CellPress

Trends in Immunology

Special issue: Systems immunology II

Opinion

A variegated model of transcription factor function in the immune system

Kaitavjeet Chowdhary[1] and Christophe Benoist[1,*]

Specific combinations of transcription factors (TFs) control the gene expression programs that underlie specialized immune responses. Previous models of TF function in immunocytes had restricted each TF to a single functional categorization [e.g., lineage-defining (LDTFs) vs. signal-dependent TFs (SDTFs)] within one cell type. Synthesizing recent results, we instead propose a variegated model of immunological TF function, whereby many TFs have flexible and different roles across distinct cell states, contributing to cell phenotypic diversity. We discuss evidence in support of this variegated model, describe contextual inputs that enable TF diversification, and look to the future to imagine warranted experimental and computational tools to build quantitative and predictive models of immunocyte gene regulatory networks.

Sense from complexity: untangling immunocyte TF networks

Cells of the immune system must mount diverse and highly specialized responses to foreign antigens, inflammatory triggers, and deviations from organ homeostasis. These responses are mediated by coordinated gene expression programs, each encompassing synergistic functionalities (homing, interaction ligands/receptors, and effector molecules). Such programs are controlled by the action of **TFs** (see Glossary), which in turn act on ***cis*-regulatory elements** (CREs), **promoters**, and **enhancers**. How TFs select their targets, are deployed in response to specific signals, or are organized combinatorially are questions of fundamental importance for the understanding and control of immunity. Given the thousands of transcriptional regulators encoded by mammalian genomes and the intricate packing of DNA in a nucleus, the complexity of the problem has led some to characterize efforts to truly comprehend TF function as futile [1]. Nevertheless, astounding advances in data generation capabilities and computational tools that are well suited to discover specific patterns within enormously complex data have enabled novel experimental strategies. These novel approaches are illuminating how TFs contribute to orchestrating immune responses and prompt a rethinking of previous conceptual frameworks. Here, we argue that a conceptual limitation has been the reductionist restriction of TFs to single functional categories. Instead, recent results support a view in which a given TF can have variegated roles across different cell states and genomic loci. Such variegation in dozens of TFs can contribute to the fine diversity of phenotypic states that individual cells adopt within a defined cell type.

We begin with a brief discussion of existing paradigms of TF function in the immune system, review recent studies supporting the variegated model of TF function, and conclude with forward-looking key questions for future investigation. We also speculate on new potential tools that would be needed to tackle such queries.

Existing models of TF function

Current paradigms tend to focus on the classification of TFs into individual functional categories. One framework, the 'collaborative-hierarchical' model, distinguishes '**LDTFs**', those with **pioneering** ability to displace **nucleosomes** and to open lineage-specific regulatory elements, from '**SDTFs**', which

Highlights

Existing models restrict transcription factors (TFs) to individual functional categories, and these apply across all contexts within a cell type.

Recent results suggest that, instead of having a single, fixed role, TFs can play varied functions across different settings, which may give rise to diversity in phenotypic cell states. We refer to this as the 'variegated model of TF function'.

This variation in function may be mediated by direct or indirect TF–TF crosstalk, or by interplay between TFs and chromatin state.

Descriptions of the many ways TFs can affect transcription emphasize that classic bind–activate (or bind–repress) scenarios vastly oversimply TF interplay and quantitative effects.

Alternatives to TF knockouts, such as natural *cis*-genetic variation, acute TF degradation, or continuous tuning of TF concentrations, may better capture direct TF targets.

New measurement tools and computational methods will be required to develop a quantitative and predictive understanding of TF-mediated gene regulatory networks in immunity.

Significance

Unlike previous models of TF function in immunity, which assigned a single role to each TF across all contexts within a cell type, recent results suggest that TFs flexibly have different roles across distinct cell states, with implications for an understanding and control of immunocyte gene regulatory networks.

 https://doi.org/10.1016/j.it.2023.05.001

[1]Department of Immunology, Harvard Medical School, Boston, MA, USA

*Correspondence:
cbdm@hms.harvard.edu (C. Benoist).

bind to elements primed by these LDTFs or collaborate with LDTFs to open latent or *de novo* enhancers [2–4]. Under this theoretical model, SDTFs lie below LDTFs in the hierarchy of TF function, with combinations of LDTFs selecting cell type-specific elements. LDTFs overlap with '**master TFs**', classically defined as factors that are both necessary and sufficient to bring about cell type-specific regulatory programs [5–8]. SDTFs can also be split into primary and secondary responder TFs, to account for the successive waves of gene activation and repression that occur after activation of any cell type [4]. Presynthesized TFs, which control primary responses, are separated from secondary responder TFs, which depend on other TFs for their expression and synthesis [4]. Importantly, by the nature of their categorization, such paradigms restrict each TF to a single functional role across all contexts and treat all individual cells within a cell type as identical.

Such TF hierarchies are not new and were highlighted a long time ago in models of *Drosophila melanogaster* development [9,10]. However, these models face several limitations. First, while conceptually convenient because of their reducing complexity, simplified models of linear operation of LDTFs followed by SDTFs do not reflect the convolutions of *in vivo* environments, where cells are simultaneously exposed to gradients of multiple inputs (cell–cell interactions, soluble mediators, and metabolites). For instance, no *in vivo* situation exists in which macrophages are solely activated by lipopolysaccharide (LPS), or T cells by CD3/CD28 engagement. Would LDTFs and their functions really remain impervious to such influences?

Second, such reductionist models ignore the many modalities through which any one TF might operate. From a theoretical standpoint, one can draw out many strategies by which TFs regulate transcription, beyond the simple bind–activate or bind–repress modes that we usually have in mind (Figure 1). There is no reason to assume that a regulator would be restricted to any one of these, especially if one considers that many TFs contain long stretches of **intrinsically disordered regions (IDRs)**, which confer functional flexibility and allow them to adapt to diverse cofactors [11]. For example, the long N-terminal region of FoxP3 is an IDR, and interacts changeably with many potential cofactors [12,13]; the TAZ1 domain of the transcriptional co-activators CBP/**p300** interacts with multiple diverse TFs with IDRs, such as CITED2 and STAT2 [11,14,15]. Furthermore, systematic deconstructions of human regulatory networks have shown that TFs are highly combinatorial in their function, with extensive network connections as well as feedback and feed-forward loops controlling regulatory activities [16].

