

RESEARCH ARTICLE

Systems-level identification of key transcription factors in immune cell specification

Cong Liu¹, Kyla Omilusik², Clara Toma², Nadia S. Kurd³, John T. Chang³, Ananda W. Goldrath², Wei Wang^{1,4*}

1 Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California, United States of America, **2** Division of Biological Sciences, University of California San Diego, La Jolla, California, United States of America, **3** Department of Medicine, University of California San Diego, La Jolla, California, United States of America, **4** Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California, United States of America

These authors contributed equally to this work.

* wei-wang@ucsd.edu



OPEN ACCESS

Citation: Liu C, Omilusik K, Toma C, Kurd NS, Chang JT, Goldrath AW, et al. (2022) Systems-level identification of key transcription factors in immune cell specification. *PLoS Comput Biol* 18(9): e1010116. <https://doi.org/10.1371/journal.pcbi.1010116>

Editor: Teresa M. Przytycka, National Center for Biotechnology Information (NCBI), UNITED STATES

Received: April 20, 2022

Accepted: August 10, 2022

Published: September 26, 2022

Copyright: © 2022 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This work was supported by the National Institutes of Health (R01AI150282 to WW; P01AI132122 to AWG and JTC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Transcription factors (TFs) are crucial for regulating cell differentiation during the development of the immune system. However, the key TFs for orchestrating the specification of distinct immune cells are not fully understood. Here, we integrated the transcriptomic and epigenomic measurements in 73 mouse and 61 human primary cell types, respectively, that span the immune cell differentiation pathways. We constructed the cell-type-specific transcriptional regulatory network and assessed the global importance of TFs based on the Taiji framework, which is a method we have previously developed that can infer the global impact of TFs using integrated transcriptomic and epigenetic data. Integrative analysis across cell types revealed putative driver TFs in cell lineage-specific differentiation in both mouse and human systems. We have also identified TF combinations that play important roles in specific developmental stages. Furthermore, we validated the functions of predicted novel TFs in murine CD8⁺ T cell differentiation and showed the importance of Elf1 and Prdm9 in the effector versus memory T cell fate specification and Kdm2b and Tet3 in promoting differentiation of CD8⁺ tissue resident memory (Trm) cells, validating the approach. Thus, we have developed a bioinformatic approach that provides a global picture of the regulatory mechanisms that govern cellular differentiation in the immune system and aids the discovery of novel mechanisms in cell fate decisions.

Author summary

The immune system protects the human body from invading pathogens and tumors. Enormous efforts have been devoted to developing preventative and therapeutic treatments based on exploitation of diverse immune cell types. Recent technological advances have made it possible to study the regulatory mechanisms at the systems level. However, the noise in the genomic measurements and the complexity of integrating multiple omics

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: A.W.G is on the SAB of ArsenalBio.

assays are still challenging. To tackle the challenges, we constructed the cell-type-specific transcriptional regulatory network for 73 mouse and 61 human immune cell types and assessed the global importance of transcription factors (TFs) using a systems biology approach called Taiji framework. We revealed driver TFs in cell lineage-specific differentiation. Furthermore, we validated the functions of predicted novel TFs in CD8⁺ T cell differentiation and showed the importance of Elf1 and Prdm9 in the effector versus memory T cell fate specification and Kdm2b and Tet3 in promoting differentiation of CD8⁺ tissue resident memory (Trm) cells, validating the approach. Our study provides a comprehensive identification of driver regulators directing immune cell development, provides a global picture of the regulatory mechanisms that govern cellular differentiation in the immune system and aids the discovery of novel mechanisms in cell fate decisions.

Introduction

The immune system protects the human body by eliminating invading pathogens as well as tumors [1,2]. Considerable effort has been devoted to developing preventative and therapeutic treatments based on exploitation of a diverse set of immune cell types. Establishing immune cell-based clinical treatments relies on research advances that reveal the fundamental mechanisms regulating the immune system, such as those that direct the differentiation from progenitor to mature immune cells and coordinate their response to the presence of various antigens. Recent technological advances have made it possible to study the regulatory mechanisms at the systems level and genomic scale [3]. However, the noise in the genomic measurements and the complexity of integrating multiple omics assays pose a great challenge of fully digesting the information and extracting the most informative signals from the extensive data.

A critical step towards achieving the goal of studying the regulatory mechanism is to uncover the key regulators that direct the specification of immune cell states. Transcription factors (TFs) are essential regulators of cell fates [4,5] and play pivotal roles in immune cell development [6]. Identification of critical TFs that drive immune cell differentiation can also provide key mechanistic insights into immune cell functions. However, a complete catalog of critical TFs in each immune cell population or in each lineage has not been established. Therefore, dissection of the genetic networks that promote immune cell differentiation and function and characterization of the genome-wide influences of the key regulators would provide novel insights into cell specification and potential strategies to impact immune function.

The global influence of a cellular regulator is conveyed through its regulatory effect on target genes, which is subsequently propagated over the genetic network. A given regulator's activity is not only affected by its own expression level but also post-translational modifications, the presence of collaborative co-factors and its targets' accessibility. Therefore, the expression level of a regulator such as a TF is not always correlated with its activity [7]. In light of this, many methods have been proposed to infer the activity of regulators using statistical or machine learning approaches. For instance, Schacht et al. [7] developed a statistical model to estimate the regulatory activity of TFs using their cumulative effects on their target genes. Arrieta-Ortiz et al. [8] used a linear model to infer the TF activity (TFA) by predicting target genes' expression levels. SCENIC [9] developed by Aibar et al. constructs the genetic network by predicting each gene's expression using the TF's expression levels and finding the most predictive TFs as the key regulators. Maslova et al. introduced AI-TAC [10] to predict ATAC-seq signals and identified the most enriched motifs when evaluating the TF importance. Although these methods were able to predict the local activity of a TF, i.e. the expression level of their

direct target genes, measuring the system-wide influence of a given TF was not their focus. As genes rarely function alone but usually cross-talk with each other to form complex regulatory logic, we argue that the global influence of a regulator should, in principle, better predict the cell state change upon perturbing the regulator than its local influence alone [11,12].

Recently, we have developed a method, called Taiji, that can successfully infer the global impact of TFs in the genetic network constructed from integrating epigenomic and gene expression data [13–15]. Focusing on the global rather than the local importance of transcriptional regulators makes Taiji robust and suitable for integrating multiomics data from noisy genomic measurements. In fact, Taiji clearly outperforms the motif enrichment analysis and the TFA approach [13–15]. In addition, several TFs predicted to be critical in regulating the differentiation of CD8⁺ T cells following infection were validated experimentally and demonstrated to have previously unappreciated roles in cell fate specification [14,15].

