of our current knowledge and technical capabilities, as well as the great promise for the future.

Discovery and characterization of key transcription factors

Without question, the greatest achievement of the immunology field to date with respect to gene regulation has been the discovery and basic characterization of a large number of transcription factors that play critical roles either in the development of specific immune cell lineages or in immune cell activation (Box 1). A brief list, which includes only a small fraction of the factors that have been conclusively shown to contribute to immune cell development and/or an immune response - E2A, Pax5, EBF, PU.1, Ikaros, GATA3, Th-POK, Tbet, Bcl6, NF-κB, STATs, IRFs – highlights this great progress. These and many other transcription factors and their associated genes were discovered by a remarkably diverse range of experimental approaches. However, their critical functions were

documented almost invariably through the generation and characterization of mutant mouse strains, occasionally following initial insights from naturally occurring mutations in humans or other species. Although the names and functions of specific transcriptional regulators would, at one time, have been recognized only by those whose research is devoted to studies of transcriptional control, it is striking to note that many transcription factor names are now recognizable by most immunologists, as the knockout strains and the transcription factor expression patterns are critical for many immunological studies. Knockout phenotypes for additional transcription factors that contribute to immune development and function continue to be reported each year, resulting in continually expanding knowledge of transcription factors that are relevant to our understanding of the immune system.

Box

Gene Regulation in the Immune System: Progress and Barriers

Most Significant Areas of Progress

- **a.** Discovery of dozens of transcription factors and their genes that have been convincingly shown in gene knockout studies to be important for immune development or function.
- **b.** Elucidation of the basic signaling pathways regulating the activity of several key transcription factors.
- **c.** Discovery of many co-regulatory proteins and chromatin proteins required for proper gene regulation in the immune system.
- **d.** Initial progress toward the elucidation of gene regulation networks regulating key immunological events.

Common Barriers to Progress

- **a.** Most studies, by necessity, focus on narrowly defined events.
- **b.** Genes that are directly targeted by transcription factors, co-regulatory proteins, and chromatin proteins remain difficult to distinguish from indirect targets.
- **c.** Chromatin and nuclear properties that cause a change in gene expression of difficult to distinguish from those properties that are a consequence of gene expression changes.
- **d.** Redundancy between factors masks important biological functions.
- **e.** Functionally important protein-protein interactions, which often occur at low affinity, are difficult to distinguish from irrelevant, sticky interactions between proteins.

The successful dissection of the basic mechanisms needed to activate specific transcription factors represents another series of major advances over the past 30 years. For example, the activation of NF-κB dimers by Type 1 and Type 2 signaling pathways and the activation of STAT transcription factors by JAK kinases have been well-documented. Mechanisms that contribute to the activation of several other transcription factors have been described. An

important caveat is that each of these factors is thought to be subject to multiple additional layers of regulation (e.g. direct post-translational modification of NF-κB dimers), with much less known about the full complement of regulatory layers and the logic through which they impact transcription factor activity, and how these may vary in different immune cell types.

Finally, for a small number of developmental transitions, initial but highly significant progress has been made toward an understanding of the networks through which multiple transcription factors, signaling pathways, and target genes orchestrate the transition. One prominent example is the basic network through which multiple transcription factors, including PU.1, E2A, Ikaros, EBF, Pax5, Foxo1, IRF4, and IRF8, act in concert with defined signaling pathways to regulate early stages of B cell development [7]. Although incomplete, the networks that have been defined form a framework on which future studies can build.

The challenges of embracing complexity

The discovery and basic characterization of a large number of transcription factors that play important roles in immune development and immune responses provides the groundwork for an advanced understanding of the logic through which transcription factors act in concert with signaling pathways, co-regulatory proteins, chromatin structure, and nuclear organization to regulate important developmental and immunological events. However, our knowledge remains severely deficient, in large part due to two general challenges faced by all laboratories studying gene regulation in the immune system (Box 1).