Third, TF categorizations are not absolute. In one study of mouse fibroblasts, analysis of genetic variants affecting **assay for transposase-accessible chromatin using sequencing (ATAC-seq)** and **H3K4me1/H3K27Ac chromatin immunoprecipitation (ChIP)-seq**, overlaid with AP-1 ChIP-seq, showed that enhancer selection depended on AP-1 factors, which are typically classified as SDTFs, such that SDTFs in one context could be LDTFs in another [17]. Similarly, TCF1 (encoded in mice by *Tcf7*) has many guises: it is a pioneer factor in the determination of the entire T cell lineage, a late SDTF at the effector/memory fork of activated $CD8^+$ T cells, and a factor that is repressed to enable **regulatory T cell (Treg)** identity [18,19]. The master TF FOXA1 is considered a canonical pioneer in mammalian endodermal lineage specification, and HNF4A a non-pioneer factor [20,21]. However, upon ectopic overexpression in a lymphoblastoid cell line, both factors were able to open previously inaccessible regions (assessed by ATAC-seq and **CUT&TAG**) and induce endodermal gene expression programs [22,23]. Furthermore, HNF4A acts as a lineage-determining pioneer factor in hepatocyte differentiation [24] and in a subpopulation of **thymic mimetic cells** [25], further demonstrating the flexibility of TF pioneering capabilities.

Finally, many apparent 'master TFs' may in fact be markers, but not drivers, of specific programs. For example, FoxP3 has lost its throne as the master regulator of the Treg lineage. While it

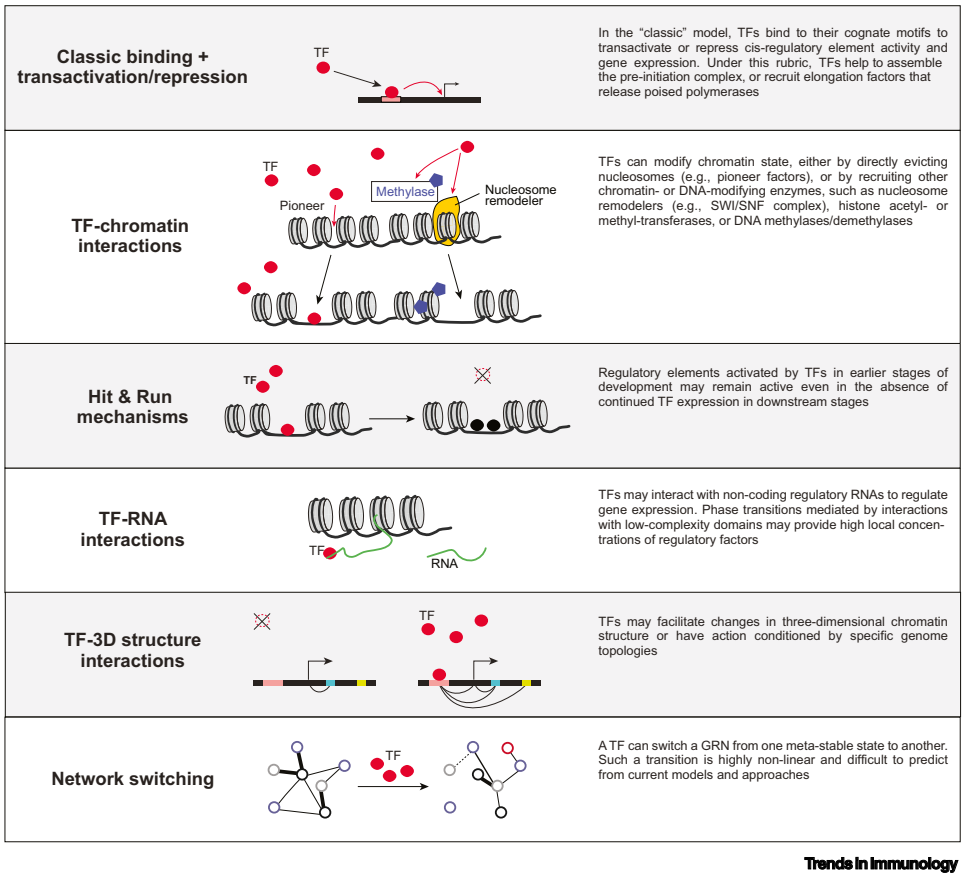

Figure 1. Beyond simple binding and transactivation: the diversity of transcription factor (TF) regulatory mechanisms. Several of these mechanisms are compiled and adapted from [23,93]. Abbreviation: GRN, gene regulatory network.

crucially contributes to Treg function, FoxP3 is neither necessary nor sufficient to establish Treg identity: **Treg-like cells** develop without it, and Treg-specific programs contain both FoxP3-independent and -dependent modules [26–32]. Moreover, while RORγ is classically considered the master regulator of **Th17 cell** differentiation, RORγ deficiency has been shown to lead to only modest changes in the chromatin landscape (by p300 and **H3K4Me3** ChIP-seq) of differentiating mouse Th17 cells. Instead, the lineage-defining effects of RORγ depend on priming by BATF, IRF4, and STAT3, which are normally considered SDTFs [33,34]. Furthermore, a comparison of RORγ target binding (by ChIP-seq) and accessibility (by ATAC-seq) in a pan-mouse immune system atlas found the targets of RORγ to be highly variable across different cell types [35]. Thus, the domain of master regulators may not equally reach across all cell conditions, furthering the limitations of absolute classification of TFs into single functional categories.

TF function: variegated not monomorphic

Whether in response to distinct external stimuli or across the continua of cell-to-cell variation, the synthesis of recent studies clarifies that the same TFs can perform distinct roles across diverse conditions (we do not enter here into the thorny debate over the boundaries between cell types and states [36–38]). However, we contend that this variegated model of TF function, depicted in Figure 2, applies across all models of mammalian cell-type configurations.