Here, we performed a comprehensive analysis on the common regulatory code in the mouse and human immune systems using Taiji. We identified key TFs critical for programming the differentiation of nine cell lineages that span diverse cell types of the immune system, thus providing the first comprehensive catalog of key regulators for a variety of cell lineages during immune system development. In addition, we uncovered TF combinations that activate in a spatiotemporal manner, behaving like transcriptional waves to orchestrate the developmental progress and cell lineage specification. We show the great predictive performance of Taiji by validating the functions of four novel TFs in murine CD8⁺ T cell differentiation and memory formation.

Results

Using assay of transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) and gene expression (RNA-seq) data for 73 mouse immune cell populations generated by the Immunological Genome Project [16], representing 9 cell types including stem cells, dendritic cells (DC), macrophages/neutrophils (MF/GN), monocytes (Mo), B cells, innate lymphoid cells (ILC), $\alpha\beta$ T cells (abT), $\gamma\delta$ T cells (gdT), and activated T cells (Act_T), we performed an integrated analysis using the Taiji pipeline [13]. Briefly, Taiji integrates gene expression and epigenetic modification data to build gene regulatory networks. Taiji first scans putative TF binding sites in each open chromatin region using motifs documented in the CIS-BP database [17], to identify active regulatory elements, including active promoters or enhancers. These TFs are then linked to their target genes as predicted by EpiTensor [18]. All of the regulatory interactions are assembled into a genetic network. Lastly, the personalized PageRank algorithm is used to assess the global influences of the TFs [19]. In the network, the node weights were determined by the z scores of gene expression levels, allocating higher ranks to the TFs that regulate more differentially expressed genes. The edge weights were set to be proportional to TFs' expression levels, the binding site peak intensity, and the motif binding affinity, thus representing the regulatory strength (Fig 1). The details of the weighting scheme are available in the “Materials and Methods” section. The superior performance of Taiji compared to other methods to identify key regulators have been confirmed using simulated data, literature evidence and experimental validations [13,15]. We constructed a network in each mouse cell type. The average number of nodes and edges of the networks were 19,340 and 773,653, respectively, including 815 (4.21%) TF nodes. On average, each TF was predicted to regulate 3996 genes, and each gene was predicted to regulate 88 TFs (S1 Fig).

PageRank score is a great indicator of cell specificity

To characterize the potential global influences of all the TFs across different cell types, we plotted the PageRank score matrix with the columns sorted by the nine cell lineages (Fig 2). We

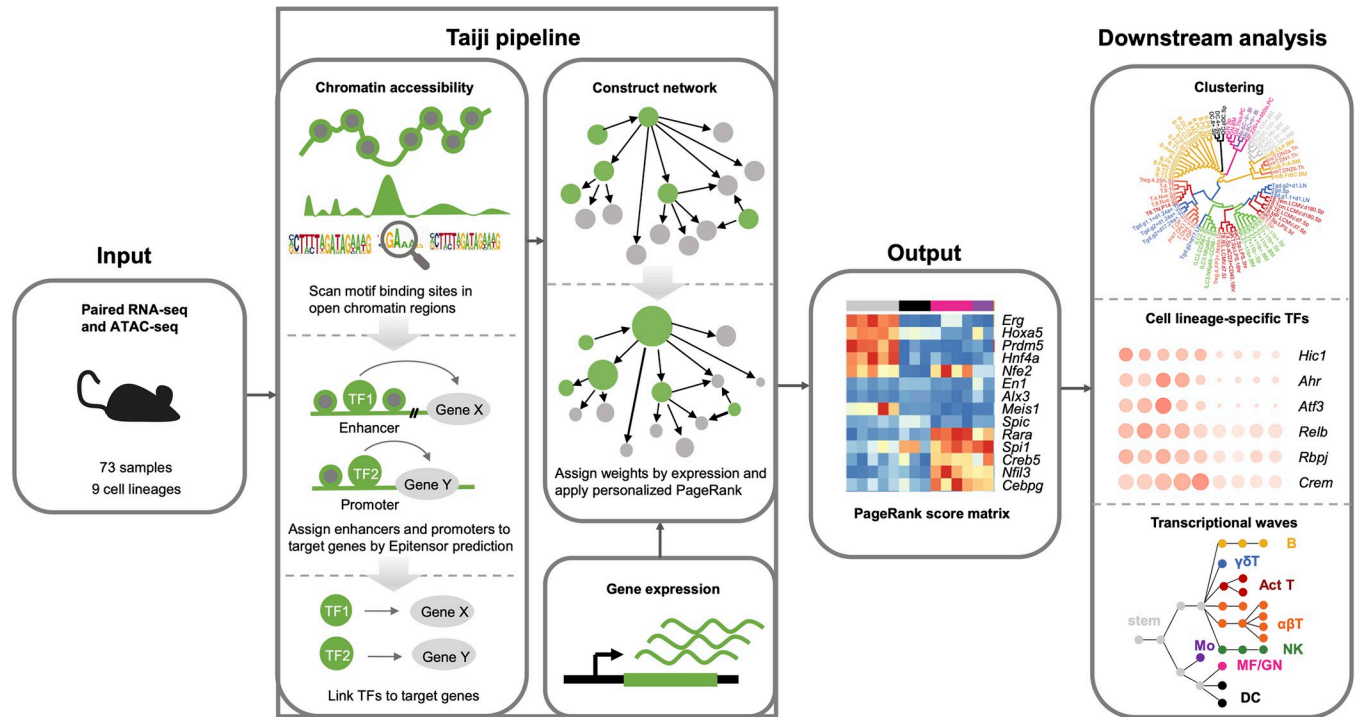


Fig 1. The workflow of study and Taiji pipeline. Matched RNA-seq and ATAC-seq data were used as input for the Taiji framework to construct a regulatory network and generate PageRank scores matrix as output (column colors code for different cell lineages, see Fig 2 for the color annotation). Downstream analyses included clustering samples to distinct cell lineages, identification of cell lineage-specific TFs, and construction of temporal transcriptional waves.

<https://doi.org/10.1371/journal.pcbi.1010116.g001>

found that different cell lineages show quite distinct TF regulatory patterns, illustrating that transcriptional regulation takes place in a cell-lineage-specific fashion. The row-wise comparison demonstrates that some TFs are specific to one particular cell lineage. For example, *Bcl11a* is essential for the survival of early B cells [20,21] and *Irf4* has been shown as a key transcription factor that directs the development and maturation of B cells including pre-B differentiation and marginal zone B cell development [22–24]. Both *Bcl11a* and *Irf4* have exclusively higher PageRank scores in B cell lineage compared to the other immune lineages. Another example is that both *Tcf7* and *Lef1* have known functions in T cell development and differentiation [25–27] and they display higher PageRank scores across T cells compared to the average.

