HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2021 October 27.

Published in final edited form as:

Nat Immunol. 2020 February ; 21(2): 110–112. doi:10.1038/s41590-019-0566-z.

Gene regulatory networks STARR-ing B cells

Alexia Martínez de Paz, **Steven Zvi Josefowicz**

Department of Pathology and Laboratory Medicine and Immunology and Microbial Pathogenesis Program, Weill Cornell Medicine, New York, NY, USA.

Abstract

A genome-wide screening of functionally active enhancers, combined with analyses of chromatin features, transcription factor binding and gene expression, reveals general principles of gene regulatory networks in activated B cells.

Cell fate transitions, both during development and in response to external stimuli, depend upon the establishment of specific transcriptional programs by gene regulatory networks (GRNs). These complex networks are based on the interplay between (1) '*trans*-acting' factors: transcription factors (TFs) and chromatin regulatory factors, including coactivators and corepressors; (2) '*cis*-acting' regulatory DNA elements, including enhancers and promoters; and (3) biophysical features of the chromatin itself, namely the accessibility of DNA in chromatin and post-translational modifications (PTMs) of histones and DNA. With their varied differentiation from a common hematopoietic progenitor and a shared capacity for rapid responses to damage or pathogen sensing, cells of the mammalian immune system represent ideal subjects for studying the complex interplay between *trans*-acting factors, regulatory DNA, and chromatin characteristics of GNRs. B cells are especially interesting because they balance developmental programming and rapid response capability with germline B cell antigen receptor (BCR) immunoglobulin rearrangement and remarkable rates of cellular proliferation in the germinal center reaction, additional processes involving extensive chromatin regulation. In this issue of *Nature Immunology*, Chaudhri et al.[1] leverage multiple '-omics' approaches to untangle the complexity of GRNs in B cells upon bacterial lipopolysaccharide (LPS) stimulation, an activation method broadly used to model B cell differentiation and germinal center entry programs.

In part because of the suitability of B cells for GRN studies, the first 'endogenous' (non-viral) tissue-specific enhancer was discovered in B cells in 1983 by Susumu Tonegawa and Walter Shaffner and their colleagues[2,3]. The enhancer, at the immunoglobulin heavy chain (*Igh*) locus, is positioned preceding the Cμ coding sequence, well downstream of the promoter and V, D and J sequences, and influences transcription only following receptor rearrangement, when it is brought into proximity with the promoter. These early findings set the stage for modern studies of regulatory DNA in general, but notably also highlighted the importance of location, physical proximity and higher-order chromatin architecture in regulatory events involving DNA elements and their associated transcriptional and

alm2806@med.cornell.edu;.

chromatin regulatory factors. Beyond early recognition of the importance of spatially linked control of regulatory DNA elements in immune receptor rearrangement and oncogenic translocations[4], recent awareness of the dynamics of chromatin architecture in 3D space makes these concepts broadly applicable.

One principal challenge of studying GRNs is the accurate identification of functional enhancer regions, DNA regulatory elements that can promote gene expression independent of the distance and orientation to the transcriptional start site. During the last decade, the development of whole-genome sequencing technologies and collaborative efforts such as the Encyclopedia of DNA Elements (ENCODE[5]) or the Immunological Genome Project (ImmGen[6]) consortia have substantially improved our ability to detect putative regulatory regions by genome-wide profiling of transcription factors and histone modifications through chromatin immunoprecipitation and sequencing (ChIP-seq), chromatin accessibility by formaldehyde-assisted isolation of regulatory elements and sequencing (FAIRE-seq[7]), DNase I hypersensitive site sequencing (DHS-seq[8]), or assay for transposase accessible chromatin and sequencing (ATAC-seq[9]). However, despite the utility of these datasets, a major limitation to understanding GRNs at a global genomic level has been the inability to define which regulatory elements are active — capable of causally driving transcriptional changes at associated genes — rather than simply associated with descriptive features of ‘active’ chromatin, including accessibility, TF binding and the ‘active’ histone PTMs H3K27ac, H3K4me1 and histone variant H3.3. Traditionally, experimental analyses of enhancer activity have been performed using reporter assays that test individual candidate regions. However, recent complementary approaches have enabled high-throughput, genomic functional assessment of enhancers: clustered regularly interspaced short palindromic repeats (CRISPR)-based regulatory element screening[10] and massively parallel reporter assays such as functional identification of regulatory elements within accessible chromatin (FIREWACh[11]) or self-transcribing active regulatory region sequencing (STARR-seq[12]) — plasmid-based approaches that allow the assessment of thousands of regions simultaneously.