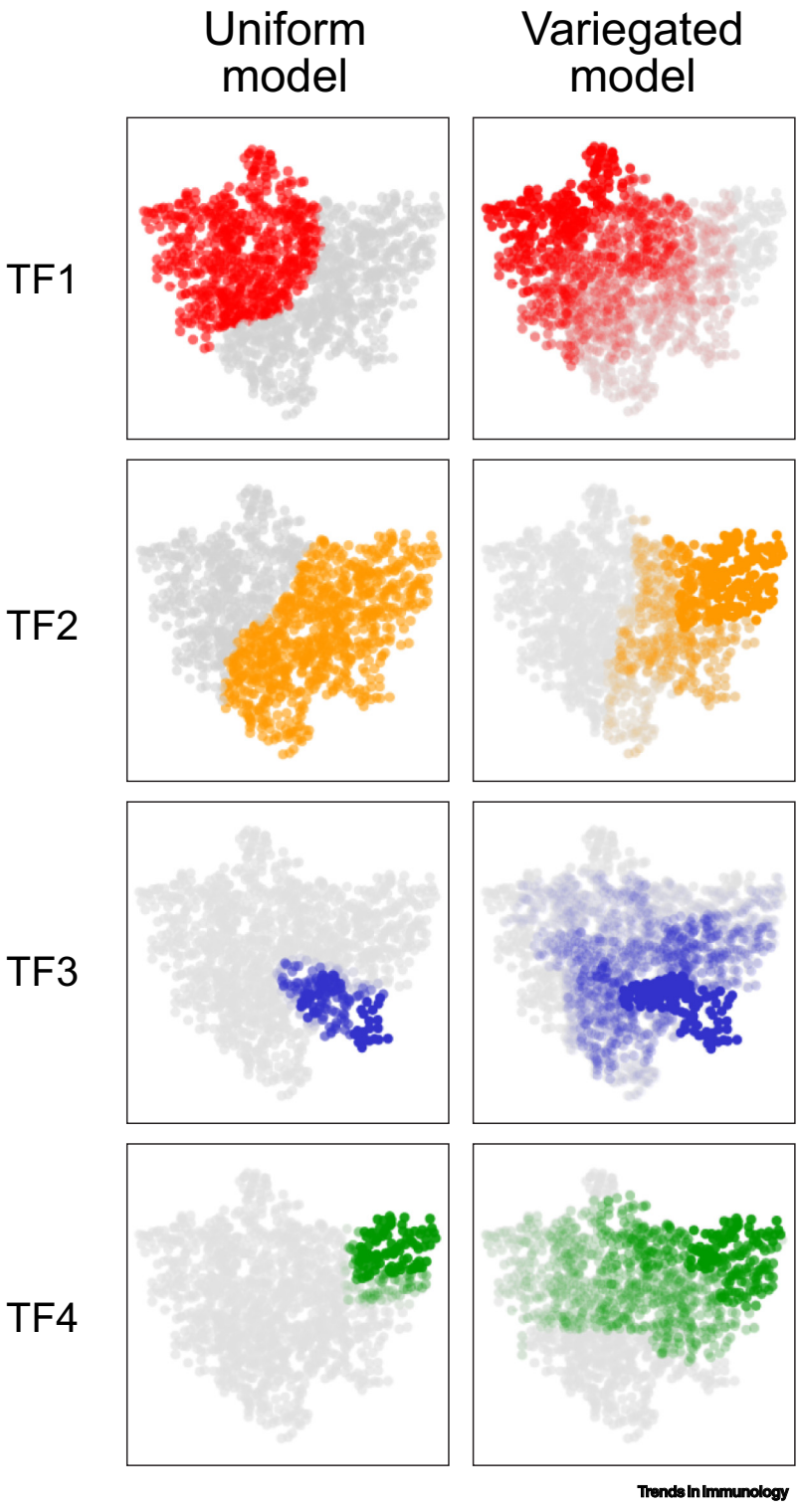

Figure 2. Variegated model of transcription factor (TF) function. Under this model, instead of each TF having uniform, monomorphic activity across all phenotypic cell states, they can have flexible, variegated functions across different settings.

Recent studies illustrate how TF functions can vary in response to extracellular stimuli. One such study leveraged natural genetic variation in **F1 hybrids** (Box 1) of different mouse strains to uncover the causal determinants of chromatin accessibility (readout by ATAC-seq) in naive and activated $CD4^+$ and $CD8^+$ T cells, following acute lymphocytic choriomeningitis virus (LCMV) infection (Armstrong strain for acute infection) [39]. A small number of TF families, including Ets, Runx, and TCF/Lef factors, were the major drivers of chromatin accessibility in both naive and activated conditions. Based on TF-binding data (measured by **CUT&RUN**), Ets1, Runx1, and TCF1 collectively occupied upward of 90% of accessible chromatin in naive T cells. The overwhelming contribution of only three TF families was surprising given the large number of TFs expressed in these cells. However, each TF family responded differently to $CD4^+$ and $CD8^+$ T cell activation. Specifically, while Ets1 acted as a 'housekeeper', by binding primarily to regions that remained unchanged in terms of accessibility (e.g., promoters of housekeeping genes) following activation, Runx1 appeared to amplify activation-dependent changes because its binding coincided with those sites that changed the most in accessibility relative to naive cells. Conversely, TCF1-bound regions became less accessible during the response to infection and sometimes became newly occupied by activation-induced TFs. This led the authors to suggest that TCF1 acts as a 'placeholder', which maintains accessibility in naive cells. Thus, three TF families that occupied most of the accessible T cell genome exhibited diverse binding behaviors, which would not have been captured by **motif enrichment analyses** and would have been refractory to knockout (KO)/perturbation approaches (difficult to interpret given their widespread functions). Of note, in a study of natural killer (NK) cells activated during acute *Toxoplasma gondii* infection in mice [40], STATs were the primary drivers of new enhancers (based on analysis of STAT motifs and binding), in addition to p300 deposition and ATAC-seq in STAT-deficient cells; furthermore, ChIP-seq analyses showed a redeployment of T-bet (a presumed LDTF) to loci containing STAT but not T-bet motifs. This observation suggested that TF crosstalk can reverse the traditional roles of LDTFs and SDTFs because TFs canonically considered to be SDTFs could control the activity of a TF otherwise classified as an LDTF; this further illustrates the variegation of TF function in response to distinct external influences.

Variegated TF function extends beyond binary comparisons of activated versus resting immunocytes. Single-cell approaches have enabled the study of TF functions across the full

Box 1. Using natural genetic variation to identify chromatin regulators genome wide

Identifying how TFs globally control immunocyte chromatin state requires genome-wide functional experiments. Several studies have collectively identified the major important players in immunocyte chromatin states by knockout (KO) of individual TFs (although surprises may still be lurking). These were continued more systematically at high throughput with RNAi or CRISPR/Cas9 screens [78–80], which also offer cell-type resolution (e.g., **Perturb-seq** or equivalent [81–83]). However, the inherent issues with gene KOs (Box 2), or the issues long recognized to plague TF overexpression experiments [84], may limit the ability of reverse-genetics approaches to identify the major regulatory nodes at a systems level.