One challenge is that most studies of molecular mechanisms focus, by necessity, on detailed analyses of narrowly defined events, which often makes it difficult to evaluate and appreciate the results in a broader context. Even when genome-wide approaches are employed to examine a proposed mechanism, the studies must remain limited in scope. One example of this challenge, as discussed below, concerns the large number of proteins that have been reported in the literature to interact with NF-κB, as well as the large number of post-translational modifications of NF-κB. Most interactions have been described in studies performed in a specific cell type and specific physiological setting. Partly for this reason, it has been difficult to evaluate the significance of each interaction from the perspective of the global or “big-picture” logic through which NF-κB’s functions are regulated. A likely reason more interactions and post-translational modifications have been reported for NF-κB than for other transcription factors is that NF-κB has been studied more extensively; it therefore is only a matter of time before a similar level of complexity emerges in the literature for other key transcription factors.

The challenge of extracting broad logic from narrowly defined studies is compounded by the second challenge, which centers on the many technical limitations currently inherent to studies of molecular mechanisms in mammalian cells. It is noteworthy that the preferred approach toward the dissection of a molecular mechanism is to study that mechanism in isolation, to avoid confounding influences of other components of the system that are not yet well-understood. As an example, our most compelling knowledge of basic molecular mechanisms, such as the mechanism of transcription initiation by RNA polymerase II,

mechanism of pre-mRNA splicing, and mechanism of DNA replication, has come from studies performed in cell-free reactions. The general approach has been to define cell-free conditions that support the molecular reaction of interest and then to reconstitute the reaction with pure components. This approach makes it possible to define the proteins and other molecules and macromolecules that are essential for the reaction, and to then analyze the reaction mechanism in biochemical and structural detail.

However, developmental stage-specific and inducible transcriptional regulation have not yet been recapitulated convincingly in a cell-free system. Furthermore, when considering the evidence that chromatin and nuclear architecture are important for proper gene regulation, combined with our imprecise knowledge of the higher order chromatin structures and nuclear features that are required for proper gene regulation and the great challenge of recapitulating native chromatin and nuclear environments in a cell-free system, a reliable cell-free approach is unlikely to be a viable option for many years. Even if a cell-free approach were possible, it would be less desirable for an analysis of a complex regulatory mechanism than it has been for analyses of simpler multi-component reactions, such as the basic transcription initiation or pre-mRNA splicing reactions. The reason for this assertion is that the results obtained in a cell-free reaction can be strongly influenced by the concentrations and specific activities of the reaction components. Altering the concentration of each transcription factor, co-regulatory protein, or chromatin protein in a complex reaction would influence the relative importance of many components of the reaction (i.e. the rate-limiting step), making it difficult to truly understand how regulation is achieved in a native in vivo setting.

Given the need to rely on in vivo experimental approaches with inherent constraints, it is important to recognize the technical limitations of current methods as we strive to advance our knowledge of gene regulation mechanisms and circuitry. One common technical challenge is distinguishing direct from indirect effects. Another challenge is distinguishing cause from effect, and a third is assessing the functional relevance and role of protein-protein interactions and post-translational modifications involving transcription factors, co-regulators, and chromatin proteins. Additional challenges are related to redundancy between transcription factors, which often makes loss-of-function experiments less informative than desired. Because of these and other technical limitations, combined with the necessary focus on narrowly defined physiological settings, most current studies of molecular mechanisms are able to provide variable levels of support for a hypothesis; the results are generally suggestive rather than conclusive and leave unanswered questions about the relevance of the mechanism being described within the broader regulatory circuitry.

As one general example, there has been considerable excitement about the role of chromatin in immune cell development and activation. Support for this important role has been provided by many studies in which immune defects have been observed in mice deficient in regulators of chromatin structure. Although these loss-of-function studies have firmly established the basic principle that chromatin is important for transcriptional regulation in the immune system, it has been difficult in most instances to convincingly elucidate the underlying molecular mechanisms. Most studies describe gene expression changes that help explain the knockout phenotype. However, it is difficult to convincingly explain, at a

mechanistic level, how loss of the chromatin protein leads to the observed gene expression changes. It can be equally difficult to define the changes that are truly relevant to the observed phenotype.