The power of the combination of such approaches to study GRNs is highlighted by Chaudhri et al.[1], who combine transcriptomics with TF and chromatin profiling to identify and characterize active enhancers governing cell fate decisions in B cells (Fig. 1). In this study, the central strategy to identify active enhancers is the combination of FAIRE-seq with STARR-seq. The first methodology detects accessible chromatin regions (much like ATAC-seq), a common surrogate for putative regulatory elements such as enhancers and promoters. Next, the analysis seeks functionally active regions among FAIRE-seq fragments through the construction of a STARR-seq library that is transfected into LPS-activated B cells. The STARR-seq plasmid reporter consists of placing the sequences to be tested downstream of a minimal promoter and before a polyadenylation site, so that active enhancers can transcribe themselves and the activity of thousands of candidate regions can be tested in parallel by RNA-seq[12].

Importantly, the authors’ annotation of functionally active enhancers — ‘STARR-seq-positive regions’, or sequences that are sufficient in stimulated B cells to drive transcription from a weak promoter in the STARR-seq plasmid — is validated by orthogonal approaches,

highlighting the suitability of this type of functional screening. First, a luciferase reporter assay in the B cell line CH12, a model for activated B cells, was used to demonstrate enhancer activity in 46 selected STARR-seq-positive regions, showing that a subset of STARR-seq-positive regions contained high-level functional activity. Second, a series of descriptive chromatin analyses were performed to reveal that STARR-seq-positive regions are especially enriched for chromatin accessibility, a stimulation-induced increase in the activity-associated histone modification H3K27ac, binding of the coactivator p300 and the architectural protein CTCF, and also a high abundance of enhancer RNAs (eRNAs).

Chromatin conformation capture techniques are useful for identifying contacts between regulatory elements that lie within distant regions in the linear genome. Here, genome-wide chromatin conformation capture, Hi-C[13], was performed to characterize the 3D chromatin organization of activated B cells. As an additional point of validation for their functional screening, STARR-seq-positive regions tend to be proximal to transcriptionally active promoters and within the same chromatin compartment or domain. It is important to note that, far from the classical notion of one enhancer controlling one gene via discrete, paired contact interactions, most active enhancer-promoter interactions are found in two main configurations: ‘multi-enhancer genes’ (genes that are connected to multiple enhancers) and ‘multigenic enhancers’ (single enhancers interacting with multiple genes). While Hi-C samples a population of cells and it is possible that the many-to-one contacts observed reflect a mix of pairwise contacts, it seems likely that the results reflect true aggregation of regulatory elements, consistent with an increasingly popular model of transcriptional control: the assembly of biomolecular condensates. This refers to the formation of membraneless organelles by liquid–liquid phase separation, in which multiple regulatory regions bound by TFs, coactivators and the transcription apparatus coalesce and concentrate to form liquid-like droplets, or ‘condensates’, to drive efficient, coregulated transcription[14]. Low-complexity (intrinsically disordered) protein domains, together with high-valency (low-affinity) protein–protein interactions between TFs, RNA polymerase II, MED1 (subunit of the mediator complex), the coactivator BRD4 and others, can drive condensate formation[14,15]. In this model, motif-binding TFs may ‘nucleate’ assembly of regulatory condensates at target loci or regulatory elements[16], with chromatin architecture dictating the genomic territories that are spatially amenable to recruitment. This model could explain another finding of Chaudhri et al.: that putative regulatory elements that are not transcriptionally active (STARR-seq-negative) were still found to interact with stimulation-responsive promoters and other regulatory elements (including STARR-seq-positive regions). While not sufficient to drive plasmid reporter expression, STARR-seq-negative regulatory elements may be recruited into condensates through histone and *trans*-acting factor characteristics and thus contribute to the spatial accumulation of the transcription apparatus.