Identification of putative driver TFs in cell lineage differentiation

We then identified 48 constitutively active TFs whose average PageRank scores across all the cell types were ranked top 10% with CVs<0.5, indicating their predictive activities across all lineages (Fig 3A). Similar to their transcriptional activities, the expression levels of these TFs remained relatively high in immune cells from different tissues and at different stages of development (S2 Fig, the circle size represents the expression levels while the color represents the PageRank scores). Functional analysis revealed that these TFs are enriched in house-keeping functions, including “biological regulation”, “metabolic process”, and “cellular process”, demonstrating the general roles of active TFs in basic cellular functions across all cell types. Notably, more than one third of these active TFs are related to the immune system processes, which is not surprising.

Furthermore, we identified potential important TFs in 73 cell types from 9 cell lineages by comparing their PageRank scores between a specific lineage and the background (see

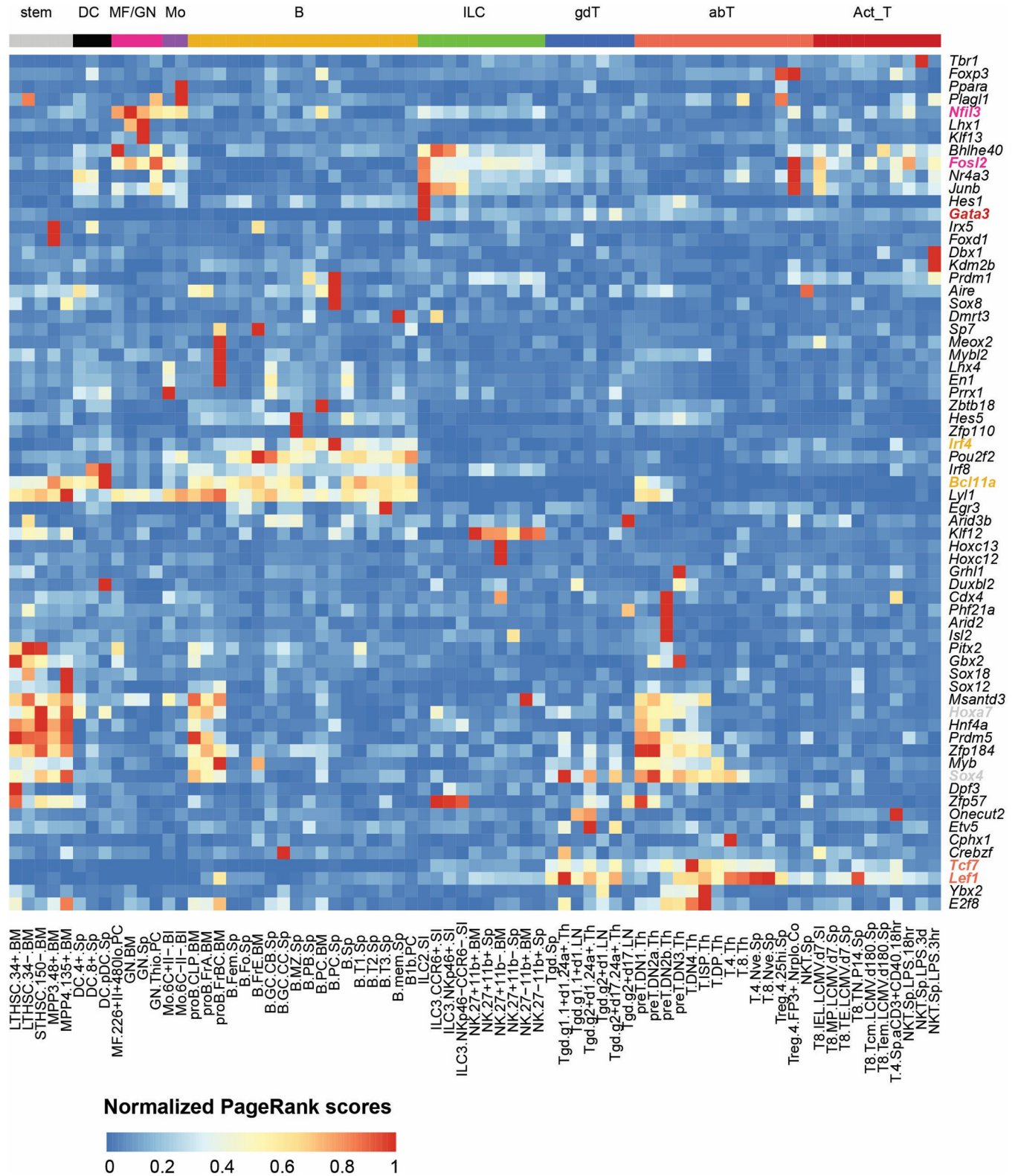
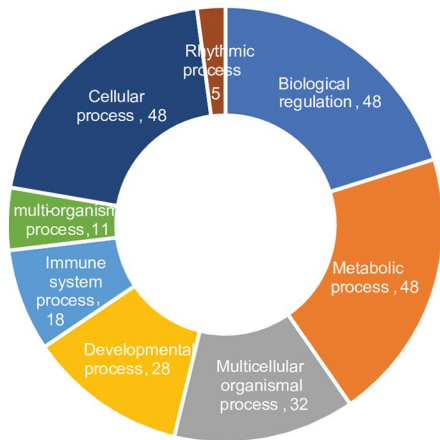


Fig 2. PageRank score is a great indicator of cell specificity. PageRank scores heatmap of 59 cell-specific TFs across 73 cell types. TFs in rows (z-normalized), cell types in columns, and color of the cell in the matrix indicates the normalized PageRank scores with red displaying high scores. Selected gene names are colored based on their cell specificity. abT, $\alpha\beta$ T cells; Act_T, activated T cells; B, B cells; DC, dendritic cells; gdT, $\gamma\delta$ T cells; ILC, innate lymphoid cells; MF, macrophages; GN, neutrophils; Mo, monocytes; stem, stem cells.