In recent years, an increasing number of investigators have taken advantage of natural genetic variation across different inbred mouse strains [17,32,39,50,85–87] (similar exploitation can be made with data sets from humans, which are heterozygous at most loci, although interpretation is more difficult since phasing and linkage disequilibrium structure introduce distinct complications, unlike inbred mouse crosses, in which the continuous sequence of each chromosome is known). In particular, F1 hybrids between distant strains offer the opportunity to identify causal regulators genome wide [17,32,39,50,85–87]. The principle of the experiment is to compare, within the same cell, with a common concentration of *trans*-regulators, the consequence of a low-affinity variant in a TF binding site in one allele compared with a high-affinity allele. Performing this comparison across all such binding sites or motif instances is equivalent to a genome-wide mutagenesis screen, powered by the 5–20 million SNPs that can distinguish any two inbred strains (and tend to be enriched at *cis*-regulatory elements) [88]. Such studies can help delineate which sequence elements are required for chromatin features at sites of interest, genome wide. This strategy adds a layer of information beyond simple motif enrichment and enables causal inference of TF contributions at scale, while maintaining physiological TF concentrations [89,90].

complexity of *in vivo* cell continua elicited by combinatorial gradients of external signals. Tregs offer a good model for this question given that their inherent reactivity to self, combined with their ability to recognize non-self, results in a broad range of cell states at baseline in lymphoid and nonlymphoid tissues [41]. These diverse states reflect their equally diverse functions in preventing autoimmunity and maintaining organismal homeostasis [41–43]. Recent work involved using **single-cell ATAC-seq (scATAC-seq)** to identify TFs controlling the diversity of mouse Tregs [32]. Using a machine learning approach (**topic modeling**) [44,45] to extract groups (topics) of open chromatin regions (OCRs) with covarying patterns of activity across single cells, the state of each cell could be decomposed into combinations of variably active chromatin programs, reflecting the integration performed by each cell *in vivo.* Analyses leveraging allele-specific scATAC-seq of Tregs from B6/Cast F1 hybrid mice identified the causal regulators of each of these programs, as also validated by Treg-specific TF KOs. Some of the major TF families that have been previously reported [39] (e.g., Ets, TCF/Lef factors) contributed to the accessibility of many Treg chromatin programs. However, many other TFs (including NF-κB, GATA, AP-1, or nuclear receptor family members) exhibited highly focal effects, affecting the accessibility of only a narrow selection of chromatin programs, themselves active in specific subpopulations. These selective impacts on chromatin program accessibility occurred despite enrichment of corresponding TF motifs across a broader range of Treg programs. Supporting the model of variegated TF activity, this finding suggested that, while TFs are present across many contexts, their functional effects are more restricted, reflecting a dependence on cell state. Further underscoring the point of differential TF effects across varied cell states, analysis of FoxP3-dependent OCRs from comparisons of FoxP3 KO and wild-type cells across the single cell space showed that FoxP3 could act as a repressor of accessibility in some chromatin programs but an activator of accessibility in others. Thus, TF function may be highly conditional on cell state, in some cases even reversing its direction in different groups of OCRs and different cell populations. We posit that variegated TF effects, dependent on cell state, may manifest as overlapping gradients of activity across the single cell chromatin space.

Determinants of TF variegation

If TF function varies across cell states, what determines context specificity? Much as numerous mechanisms can explain the activity of a given TF (Figure 1), TF interactions can arise through several modalities (Figure 3). These diverse modalities of TF–TF interactions might then account for high variability across cells, and for the overlapping gradients of TF function that are observed in single-cell data sets. Moreover, combinatorial integration of several TFs is a long-appreciated mechanism to achieve specificity, going back to the first descriptions of regulatory control by TFs [2,3,9,10,16,46]. Recent examples include distinct compositions of LDTFs at B cell functional versus nonfunctional enhancers [47], colocalization of bZIP and T-box motifs specifically at sites with increased Runx1 occupancy in activated T cells [39], and activating or repressive FoxP3 regulatory effects mediated by distinct FoxP3-cofactor ensembles [13]. Thus, state-specific TF function might be conditioned by the availability of partner TFs; possible mechanistic interactions are schematized in Figure 3.

Persistent expression of a TF is not always essential to modulate the action of other factors: in so-called ‘hit & run’ mechanisms, regions opened earlier during differentiation can remain active in the absence of the continued expression of the pioneer factor. For example, in mature Tregs, FoxP3 binds to loci opened first by other Forkhead factors (e.g., FoxO1) in mouse thymic T cell differentiation [48]. **Nucleosome remodeling** can account for pioneer events around lineage commitment that open loci to subsequent TF action, but remodeling is also involved at later stages. For instance, in response to activation signals in mouse fibroblasts, AP-1 family TFs can recruit nucleosome remodelers of the BAF complex to mediate enhancer selection, which