The role of biomolecular condensates in transcription has been studied mainly in the context of super enhancers — clustered enhancers controlling key genes associated with cell functional identity[17]. It will be of interest to test if multi-gene enhancers are functioning through similar mechanisms. The presence of high levels of acetylated nucleosomes in the multi-gene enhancer configuration could be reflecting a higher concentration of factors (coactivators, histone acetyltransferases) that create biophysical characteristics amenable

to condensate formation. These especially robust chromatin regulatory features could be related to the coordinated activation of multiple genes (related to aerobic metabolism or ribosome biogenesis) involved in processes necessary to manage the transition from quiescent B cells to the more proliferative state of activated B cells.

Chaudhri and colleagues' analysis of genomic chromatin occupancy of B cell lineage-determining TFs and motif enrichments reveals that enhancers can be poised (made accessible) by the binding of a small number of lineage TFs (including EBF1, PAX5, TCF3, PU.1, ETS1 and PRDM1), but that enhancer activation is associated with extensive collaborative binding of these TFs along with additional signal-responsive TFs such as AP-1, IRF4, STAT and NF-κB. Along these lines, enhancer features that promote condensate formation, namely number and density of TF binding sites (reflecting the coordinated binding of multiple TFs), are also found to increase transcriptional activity in luciferase reporter assays[18]. It will be interesting to experimentally test whether combinations of the enriched motifs found here are able to recruit TFs necessary to nucleate the formation of condensates and a strong transcription response. Further functional characterization, including the use of DNA-editing technologies and focused studies of specific regulatory elements, will enrich and extend the general principles of GRNs in activated B cells described here. Overall, the work by Chaudhri et al. demonstrates the potential of integrative, multi-dataset analyses for studying GRNs, revealing extensive collaboration between TFs (both lineage and stimulation responsive), chromatin features and architecture, and regulatory DNA sequences. Further, the study represents an important mineable resource with the potential to answer (and raise!) additional questions on transcriptional and epigenetic control of B cell activation.

References

1. Chaudhri VK, Dienger-Stambaugh K, Wu Z, Shrestha M & Singh H Nat. Immunol 10.1038/s41590-019-0565-0 (2020).
2. Banerji J, Olson L & Schaffner W Cell 33, 729–740 (1983). [PubMed: 6409418]
3. Gillies SD, Morrison SL, Oi VT & Tonegawa S Cell 33, 717–728 (1983). [PubMed: 6409417]
4. Pettersson S, Cook GP, Brüggemann M, Williams GT & Neuberger MS Nature 344, 165–168 (1990). [PubMed: 2106628]
5. Dunham I et al. Nature 489, 57–74 (2012). [PubMed: 22955616]
6. Yoshida H et al. Cell 176, 897–912.e20 (2019). [PubMed: 30686579]
7. Giresi PG, Kim J, McDaniell RM, Iyer VR & Lieb JD Genome Res 17, 877–885 (2007). [PubMed: 17179217]
8. Song L & Crawford GE Cold Spring Harb. Protoc 10.1101/pdb.prot5384 (2010).
9. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ Nat. Methods 10, 1213–1218 (2013). [PubMed: 24097267]
10. Sanjana NE et al. F. Science 353, 1545–1549 (2016).
11. Murtha M et al. Nat. Methods 11, 559–565 (2014). [PubMed: 24658142]
12. Arnold CD et al. Science 339, 1074–1077 (2013). [PubMed: 23328393]
13. Lieberman-Aiden E et al. Science 326, 289–293 (2009). [PubMed: 19815776]
14. Sabari BR et al. Science 361, eaar3958 (2018).
15. Cho WK et al. Science 361, 412–415 (2018). [PubMed: 29930094]
16. Boija A et al. Cell 175, 1842–1855.e16 (2018). [PubMed: 30449618]
17. Whyte WA et al. Cell 153, 307–319 (2013). [PubMed: 23582322]