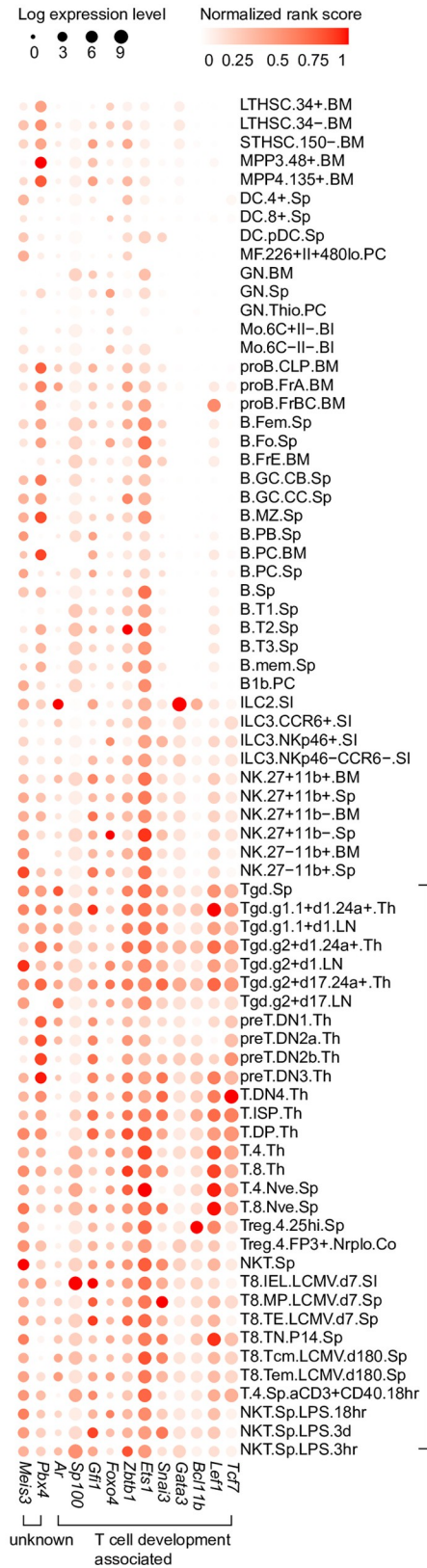
<https://doi.org/10.1371/journal.pcbi.1010116.g002>

a



Rela, Zfx, Sp3, Sp1, Zbtb17, Elf1, Gabpa, Atf2, Irf3, Sp2, Fli1, Elk4, Usf1, Foxj3, Elf2, Zfp263, Nrf1, E2f4, Zfp398, Nfyc, Nfya, Zfp260, Arnt, Ikzf1, Thap11, Mlx, Foxn3, Ctcf, Maz, Sp4, Pknox1, Elk3, Atf7, Zfp281, Klf7, Zfp212, Stat3, Zfp566, Zfp287, Stat6, Klf3, Sp110, Irf2, Smad3, Trp53, Irf1, Klf10, Patz1, Etv6

b



c

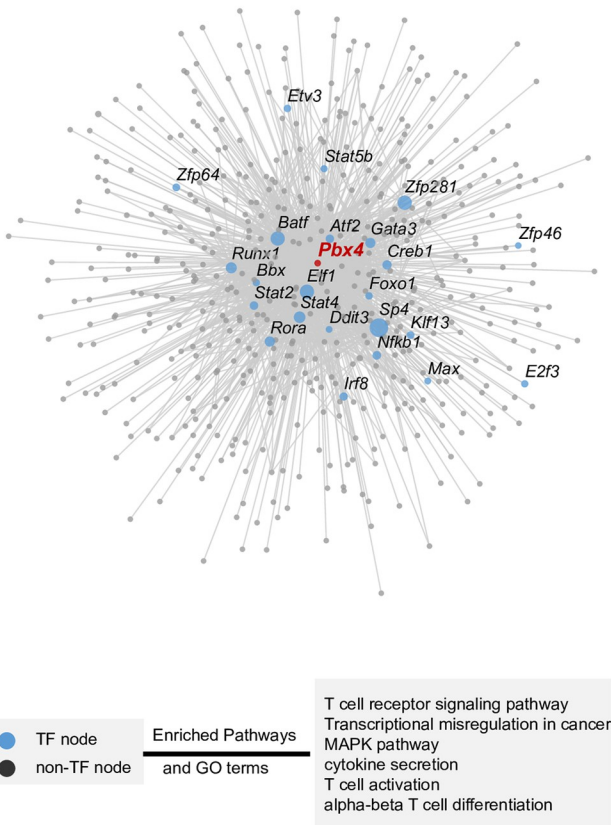


Fig 3. Identification of putative driver TFs in cell lineage specification during mouse immune system development. (A) Functional classification of 48 constitutively active TFs reveals their functions in essential biological processes. (B) Identified putative driver TFs in T cell lineage, including eleven TFs known to be related to T cell development, and two TFs with unknown functions. Circle size represents the logarithm of gene expression, and the color represents the normalized PageRank score. (C) The network for *Pbx4*, a previously unknown putative driver TF in T cell lineage, and its regulatees in the T8. IEL.LCMV.d7.SI. Each node represents one regulatee of *Pbx4* with TF highlighted in blue. Larger node size represents a higher PageRank score in the network. The bottom box shows the four enriched Gene Ontology (GO) terms and KEGG pathway for *Pbx4*'s regulatees.

<https://doi.org/10.1371/journal.pcbi.1010116.g003>

“Methods” for details). To our knowledge, this is the first systematic mapping of the complex transcriptional regulation that occurs during mouse immune system development (S3 Fig). The identified putative driver TFs include many well-known key regulators. For instance, 11 of 13 identified T cell-specific TFs have been previously reported to play a pivotal role during the T cell development and differentiation in the mouse immune system [28–30] (Fig 3B), such as *Lef1* and *Tcf7* that are critical for CD8⁺ T cell memory formation and CD4⁺ T cell differentiation [31–35], and *Bcl11b*, an essential TF for the development of $\alpha\beta$ T cells and most $\gamma\delta$ T cells [36].

To systematically validate the predictions of these analyses, we performed literature search in 5 lymphoid hematopoietic lineages that have been extensively studied, including the $\alpha\beta$ T cells, Act_T cells, B cells, $\gamma\delta$ T cells, and ILC. Forty-two of 50 (84%) TFs identified by our method are shown to be associated with regulation of development or normal functions of these cell lineages (S2 Table). For example, *Tcf4* is a well-known transcriptional activator in B cell differentiation and regulates the expression of genes critical for germinal center (GC) B cell development [37–40].

The lineage specificity was decided by assessing whether the TF PageRank scores are statistically significantly higher in one lineage than the other lineages, i.e. the p-value smaller than 0.001 and the log₂ fold change larger than 1. It is possible that some TFs are important in several cell lineages. Take *Bcl11a* as an example, it exhibits high PageRank scores in several cell lineages: stem, DC cells and B cells, and its role in all these three lineages have been previously described [20,21,41–43]. We found that several putative driver TFs are highly conserved in three T cell subsets. *Lef1*, *Gata3*, and *Bcl11b* were ranked top 10 in all 3 T cell sub-lineages, demonstrating that interplay and co-expression of these regulators are essential for the development and differentiation of several T cell subsets including CD4⁺ and CD8⁺ T cells.

We have also identified TFs with previously undefined functions in immune cell differentiation. For example, *Pbx4*, identified as a key regulator for T cell development, has not yet been studied. We predicted that *Pbx4* regulates a number of high-rank TFs, including several aforementioned constitutively active TFs, i.e., *Sp4*, *Elf1*, and *Zfp281* (Fig 3C). The functional enrichment analysis of *Pbx4*'s regulatees suggests that they participate in T cell receptor signaling pathway, T cell activation, and $\alpha\beta$ T cell differentiation (Fig 3C).