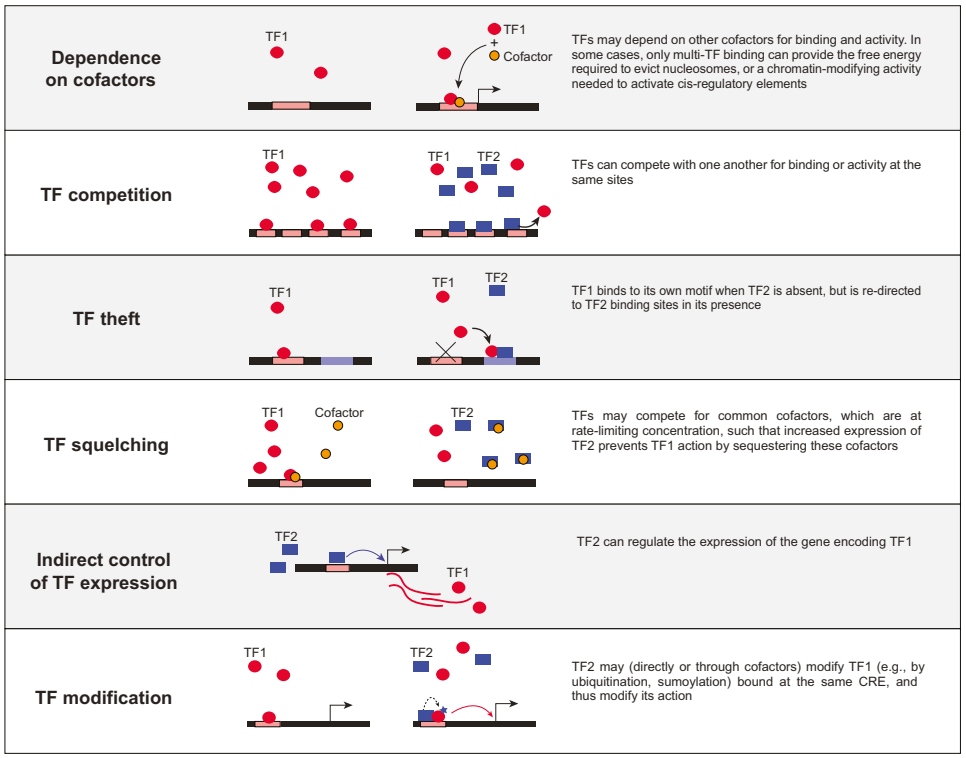

Figure 3. Theoretical mechanisms of transcription factor (TF)–TF interactions and crosstalk: how one TF can modify the activity of another. The schematics depict how a secondary TF ('TF2') or cofactor can modify the activity of a first TF ('TF1'), directly or indirectly.

then enable further function at serum-responsive genes [17]. Similarly, during mouse T cell activation, most sites bound by Ets1 in naive $CD4^+$ and $CD8^+$ T cells remain unchanged in accessibility, but a small subset of Ets1-bound sites undergo a dramatic increase in accessibility [39]. These sites lack canonical Ets1 motifs and colocalize with a distinct group of activation-induced TFs as well as Brg1, a component of the SWI/SNF chromatin remodeling complex [39]. Thus, the specific activity of these Ets1-bound sites depends on both chromatin remodeling and a specific complement of TFs. Notably, despite the influence of chromatin remodelers, motifs from other TFs that are bound at these locations still have a positive effect on chromatin accessibility based on F1 genetic variation analysis [39]; this suggests that TFs are not simply recruited to accessible sites, but also independently contribute to chromatin states at their target enhancers.

From another angle, TF–TF interactions need not operate directly at TF-binding sites: changes in cell state can also modulate TF activities remotely. For instance, in PU.1 overexpression experiments and in surveys across mouse early T cell development, ChIP-seq experiments showed that PU.1 (the main function of which is as a LDTF for myeloid and B cells) can redirect Satb1 and Runx1 binding to lower affinity target regions within PU.1-bound sites via TF 'theft' [49]. Given that PU.1 expression is limited to a specific time window before final T lineage commitment, this process provides stage-specific control of otherwise 'constitutively' bound TFs [49]. Similarly, upstream regulators can provide contextual information by regulating the expression of other TFs. For example, a recent study found that FoxP3 repressed the closely related conventional T cell (Tconv) lineage program by downregulating the expression of *Tcf7* (encoding TCF1, a positive regulator of Tconv-like chromatin accessibility)

[50]. However, how much of Treg identity is truly determined by this mechanism is debated [51]. Thus, whether by redirecting TF binding or modulating TF expression, one TF can affect the activity of another without directly binding to the same genomic position.

Another mechanism by which TFs can provide contextual information is by modifying chromatin state via the epigenetic recruitment of methyltransferases and/or demethylases to control **DNA CpG methylation** (the forms and roles of DNA methylation have been recently discussed [52–54]). In a recent methylation profiling study across six lymphoid (B, $CD4^+$ T, $CD8^+$ T, and NK cell) and myeloid (monocyte and granulocyte) immunologic cell types in healthy humans, lineage-specific TFs bound to cell type-specific hypomethylated regions [55]. Hypomethylated sites bound by LDTFs in one cell type were correspondingly methylated in other cell types. This is relevant because, generally, methylation can directly inhibit TF binding [56], thus providing negative regulatory control of TF function. However, TF function does not exclusively occur subsequent to chromatin remodeling: TFs can themselves regulate both chromatin remodeling and methylation [57]. For example, PU.1, previously reported to direct site-specific methylation [58], has been reported to bind to methylated regions in each of the profiled human cell types that were discussed [55], consistent with the hypothesis that PU.1 might partially control lineage specification via methylation of lineage-inappropriate loci. Notably, based on nucleosome-binding arrays, PU.1 can also recruit nucleosome remodelers via its IDR [23,59]. Therefore, the dichotomy that is often drawn between sequence-specific TF binding and chromatin mechanisms is better described as a feedback loop in which TFs can both modify and respond to changes in chromatin to diversify their regulatory roles. Each chromatin-modifying process provides an additional layer of regulatory information, which can act at different timescales.

A gene regulation 'wish list'

As recent studies suggest, if the regulatory landscape is a ballet of dozens of interacting TFs, all varying along cross-cutting gradients of activity in the space of individual cell phenotypes, should we also give up on truly understanding immunological gene regulation? There is hope, perhaps, in that the convoluted architectures recently uncovered do show structure and are reproducible between individuals. In truth, it is an exciting time to be studying immune gene regulation, and the vistas of the activity of the genome that we observe, even if bewildering, are awe inspiring. What technologies might we wish for to answer open questions regarding chromatin landscapes and transcriptional regulation? Here, we select a 'wish list' (by no means exhaustive) of such questions and possible solutions, some already emerging, others still far off.

Can we move beyond binary descriptions of TF binding toward quantitative measurements of *in vivo* genome-wide affinities, and understand residence times of TFs at specific loci? Investigators must currently make a trade-off between using genomic TF-binding assays to gain descriptive information on TF binding and genomic localization *in vivo,* or make quantitative measurements of TF biophysical parameters *in vitro* (affinities and dissociation constants) under conditions that bear little resemblance to the physicochemical milieu of the nucleus of a cell. Two recent studies [60,61] measured genome-wide ATAC-seq or TF binding across a range of TF concentrations, either by tunable expression or by varying the amount of TF spiked into each reaction. The changes in accessibility or binding signals across a range of TF concentrations enabled derivation of per-locus TF affinity parameters (e.g., K_d and Hill coefficients), providing an initial proof-of-principle toward genome-wide kinetic and affinity measurements, but which require robust validation [60,61]. It will also be important to ascertain true co-binding of two TFs (not only similar localization in parallel immunoprecipitation experiments). One could also envision imaging solutions that read out the interplay of different TFs at a given enhancer in real time to grasp binding and interaction parameters at and within individual regulatory elements.