18. Shrinivas K et al. Mol. Cell 75, 549–561.e7 (2019). [PubMed: 31398323]

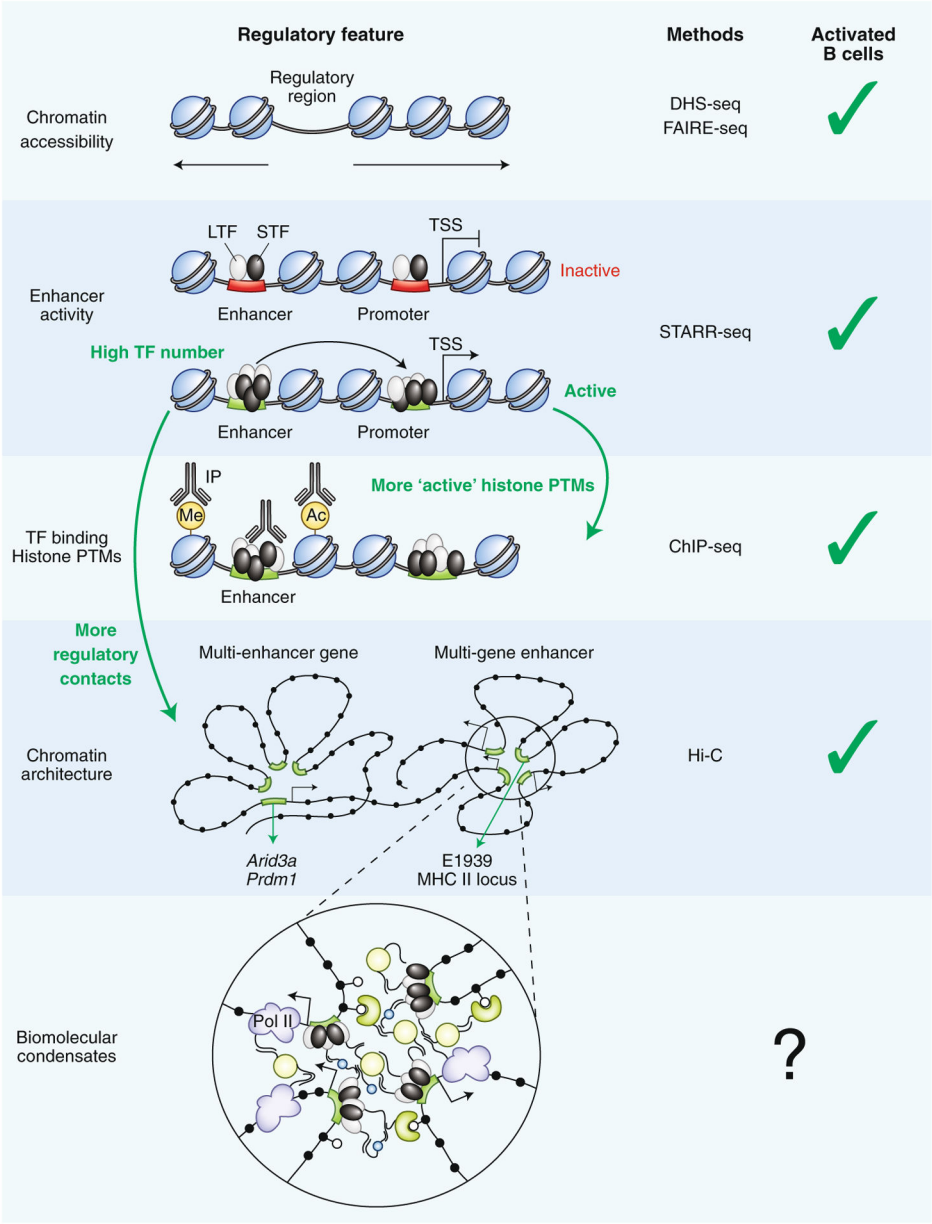

Fig. 1 |. High-throughput analyses of regulatory features define general characteristics of gene regulatory networks in lPS-activated B cells.

Active enhancers are characterized as highly accessible, with combinatorial binding (three or more) of lineage-defining (LTF) and signaling-induced (STF) transcription factors, and by the presence of activity-related histone modifications. Chromatin architecture analysis shows the existence of active regulatory hubs in two main configurations: 'multi-enhancer genes', as in the *Arid3a* or the *Prdm1* loci, and 'multi-gene enhancers', as in the enhancer E1939, which controls genes of the major histocompatibility complex (MHC) class II. This cooperativity in transcription regulation suggests the potential involvement of transcriptional condensates in B cell fate transitions.