During development, hematopoietic stem cells (HSCs) differentiate into two different types of immune cells—myeloid and lymphoid lineages. Myeloid cells include Mo, DC, neutrophil, GN, and MF while lymphocytes include ILC, T and B lineages. T cell lineage can be further divided into abT, gdT, and Act_T. To identify TFs that are specific to each cell lineage, we first grouped and compared samples originated from myeloid lineage against samples from lymphoid lineage. The student t tests with a P value cutoff of 0.005 and log₂ fold change cutoff of 0.5 were used to identify TFs whose ranks change significantly in myeloid lineage compared with lymphoid lineage and vice versa, thus obtaining a list of myeloid-specific TFs and lymphoid-specific TFs (Fig 4). The functional relevance of the predicted specific TFs is supported by previous studies. For example, we have identified *Hhex* and *Spi1* as putative specific regulators for myeloid cells, which are known to be associated with myeloid leukocyte differentiation [44,45]. The enriched Gene Ontology (GO) terms showed that the identified lymphoid-specific

TFs are involved in T cell receptor V(D) J recombination, B cell lineage commitment, and cellular response to interleukin-4.

To analyze the finer cell lineage specificity of the predicted putative driver TFs, we further compared each sub-cell lineage with other cell lineages originated from the same progenitors and used the student t tests with a P value cutoff of 0.005 and \log_2 fold change cutoff of 0.5 to identify cell lineage-specific putative driver TFs. In total, 241 cell lineage-specific putative driver TFs were found and the top 10 TFs are shown in Fig 4. GO terms and pathway enrichment analysis is performed for each group of cell lineage-specific putative driver TFs (S4 Table). Many well-known cell lineage-specific markers appeared in the list, e.g., *Fosl2* [46] and *Runx3* [47,48] for NK cells and *Bcl11a* [43] for DC. In addition, a number of novel TFs were also found from our analysis, such as *Zfp296* for B cell, *Jdp2* for NK cell and *Ar* for Act T cells. Taken together, these results provide a comprehensive map for the future mechanistic study of mouse immune system development.

Transcriptional waves during immune cell development

In addition to the tissue specificity, we have also analyzed the dynamic activity of TFs during immune cell development. To identify the TFs which show similar temporal activity patterns, we clustered the TFs based on the normalized PageRank scores across samples. First of all, we performed the principal component analysis (PCA) for dimension reduction of the TF score matrix. We retained the first 30 principal components for further clustering analysis, which explained 75% variance (S4 Fig). We used the k-means algorithm for clustering analysis. To find the optimal number of clusters and similarity metric, we performed the Silhouette analysis to evaluate the clustering quality using five distance metrics: Euclidean distance, Manhattan distance, Kendall correlation, Pearson correlation, and Spearman correlation (S5 Fig). Pearson correlation was the most appropriate distance metric since the average Silhouette width was the highest among the five distance metrics. Based on these analyses, we identified 40 distinct dynamic patterns of TF activity during immune cell development (S6 Fig).

The 40 clusters represent transcriptional waves that orchestrate tissue differentiation. Some TFs are predicted to be active throughout many developmental stages, represented by cluster C21 and C29 (Fig 5A and 5B). In both clusters, TFs exhibit relatively high PageRank scores and show active activity across almost all stages. Example TFs include constitutively active genes identified earlier, such as *Sp1* [49], *Zfx* [50], and *E2f4* [51], which are essential for immune system development and maintenance of normal functions of various immune cells.

We also found clusters that include genes responsible for the differentiation of specific cell lineages. For example, C7 (Fig 5C) is particularly active in DC samples, suggesting its critical role in DC cell differentiation. Many TFs that are crucial for DC cell development were recovered in C7, such as *Stat1* [52], *Irf7* [53], and *Batf* [54]. Another instance of cell lineage-specific clusters is C27 (Fig 5D), C27 is exclusively active in stem cells, particularly in 3 HSCs (LTHSC.34-.BM, LTHSC.34+.BM, STHSC.150-.BM,). TFs in this cluster are associated with functions specific to early development of HSC including ten *Hox* [55] family TFs and two *Sox* [56] family TFs.

In addition to cell lineage-specific transcriptional waves, there are also developmental stage-specific transcriptional programs that appear to drive the differentiation of individual stages. We highlighted 4 clusters as examples, C15, C38, C10, and C12, which are potentially responsible for the B.FrE.BM, preT.DN3.Th, T8.Tem.LCMV.d180.Sp, and ILC2.SI development, respectively (Fig 5E–5H). TFs in these clusters exhibit transient transcriptional spikes, likely related to their stage-specific functions, thus activating at different stages in immune system development. C38 includes three *Duxbl* family members (*Duxbl1*, *Duxbl2*, and *Duxbl3*),