One challenge is adapting such solutions to be tractable in primary cells (rather than tumor cell lines) and to make them scalable to assaying multiple TFs simultaneously.

Moreover, what are the direct targets, and what are the binding preferences of each TF *in vivo*? Currently, most summaries of TF binding rely on **position-weight matrices**, which are derived from *in vitro* binding assays or ChIP-seq experiments. However, such descriptions reflect only major patterns in binding, and may not capture the full range of *in vivo* binding preferences of a given TF. Moreover, an open question is how faithfully ChIP-seq represents true *in vivo* binding [62,63]. In addition, new tools will be needed to disentangle primary from secondary TF effects. Instead of TF gene KO experiments, inducible TF protein degradation, such as by **degrons** [64], **proteolysis-targeting chimeras (PROTACs)** [65], or other chemical means [66], might enable an improved assessment of TF control in a time-resolved manner because acute depletion would allow study of the immediate direct dependencies without the compensatory secondary effects of brutal KOs (Box 2). New developments are warranted to sample changes over time in response to TF reduction to quantitatively measure biochemical kinetic parameters at TF targets [67]. Recent work has also developed CRISPR/Cas9-based approaches to identify upstream CREs and *trans*-regulatory factors controlling the expression of key immune response genes [68,69]. These will be important for mapping complete networks of gene expression responses that both impact, and are impacted by, a given TF.

How can we develop a predictive and deeply granular understanding of immunocyte gene regulatory networks? Tackling this challenge will require the integration of all the questions and approaches discussed in the preceding text. New advances in multimodal data integration are beginning to reconstruct gene regulatory networks from single-cell genomics data [70–72]. These will need to be further developed in conjunction with experimental perturbations of putative regulators and kinetic measurements with or without immunologically relevant stimuli; in turn, this may enable movement toward mechanistically grounded predictions of network responses to varied inputs. Given that changing every possible configuration of such regulatory networks is impossible, model-guided experimentation will be required to guide future investigation.

Box 2. Assessing TF function (TF KOs are not always the answer)

Genetic KOs have generally been considered the gold standard for discerning the causal chain of TF actions. However, this approach is not without limitations, which are worth considering in the design of experiments to assess TF function. First, both direct and indirect effects manifest in TF KOs. Given that transcriptional regulation is mediated by complex, interconnected networks, removing one node from the network may lead to compensatory changes or propagation of downstream effects, which may mask the direct function of the TF in question. Inducible TF KOs or more acute chemical degradation strategies [66] may provide one solution to this problem. These may still be limited by need for development of reagents to cover all proteins of interest, portability of such systems to *in vivo* contexts, and specificity and efficiency of protein degradation [89]. However, evaluating function over time may be another necessary component for isolating the most proximal effects. The appropriate timescale may vary depending on whether the TF is presynthesized or induced only in response to some stimulus [4].

Second, TF KOs provide information outside the physiological range of TF function [89]. TFs operate within tightly controlled concentrations [8] with nonlinear responses to changes in TF occupancy. Instead of treating TF function as a binary variable, reading TF function across a range of finely controlled TF concentrations [60] may better reflect *in vivo* mechanisms, moving from qualitative effects to quantitative parameters (e.g., affinities and dissociation constants). Leveraging variation in *cis* (see Box 1) may be another complementary strategy in this regard.

Finally, parsing the relative contributions of different members of the same TF family or of TF paralogs remains an important challenge. Redundancy increases the rate of false negatives in TF KO experiments. This can partially be overcome by coupling KO and overexpression studies [91] and by multifamily member perturbations [92], but the caveats of nonphysiological concentrations and ability to scale perturbations remain.

Finally, as illustrated with advances in image recognition, or protein structure prediction by **AlphaFold**, advances in Artificial Intelligence (AI) may become increasingly important in rendering the complexity of immune TF networks tractable. Given their inherent ability to deal with highly complex inputs, **deep learning** and other novel AI approaches have already begun to tackle the *cis*-regulatory grammar of the genome at the sequence level, with spectacular results in several cases [73–75]: identifying new modes of TF cooperativity and binding for mammalian pluripotency TFs [74], engineering new promoter sequences [76], and capturing decades of mouse immunocyte gene regulation biology within a single model [77]. Such models may prove crucial to identify, in quantitative terms, multiple modes of TF function. Deep models might also uncover multiple types of sequence feature that are predictive of TF binding or motif interactions, and enable *in silico* predictions of regulatory changes [74,75]. Development of 'transparent' models that provide not only predictive capability, but also interpretable views into the underlying biological logic will be especially useful in informing our understanding of how the complexity of such networks is constructed from its constituent parts [75], how they are affected by genetic variation, and to enable their targeted modification.

Concluding remarks

Immunology has provided a fertile ground for uncovering basic mechanisms of TF function. New results make clear that TFs are not monomorphic but instead flexible in their functional capabilities, even within a cell type. TFs can further condition the function of other factors in response to external inputs, creating an interconnected loop between environmental stimuli, signaling cascades, TF regulation, and chromatin states. It is possible that the variegation is more marked in immunocytes, which must flexibly and rapidly adapt to changing environments, compared with in neurons, which operate more in fixed locations. The next frontier (see Outstanding questions) will be to combine advances in genome-wide measurements of gene regulatory function with computational advances in analyzing these functions, aiming to arrive at a complete, predictive understanding of immunocyte gene regulatory networks.

Outstanding questions

How can we move beyond binary descriptions of TF binding toward quantitative measurements of *in vivo* genome-wide affinities, and understand residence times of TFs at specific loci?

What are the direct targets of each TF?

What are the true binding preference (s) of each TF *in vivo*?

How does TF binding vary across single cells, or in a single cell over time?

Which sites are truly co-bound by multiple TFs and what are the relative dynamics of binding of each interaction, at single-regulatory element resolution?

Do single TFs, or even TF pairs, really exist, or do they mostly exist within multi-molecular complexes?

How can we develop a predictive and complete understanding of gene regulatory networks, for even one immunocyte?

How can new computational methods, such as deep learning and other AI tools, be best deployed to further understand immunologic gene regulation?