When studying gene regulation by chromatin, distinguishing cause from effect, as well as direct and indirect effects, represent major challenges. For example, changes in histone modification or DNA methylation levels may be observed at genes that are aberrantly expressed in cells deficient in a transcription factor or chromatin regulator. However, changes in DNA methylation levels and the levels of some histone modifications may often be a consequence rather than a cause of gene expression changes. Chromatin immunoprecipitation (ChIP) results may further show that the protein of interest can bind the aberrantly expressed genes, and it may be possible to use co-immunoprecipitation to provide evidence that the chromatin regulator interacts with a transcription factor that is thought to regulate the target genes. Each of these results can provide support for a hypothesis, but such results are usually far from conclusive. For example, many published results over the past 20 years have suggested that transcription factors and chromatin regulators interact with genomic sites at which they do not function, in addition to their functionally important targets [8, 9]. Furthermore, transcription factors often interact inefficiently and at low affinity with their important co-regulatory proteins, making it difficult to distinguish functionally important interactions from irrelevant ‘sticky’ interactions in co-immunoprecipitation experiments.

The phosphorylation-dependent interaction of NF-κB with the p300 and CBP co-regulatory proteins serves as an example of the long way we need to go to understand gene regulation mechanisms and circuitry in the immune system. This example is noteworthy because it illustrates the limitations of our knowledge, even with respect to one of the most compelling and best-studied transcription factor/co-regulatory protein interactions in the immune system. p300 and CBP are large protein paralogs containing protein acetyltransferase domains and many other protein domains that remain poorly understood. A direct interaction between the RelA member of the NF-κB family and both p300 and CBP was found to be dependent on the phosphorylation of serine 276 (S276) in the Rel homology region of RelA [10, 11]. This phosphorylation event leads to a conformational change that allows RelA S276 and a RelA transcriptional activation domain to interact with two domains of p300/CBP. Notably, a mouse strain containing an alanine substitution mutation in RelA S276 was generated and exhibited greatly reduced activation of a subset of NF-κB target genes [12], providing strong evidence that both the phosphorylation event and the p300/CBP interaction are biologically important. This model has been further strengthened by recent structural and functional studies [13].

Although the RelA-p300/CBP interaction has been extensively studied, the results summarized above represent only the initial steps toward an understanding of this interaction, with many mechanistic and biological questions remaining unanswered. Mechanistically, if S276 phosphorylation is required for the activation of only a subset of NF-κB target genes, does NF-κB induce its remaining target genes via a p300/CBP-independent mechanism? If so, what is that mechanism? Are p300 or CBP fully dispensable for the activation of those target genes or are these co-regulatory proteins recruited by a

different transcription factor? Furthermore, how do p300 and CBP contribute to transcriptional activation by NF-κB? Do they primarily function by acetylating histone tails, by acetylating non-histone proteins, or is the co-activation function at these genes independent of protein acetylation? Are other domains of the large p300 and CBP proteins required for co-activation of NF-κB target genes and, if so, how do these domains function in this context? Are there differences between p300 and CBP or are these proteins fully redundant at NF-κB target genes? At S276 phosphoryation-dependent genes, is p300 or CBP the only co-activator needed for NF-κB activation or are other co-regulatory interactions with NF-κB also essential? Is S276 phosphorylation important primarily at inducible genes activated by RelA-p50 NF-κB heterodimers or is this activation mechanism also relevant to other RelA-containing dimers?

Biologically, how does RelA S276 phosphorylation help regulate the selective activation of NF-κB target genes in different cell types and physiological contexts? For example, are there stimuli that induce NF-κB nuclear translocation but not signaling pathways that lead to S276 phosphorylation? If so, what are the stimuli and does the absence of S276 phosphorylation play a major role in shaping the transcriptional response to these stimuli? Conversely, which stimuli induce RelA S276 phosphorylation and does a careful examination of the S276 phosphorylation-dependent target genes “make sense” in explaining the biological response to these stimuli?