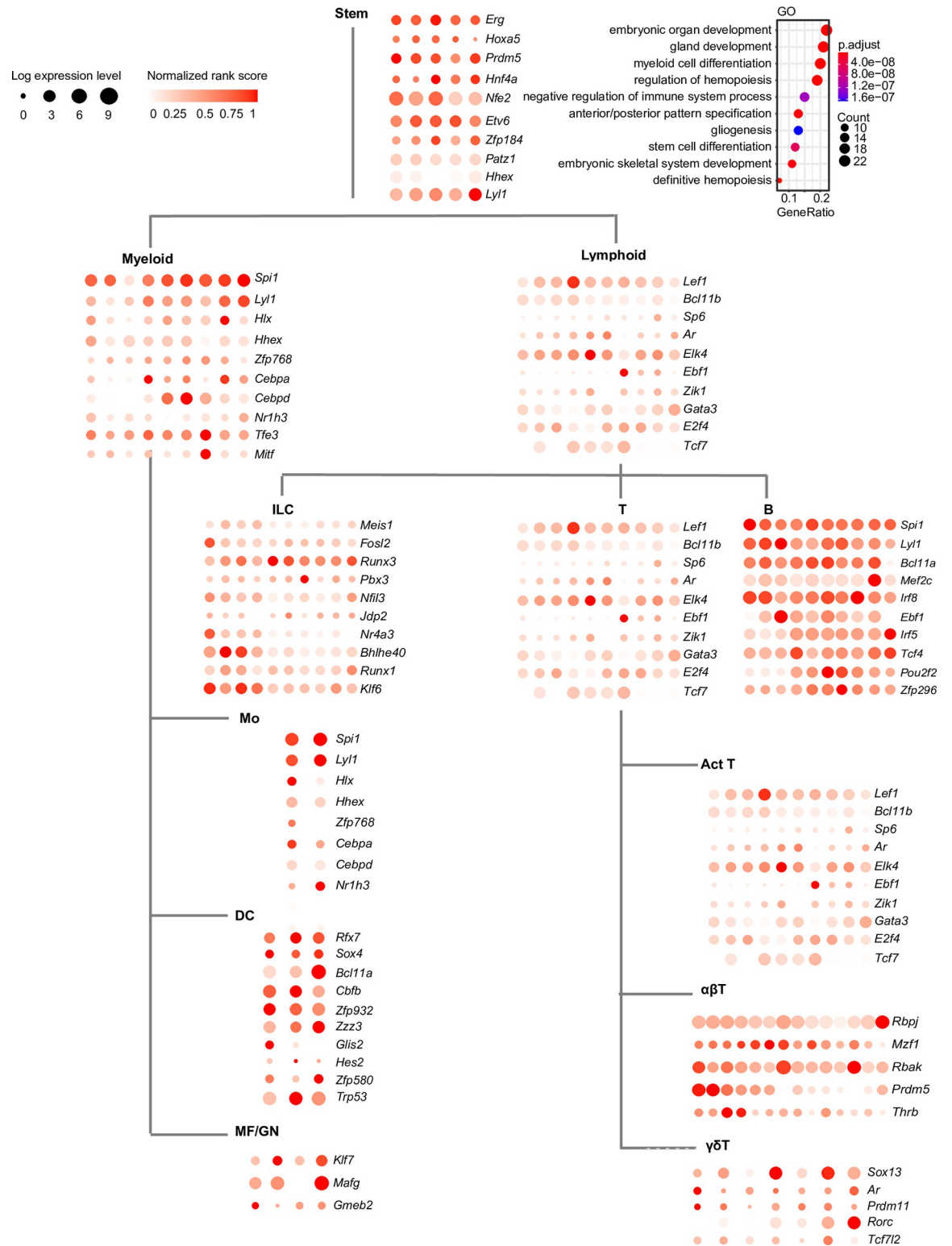


Fig 4. Cell lineage-specific putative driver TFs in the mouse immune system. A P-value of 0.005 and log2 fold change of 0.5 were used to call the putative driver TFs in each cell lineage. The bubble plots show the top 10 putative driver TFs for “Lymphoid” and “T” lineage, only “Act_T” samples are shown as example due to limited space. The GO term enrichment analysis for stem cells is shown as an example.

<https://doi.org/10.1371/journal.pcbi.1010116.g004>

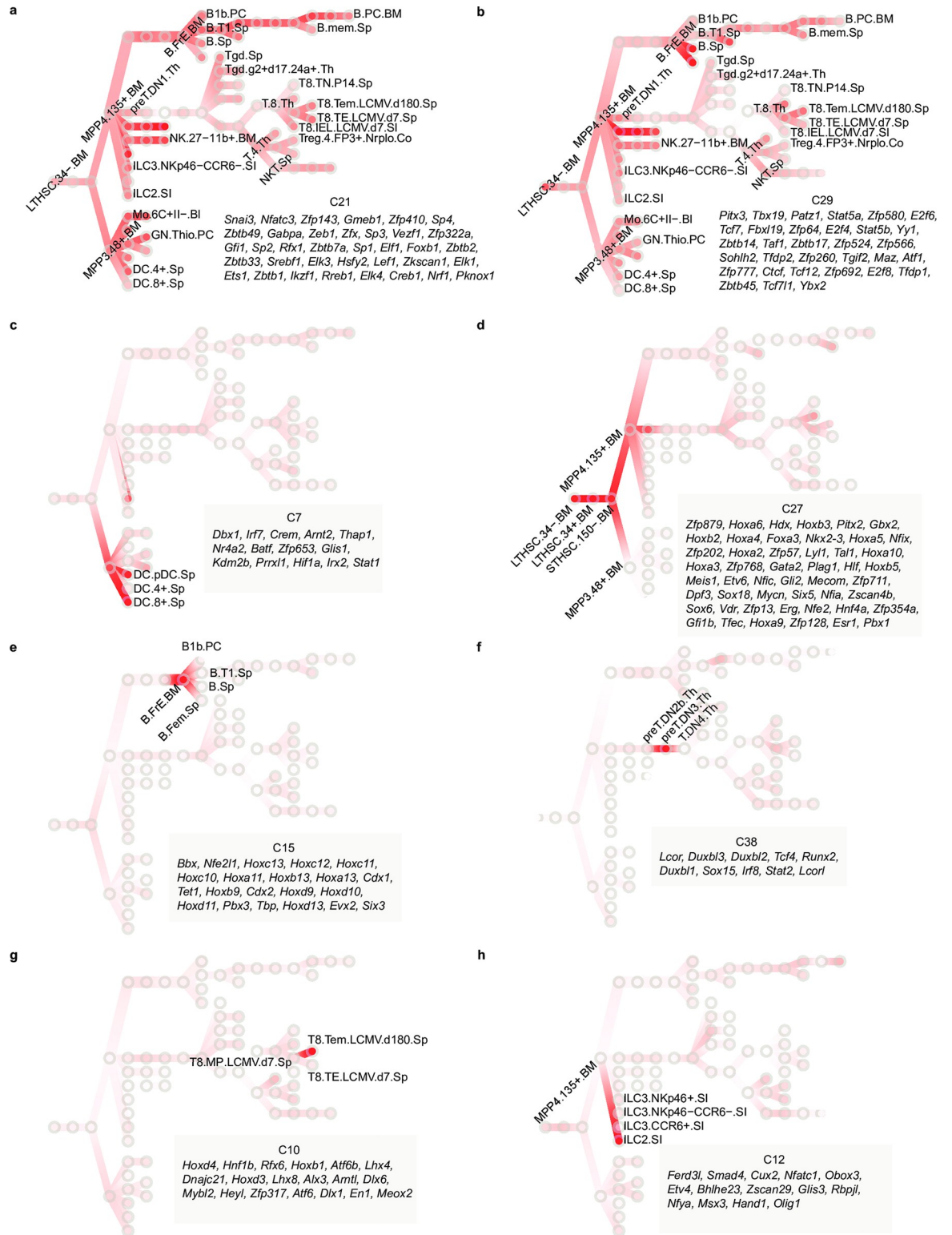


Fig 5. Transcriptional waves regulate the development of the mouse immune system. (A,B) Two transcriptional waves C21 and C29, which are active in all developmental stages. (C,D) Two cell lineage-specific transcriptional waves. C7 is active in DC lineage and C27 is active in stem cells. (E-H) Four examples of stage-specific waves in immune cell development. C15 is active in B.FrE.BM; C38 is active in pre.T.DN3.Th; C10 is active in T8.Tem.LCMV.d180.Sp; C12 is active in ILC2.SI. Circles in each panel represent specific cells in the developmental stages. Color indicates normalized PageRank scores with red displaying high values (color bar is the same as Fig 4). TF members of the cluster are shown on the bottom right of the panel.