Acknowledgments

Citations in this opinion piece are not meant to be complete, but rather illustrative, and we apologize to colleagues whose ideas or discoveries are not credited as they should be. This work was funded by grants from the NIH to C.B. (AI150686, AI165697, and AI125603) and a grant from the JPB Foundation. K.C. was supported by NIGMS grants T32GM007753 and T32GM144273 and a Harvard Stem Cell Institute MD/PhD Training Fellowship.

Declaration of interests

None declared by authors.

References

1. Wasserman, W.W. *et al.* (2004) Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* 5, 276–287
2. Heinz, S. *et al.* (2015) The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* 16, 144–154
3. Troutman, T.D. *et al.* (2021) Exploiting dynamic enhancer landscapes to decode macrophage and microglia phenotypes in health and disease. *Mol. Cell* 81, 3888–3903
4. Pope, S.D. *et al.* (2018) Emerging principles of gene expression programs and their regulation. *Mol. Cell* 71, 389–397
5. Davis, R.L. *et al.* (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987–1000
6. Takahashi, K. *et al.* (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676
7. Vierbuchen, T. *et al.* (2010) Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035–1041
8. Lee, T.I. *et al.* (2013) Transcriptional regulation and its misregulation in disease. *Cell* 152, 1237–1251
9. Biggin, M.D. *et al.* (1989) Transcription factors and the control of *Drosophila* development. *Trends Genet.* 5, 377–383
10. Mitchell, P.J. *et al.* (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245, 371–378
11. Ferrie, J.J. *et al.* (2022) 'Structure'-function relationships in eukaryotic transcription factors: the role of intrinsically disordered regions in gene regulation. *Mol. Cell* 82, 3970–3984
12. Rudra, D. *et al.* (2012) Transcription factor Foxp3 and its protein partners form a complex regulatory network. *Nat. Immunol.* 13, 1010–1019
13. Kwon, H.K. *et al.* (2017) Different molecular complexes that mediate transcriptional induction and repression by FoxP3. *Nat. Immunol.* 18, 1238–1248
14. De Guzman, R.N. *et al.* (2004) Interaction of the TAZ1 domain of the CREB-binding protein with the activation domain of CITED2: regulation by competition between intrinsically unstructured ligands for non-identical binding sites. *J. Biol. Chem.* 279, 3042–3049

15. Wojciak, J.M. *et al.* (2009) Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. *EMBO J.* 28, 948–958
16. Gerstein, M.B. *et al.* (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91–100
17. Vierbuchen, T. *et al.* (2017) AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. *Mol. Cell* 68, 1067–1082
18. Zhao, X. *et al.* (2022) TCF1 in T cell immunity: a broadened frontier. *Nat. Rev. Immunol.* 22, 147–157
19. Gounari, F. *et al.* (2022) TCF-1: a maverick in T cell development and function. *Nat. Immunol.* 23, 671–678
20. Cirillo, L.A. *et al.* (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* 9, 279–289
21. Karagianni, P. *et al.* (2020) Bookmarking by non-pioneer transcription factors during liver development establishes competence for future gene activation. *Cell Rep.* 30, 1319–1328
22. Hansen, J.L. *et al.* (2022) A test of the pioneer factor hypothesis using ectopic liver gene activation. *eLife* 11, e73358
23. Kim, S. *et al.* (2023) Deciphering the multi-scale, quantitative cis-regulatory code. *Mol. Cell* 83, 373–392
24. Parviz, F. *et al.* (2003) Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat. Genet.* 34, 292–296
25. Michelson, D.A. *et al.* (2022) Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. *Cell* 185, 2542–2558
26. Sugimoto, N. *et al.* (2006) Foxp3-dependent and -independent molecules specific for CD25+CD4+ natural regulatory T cells revealed by DNA microarray analysis. *Int. Immunol.* 18, 1197–1209
27. Hill, J.A. *et al.* (2007) Foxp3 transcription-factor-dependent and -independent regulation of the regulatory T cell transcriptional signature. *Immunity* 27, 786–800
28. Lin, W. *et al.* (2007) Regulatory T cell development in the absence of functional Foxp3. *Nat. Immunol.* 8, 359–368
29. Ohkura, N. *et al.* (2012) T cell receptor stimulation-induced epigenetic changes and Foxp3 expression are independent and complementary events required for Treg cell development. *Immunity* 37, 785–799
30. Charbonnier, L.M. *et al.* (2019) Functional reprogramming of regulatory T cells in the absence of Foxp3. *Nat. Immunol.* 20, 1208–1219
31. Zemmour, D. *et al.* (2021) Single-cell analysis of FOXP3 deficiencies in humans and mice unmasks intrinsic and extrinsic $CD4^+$ T cell perturbations. *Nat. Immunol.* 22, 607–619
32. Chowdhary, K. *et al.* (2023) An interwoven network of transcription factors, with divergent influences from FoxP3, underlies Treg diversity. *bioRxiv* Published online May 21, 2023. https://doi.org/10.1101/2023.05.18.541358
33. Ciofani, M. *et al.* (2012) A validated regulatory network for Th17 cell specification. *Cell* 151, 289–303
34. Sungnak, W. *et al.* (2019) Multilayer regulation of CD4 T cell subset differentiation in the era of single cell genomics. *Adv. Immunol.* 141, 1–31
35. Yoshida, H. *et al.* (2019) The cis-regulatory atlas of the mouse immune system. *Cell* 176, 897–912
36. Mujal, A.M. *et al.* (2019) Immunity as a continuum of archetypes. *Science* 364, 28–29
37. Zemmour, D. *et al.* (2020) $CD4^+$ teff cell heterogeneity: the perspective from single-cell transcriptomics. *Curr. Opin. Immunol.* 63, 61–67
38. Zeng, H. (2022) What is a cell type and how to define it? *Cell* 185, 2739–2755
39. Zhong, Y. *et al.* (2022) Hierarchical regulation of the resting and activated T cell epigenome by major transcription factor families. *Nat. Immunol.* 23, 122–134
40. Sciumè, G. *et al.* (2020) Rapid enhancer remodeling and transcription factor repurposing enable high magnitude gene induction upon acute activation of NK cells. *Immunity* 53, 745–758
41. Muñoz-Rojas, A.R. *et al.* (2021) Tissue regulatory T cells: regulatory chameleons. *Nat. Rev. Immunol.* 21, 597–611
42. Zemmour, D. *et al.* (2018) Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* 19, 291–301
43. Campbell, D.J. *et al.* (2011) Phenotypical and functional specialization of FOXP3+ regulatory T cells. *Nat. Rev. Immunol.* 11, 119–130
44. Blei, D.M. *et al.* (2003) Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022
45. Bravo Gonzalez-Blas, C. *et al.* (2019) cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* 16, 397–400
46. Thanos, D. *et al.* (1995) Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091–1100
47. Chaudhri, V.K. *et al.* (2020) Charting the cis-regulome of activated B cells by coupling structural and functional genomics. *Nat. Immunol.* 21, 210–220
48. Samstein, R.M. *et al.* (2012) Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* 151, 153–166
49. Hosokawa, H. *et al.* (2018) Transcription factor PU.1 represses and activates gene expression in early T cells by redirecting partner transcription factor binding. *Immunity* 48, 1119–1134
50. van der Veeken, J. *et al.* (2020) The transcription factor Foxp3 shapes regulatory T cell identity by tuning the activity of trans-acting intermediaries. *Immunity* 53, 971–984
51. Ramirez, R.N. *et al.* (2022) FoxP3 associates with enhancer-promoter loops to regulate Treg–specific gene expression. *Sci. Immunol.* 7, eabj9836
52. Moore, L.D. *et al.* (2013) DNA methylation and its basic function. *Neuropsychopharmacology* 38, 23–38
53. Lio, C.J. *et al.* (2019) TET enzymes and 5hmC in adaptive and innate immune systems. *Front. Immunol.* 10, 210
54. Mattei, A.L. *et al.* (2022) DNA methylation: a historical perspective. *Trends Genet.* 38, 676–707
55. Roy, R. *et al.* (2022) DNA methylation signatures reveal that distinct combinations of transcription factors specify human immune cell epigenetic identity. *Immunity* 55, 1135
56. Kaluscha, S. *et al.* (2022) Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* 54, 1895–1906
57. Zeitlinger, J. (2020) Seven myths of how transcription factors read the cis-regulatory code. *Curr. Opin. Syst. Biol.* 23, 22–31
58. de la Rica, L. *et al.* (2013) PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol.* 14, R99
59. Frederick, M.A. *et al.* (2023) A pioneer factor locally opens compacted chromatin to enable targeted ATP-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* 30, 31–37
60. Naqvi, S. *et al.* (2023) Precise modulation of transcription factor levels reveals drivers of dosage sensitivity. *Nat. Genet.* 55, 841–851
61. Neikes, H.K. *et al.* (2023) Quantification of absolute transcription factor binding affinities in the native chromatin context using BANC-seq. *Nat. Biotechnol.* Published online March 27, 2023. https://doi.org/10.1038/s41587-023-01715-w
62. Teytelman, L. *et al.* (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18602–18607
63. Worsley, H.R. *et al.* (2014) Non-targeted transcription factors motifs are a systemic component of ChIP-seq datasets. *Genome Biol.* 15, 412
64. Macdonald, L. *et al.* (2022) Rapid and specific degradation of endogenous proteins in mouse models using auxin-inducible degrons. *eLife* 11, e77987
65. Bekes, M. *et al.* (2022) PROTAC targeted protein degraders: the past is prologue. *Nat. Rev. Drug Discov.* 21, 181–200
66. Jaeger, M.G. *et al.* (2021) Fast-acting chemical tools to delineate causality in transcriptional control. *Mol. Cell* 81, 1617–1630
67. Chen, W. *et al.* (2022) Live-seq enables temporal transcriptomic recording of single cells. *Nature* 608, 733–740
68. Simeonov, D.R. *et al.* (2017) Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* 549, 111–115
69. Freimer, J.W. *et al.* (2022) Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *Nat. Genet.* 54, 1133–1144