It is important to reiterate that these fundamental unanswered questions are relevant to one of the best-studied post-translational modifications and protein-protein interactions regulating transcription in the immune system. Hundreds of other post-translational modifications and protein-protein interactions have been reported; a small number have been rigorously validated but await extensive analysis, and many others are still in need of rigorous validation. In particular, mouse strains that selectively disrupt a post-translational modification or protein-protein interaction have been generated in only a small number of instances. It is also noteworthy that the NF-κB family is one of the smallest transcription factor families known to be important in the immune system. Other transcription factor families, such as the C2H2 zinc finger, Hox, and Ets families, contain far more members, thereby further complicating efforts to understand the biological functions and mechanisms of action of each family member. Clearly, there is much to learn.

One additional topic of great interest to immunologists is the extent to which the development and function of cells of the immune system are regulated by true epigenetic mechanisms. Although epigenetics is sometimes used to refer to any contribution of chromatin to gene regulation, a more formal definition of epigenetic regulation is a cell state (or gene expression state) that can be maintained through cell division in the absence of the factors responsible for establishing the state. In the immune system, the best-characterized example of epigenetic regulation concerns the silencing of the *Cd4* gene in CD8+ thymocytes. A dedicated silencer region in the *Cd4* locus is known to play a central role in the initiation of *Cd4* silencing during CD8+ T cell development [14]. However, in mature CD8+ T cells, the silent state of the *Cd4* locus and the chromatin properties that characterize the silent state can be stably maintained in the absence of the silencer [14, 15]. This finding suggests that the silencer helps establish a repressive chromatin structure that is stably

maintained by a true epigenetic mechanism. Evidence has been presented that epigenetic mechanisms are used to regulate many developmental states, and may also help regulate memory responses in cells of both the innate and adaptive immune systems [16, 17]. However, only a small number of experimental tests of these hypotheses have been reported, highlighting the need for further investigation.

And yet, with enthusiasm, here we go!

The purpose of this status report is to highlight the fact that, despite great progress, especially in the discovery of transcription factors that are important for the development and function of cells of the immune system, much remains to be learned about both the global and gene-specific logic of transcriptional regulation. As emphasized, the two main reasons for our current gaps in understanding are the complexity of gene regulation mechanisms combined with a large number of unavoidable technical limitations. It can be argued that a third reason for our limited knowledge is that the molecular immunology community has not had sufficient time to uncover the regulatory logic, with the first mechanistic advances occurring a mere 30 years ago. Also of relevance, the first human genome sequences were reported only 13 years ago. Since that time, the gene regulation and genomics fields have dedicated an enormous amount of effort to the development of powerful experimental and computational methods to take advantage of genome sequence information. After each sophisticated new technology is developed, a substantial amount of time is needed for researchers to learn how to use it to extract meaningful new insights.

With this in mind, it can be argued that this is an especially exciting time for molecular immunologists studying gene regulation, as we remain immersed in efforts to determine how to extract important mechanistic and biological insights from the powerful genomics and proteomics methodologies that have been developed over the past several years. Sophisticated genomics approaches are essential for efforts to uncover the global logic through which transcription regulates biologically important events, but conventional molecular and biochemical studies will also continue to be necessary and of great value. The recent emergence of RNA sequencing is especially noteworthy, given its ability to provide the information that forms the foundation for any study of transcriptional circuitry: that is, accurate and quantitative knowledge of the relative abundances of all transcripts in a cell population or single cell. The discovery of the CRISPR/Cas9 system also holds great promise for the future for the manipulation of genes and genomes with unprecedented ease and speed [18]; this system will make it easier, for example, to scrutinize the relevance of post-translational modifications and protein-protein interactions, by examining the impact of targeted mutations of key residues in mice and other relevant models. Furthermore, continual improvements in single-cell methods and methods for studying the relationship between transcription and nuclear organization will enhance our understanding of transcriptional dynamics and the mechanisms through which intra- and inter-chromosomal interactions and nuclear compartmentalization contribute to transcriptional control. Together, these diverse experimental strategies should gradually reveal the logic through which individual genes and groups of genes are regulated in key immunological settings, as well as the mechanisms by which immune development and immune responses are

coordinated by interactions between individual cells and between cell populations and tissues.