<https://doi.org/10.1371/journal.pcbi.1010116.g005>

which mediate elimination of pre-T cells that fail β -selection [57]. Several *Hox* family TFs (*Hoxd4*, *Hoxb1*, *Hoxd3*) in C10 show significantly higher Pagerank score in CD8⁺ effector memory T cell (T_{EM}) than that in others, suggesting their putative role in regulating T_{EM} development. Although previous evidence shows that *Hoxb4* overexpression can increase the central memory population in CD4⁺ T cell [58], the mechanism of *Hox* family TFs in CD8⁺ T cell has not been thoroughly explored. As the mechanism of immune system development remains to be completely defined, our discovery, therefore, provides an invaluable resource for future studies.

Taiji identifies key TFs important for differentiation of effector and memory CD8⁺ T cell populations

Memory cells that mount a rapid and protective response upon reinfection are the fundamental basis for vaccines, long-term immunity and health [59]. To examine key regulators of memory formation, we further investigated TFs that are important for directing differentiation of short-lived effector versus long-lived memory cells. We identified putative driver TFs in 3 memory cell populations (B.mem.Sp, T8.Tcm.LCMV.d180.Sp, T8.Tem.LCMV.d180.Sp) with all other cells as background following the similar procedure to the identification of cell lineage-specific TFs (see [Methods](#)). We then performed literature search on the top 10 putative driver TFs ([S3 Table](#)) and found that half of them are well-known regulators of the differentiation and maintenance of normal functions of memory T cells and B cells, such as *Foxo1* [60–67], *Irf3* [68,69], and *Ets1* [70], while others remain unknown and warrant a more thorough investigation, such as *Cphx1*, *Pou4f1*, and *Osr2*.

We selected 2 TFs predicted by Taiji to regulate CD8⁺ T cell differentiation with previously unknown functions in CD8⁺ T cell differentiation and memory formation—*Elf1* and *Prdm9*—to further validate our analysis. Congenically distinct P14 T cells (expressing a TCR specific for the LCMV glycoprotein gp33) were transduced with retroviral vector encoding hairpin RNAs (shRNAs) targeting our gene of interest (*Elf1* or *Prdm9*) or a negative control (*Cd19*), mixed at a 1:1 ratio then transferred into recipients that were subsequently infected with LCMV-Armstrong (Figs [6A](#) and [S7A](#)). On day 7 of infection, the effector P14 T cell populations in the spleen that were deficient for *Elf1* or *Prdm9* displayed an increased frequency of memory precursor (KLRG1⁻CD127⁺) cells compared to the control population (Fig [6B](#) and [6C](#)). Furthermore, at day 21 of infection, the memory T cell population was comprised of an increased proportion of T_{EM} (CD62L⁻CD127⁺) when *Elf1* expression was knocked down and an increased frequency T_{CM} (CD62L⁺CD127⁺) and a decreased frequency of t-T_{EM} (CD62L⁻CD127⁻) when *Prdm9* was knocked down (Fig [6B](#) and [6C](#)). These data validate the functions of *Elf1* and *Prdm9* in CD8⁺ T cell differentiation and suggest a role for these TFs in promoting terminal differentiation.

Tissue-resident lymphocytes are a diverse population of non-recirculating lymphocytes lodged in tissues that provide localized protective immunity and immunosurveillance [71]. As the importance of this population to host protection is clear, we next sought to define regulators important for tissue residence. To identify putative tissue-residency-specific TFs, we focused on 6 lymphocyte populations originating from the small intestine (ILC3.NKp46+.SI, ILC3.NKp46-CCR6-.SI, ILC3.CCR+.SI, ILC2.SI, Treg.4.FP3+.Nrpl.Co, T8.IEL.LCMV.d7.SI) with 8 similar cells in the circulation as background (T8.Tcm.LCMV.d180.Sp, T8.Tem.LCMV.d180.Sp, T8.MP.LCMV.d7.Sp, T8.TE.LCMV.d7.Sp, Treg.4.25hi.Sp, NK.27+11b-.Sp, NK.27+11b+.Sp, NK.27-11b+.Sp). More than 60% of the identified TFs have been identified as regulators within the tissue-resident lymphocyte populations by previous studies ([S3 Table](#)). For instance, *Hic1* is critical for establishing and maintaining T_{RM} cells in the gut [72–74], and *Ahr* is required for long term persistence of T_{RM} in the epidermis [75].