70. Kamimoto, K. *et al.* (2023) Dissecting cell identity via network inference and in silico gene perturbation. *Nature* 614, 742–751
71. Kartha, V.K. *et al.* (2022) Functional inference of gene regulation using single-cell multi–omics. *Cell Genom.* 2, 100166
72. González-Blas, C.B. *et al.* (2022) SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *bioRxiv* Published online August 19, 2022. https://doi.org/10.1101/2022.08.19.504505
73. Eraslan, G. *et al.* (2019) Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* 20, 389–403
74. Avsec, Z. *et al.* (2021) Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* 53, 354–366
75. Novakovsky, G. *et al.* (2023) Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nat. Rev. Genet.* 24, 125–137
76. Vaishnav, E.D. *et al.* (2022) The evolution, evolvability and engineering of gene regulatory DNA. *Nature* 603, 455–463
77. Maslova, A. *et al.* (2020) Deep learning of immune cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* 117, 25655–25666
78. Laufer, C. *et al.* (2013) Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat. Methods* 10, 427–431
79. Kemmeren, P. *et al.* (2014) Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 157, 740–752
80. Parnas, O. *et al.* (2015) A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* 162, 675–686
81. Dixit, A. *et al.* (2016) Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167, 1853–1866
82. Adamson, B. *et al.* (2016) A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 167, 1867–1882
83. Rubin, A.J. *et al.* (2019) Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* 176, 361–376
84. Taverner, N.V. *et al.* (2004) Identifying transcriptional targets. *Genome Biol.* 5, 210
85. Heinz, S. *et al.* (2013) Effect of natural genetic variation on enhancer selection and function. *Nature* 503, 487–492
86. Link, V.M. *et al.* (2018) Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. *Cell* 173, 1796–1809
87. van der Veeken, J. *et al.* (2019) Natural genetic variation reveals key features of epigenetic and transcriptional memory in virus-specific CD8 T cells. *Immunity* 50, 1202–1217
88. Keane, T.M. *et al.* (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289–294
89. Housden, B.E. *et al.* (2017) Loss-of-function genetic tools for animal models: cross-species and cross-platform differences. *Nat. Rev. Genet.* 18, 24–40
90. Johanson, T.M. *et al.* (2022) Parsing the transcription factors governing T cell immunity. *Nat. Immunol.* 23, 3–4
91. Carnevale, J. *et al.* (2022) RASA2 ablation in T cells boosts antigen sensitivity and long-term function. *Nature* 609, 174–182
92. Bruno, L. *et al.* (2019) Selective deployment of transcription factor paralogs with submaximal strength facilitates gene regulation in the immune system. *Nat. Immunol.* 20, 1372–1380
93. Rothenberg, E.V. (2021) Logic and lineage impacts on functional transcription factor deployment for T-cell fate commitment. *Biophys. J.* 120, 4162–4181