References

1. Parslow TG, et al. Structure of the 5′ ends of immunoglobulin genes: a novel conserved sequence. Proc Natl Acad Sci. 1984; 81:2650–2654. [PubMed: 6425835]
2. Singh H, et al. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature. 1986; 319:154–158. [PubMed: 3079885]
3. Staudt LM, et al. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature. 1986; 323:640–643. [PubMed: 3095662]
4. Banerji J, et al. A lymphocyte-specifi cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell. 1983; 33:729–740. [PubMed: 6409418]
5. Gillies SD, et al. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell. 1983; 33:717–728. [PubMed: 6409417]
6. Sen R, Baltimore D. Inducibility of kappa immunoglobulin enhancer-binding protein Nf-kappa B by a posttranslational mechanism. Cell. 1986; 47:921–928. [PubMed: 3096580]
7. Mercer EM, et al. Factors and networks that underpin early hematopoiesis. Semin Immunol. 2011; 23:317–325. [PubMed: 21930392]
8. Walter J, et al. Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in Drosophila embryos. Genes Dev. 1994; 8:1678–1692. [PubMed: 7958848]
9. Carey MF, et al. Confirming the functional importance of a protein-DNA interaction. Cold Spring Harb Protoc. 2012; 2012:733–757. [PubMed: 22753608]
10. Zhong H, et al. Phosphorylation of NF-κB p65 by PKA stimulates transcriptional activity by promoting a novel bivalent interaction with the co-activator CBP/p300. Mol Cell. 1998; 1:661–671. [PubMed: 9660950]
11. Zhong H, et al. Phosphorylation of nuclear NF-κB governs its association with either HDAC-1 or CBP/p300: a mechanism for regulating the transcriptional activity of NF-κB. Mol. Cell. 2002; 9:625–636.
12. Dong J, et al. Epigenetic regulation of NF-κB dependent gene expression. Genes Dev. 2008; 22:1159–1173. [PubMed: 18408078]
13. Mukherjee SP, et al. Analysis of the RelA:CBP/p300 interaction reveals its involvement in NF-κB-driven transcription. PLoS Biol. 2013; 11:e1001647. [PubMed: 24019758]
14. Gialitakis M, et al. The epigenetic landscape of lineage choice: lessons from the heritability of CD4 and CD8 expression. Curr Top Microbiol Immunol. 2012; 356:165–188. [PubMed: 21989924]
15. Zou YR, et al. Epigenetic silencing of CD4 in T cells committed to the cytotoxic lineage. Nat Genet. 2001; 29:332–336. [PubMed: 11687799]
16. Foster SL, et al. Gene-specific control of inflammation by TLR-induced chromatin modifications. Nature. 2007; 447:972–978. [PubMed: 17538624]
17. Otsuni R, et al. Latent enhancers activated by stimulation in differentiated cells. Cell. 2013; 152:157–171. [PubMed: 23332752]
18. Yang H, et al. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell. 2013; 154:1370–1379. [PubMed: 23992847]

Highlights

- Identification of many transcription factors relevant to immune development and function
- Identification of regulatory signaling pathways, co-regulators and relevant chromatin modifiers
- Progress necessary in elucidation of gene regulation networks and cell-type specific regulation
- The complexity of gene regulation circuitry and current technical limitations present barriers
- Inroads into understanding gene-specific and global logic provided by new technologies