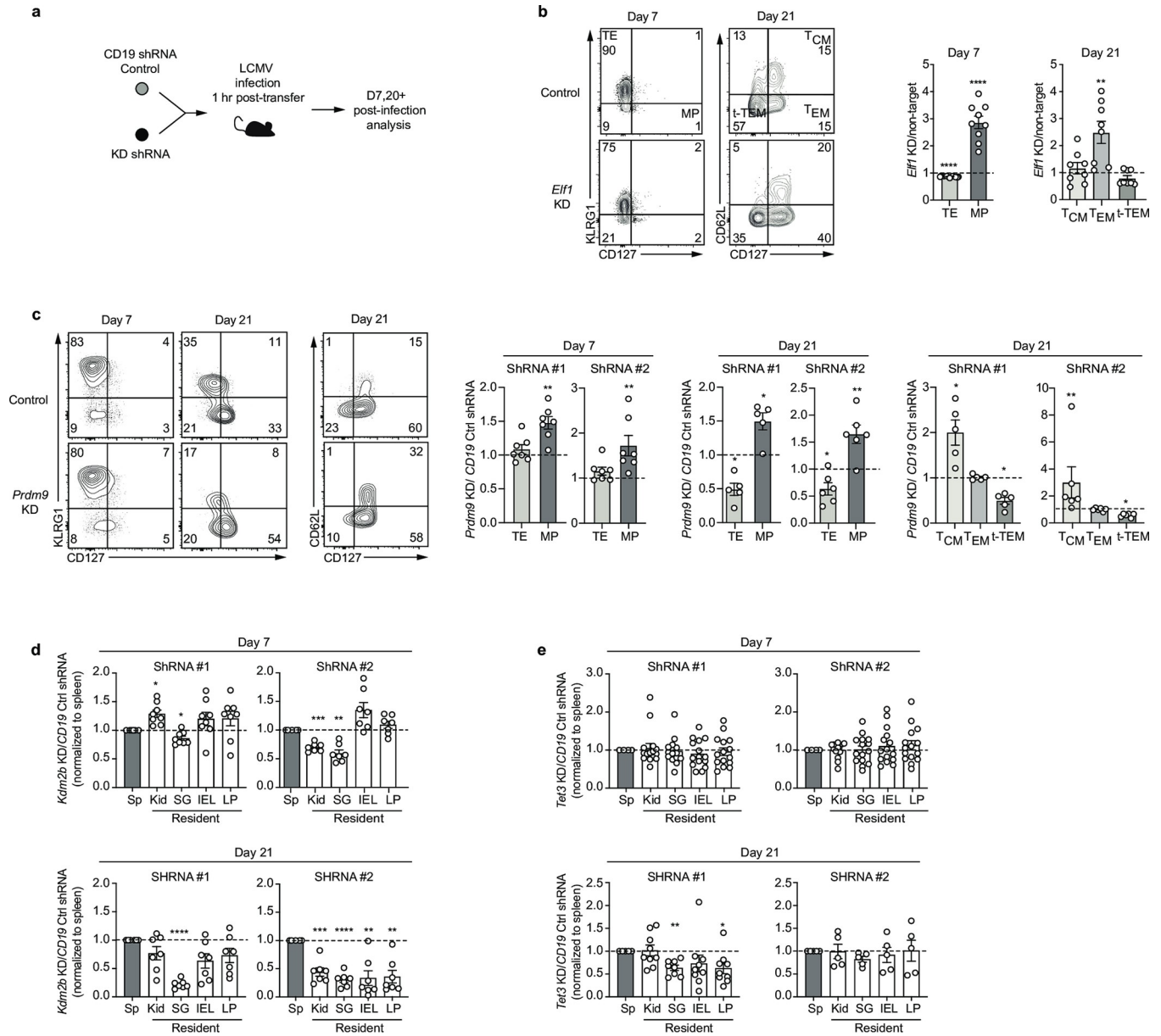


Fig 6. Validation of novel TFs' roles in CD8⁺ T cell differentiation. (A) Experimental schematic in which congenically distinct P14 were transduced with gene-specific (knock-down, KD) or *Cd19* control shRNA encoding retroviruses, mixed 1:1 then transferred into recipient mice subsequently infected with LCMV. (B) The ratio of TE (KLRG1^{hi}CD127^{lo}) or MP (KLRG1^{lo}CD127^{hi}) and T_{CM} (CD62L^{hi}CD127^{hi}), T_{EM} (CD62L^{lo}CD127^{hi}) or t-T_{EM} (CD62L^{lo}CD127^{lo}) (gating strategy, left) within the *Eif1* KD and control P14 population recovered from the spleen on day 7 (middle) or day 21–27 (right) of infection. (C) Representative flow cytometry staining (left) and quantification of the ratio of TE and MP and T_{CM}, T_{EM} or t-T_{EM} (right) within the *Prdm9* KD and control P14 population recovered from the spleen on day 7 or day 21. (D) The ratio of recovered *Kdm2b* KD and control P14 cells in indicated tissues on day 7 (top) and 21 (bottom) of infection. (E) The ratio of recovered *Tet3* KD and control P14 cells in indicated tissues on day 7 (top) and 21 (bottom) of infection. Data are cumulative of 2–3 experiments with n = 3–5 per group. Graphs show mean ± SEM. A one-sample two-tailed t-test, two-tailed ratio paired t-test (B,C) or a one-way paired ANOVA test (D,E) was used to determine statistical significance; *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

<https://doi.org/10.1371/journal.pcbi.1010116.g006>

Using this approach, we also identified several genes with undefined function in tissue residency. We validated the role of 2 of these genes, *Kdm2b* and *Tet3*, in promoting differentiation of CD8⁺ T_{RM}. *Kdm2b* and *Tet3* were both documented in the CIS-BP database to recognize specific DNA motifs and were thus included in our study. We selected these two chromatin

remodeling factors because epigenetic modification plays important roles in cell differentiation and they have not been reported to function in CD8⁺ T cell differentiation. Their available motifs and identified importance from the Taiji analysis gave us an opportunity to study their importance in CD8⁺ T cell differentiation for the first time. As above, congenically distinct P14 T cells expressing shRNAs for our gene of interest (*Kdm2b* or *Tet3*) or for a negative control (*Cd19*) were mixed 1:1 and transferred into recipients that were infected with LCMV-Armstrong (Figs 6A and S7A). Circulating T cell populations of the spleen were compared to T_{RM} the salivary gland (SG), kidney (Kid), and small intestine intraepithelial layer (IEL) or lamina propria (LP) on day 7 and 21 of infection. By day 21, T cells lacking *Kdm2b* resided at a reduced frequency in all tissues examined compared to control cells (Fig 6D). Furthermore, the proportion of T cells in the gut expressing the canonical T_{RM} markers CD69 and CD103 was reduced with decreased *Kdm2b* expression compared to control cells (S4B Fig). *Tet3*-deficient T cells were also impaired in their ability to populate the SG and LP (Fig 6E) at day 21 and showed an early defect in differentiation with reduced CD69 and CD103 expression on day 7 of infection compared to control IEL and LP T cells. Thus, the TFs, *Kdm2b* and *Tet3*, function to promote generation of CD8⁺ T_{RM} populations and represent newly identified regulators of T_{RM} identified regulators of this important T cell memory population, further supporting the tremendous predictive value of Taiji.

Based on the regulatory network Taiji generated, we were able to extract regulatees of selected TFs in the corresponding cell states and performed functional analysis of the regulatees. The regulatees of all four selected TFs (*Elf1*, *Prdm9*, *Kdm2b*, *Tet3*) are related to T cell differentiation based on the enriched GO terms including alpha-beta T cell activation, T cell receptor signaling pathway, Th17 cell differentiation. The regulatees of the four novel TFs have functional overlap with the regulatees of known TFs (we selected *Foxo1*, *Ikzf3*, *Hic1*, *Ahr* as examples). The regulatees of the known TFs are also enriched with T cell-related functional terms like CD4-positive and alpha-beta T cell differentiation and activation, T cell receptor signaling pathway, and positive regulation of cytokine production. The results are summarized in S7 Table.

Systematic comparison between human and mouse immune cells reveals conserved regulatory code

To systematically compare the regulatory code, we used Taiji to construct regulatory networks in human immune cell populations [76,77] similar to those we generated in the mouse immune system and obtained a PageRank matrix. Twenty-four human samples were paired with 29 mouse samples. These 53 samples spanning 8 immune cell lineages were used for all the downstream analysis in Fig 7.

The hierarchical clustering based on PageRank scores of 733 common TFs showed diverse driving forces in clustering across cell lineages (Fig 7A). Some samples with similar cell identity are more likely to be clustered together regardless of the species they come from. For instance, most B cells from both mouse and human are clustered; DCs and monocytes also show proximity within the same cell lineage. Other cell lineages including several T cell lineages are more inclined to be clustered by species.

CD4⁺ and CD8⁺ T cells show more complex patterns indicative of the cell lineage heterogeneity both within and across the source organisms. The cell-lineage specific TFs for human and mouse are identified following the same method described previously in “Materials and Methods: Identification of putative driver TFs in cell lineage specification”. The comparison of specific TFs between human and mouse revealed similar TF regulatory profiles across species (Fig 7C). Twenty-four percent of total B cell-specific TFs are shared between the two

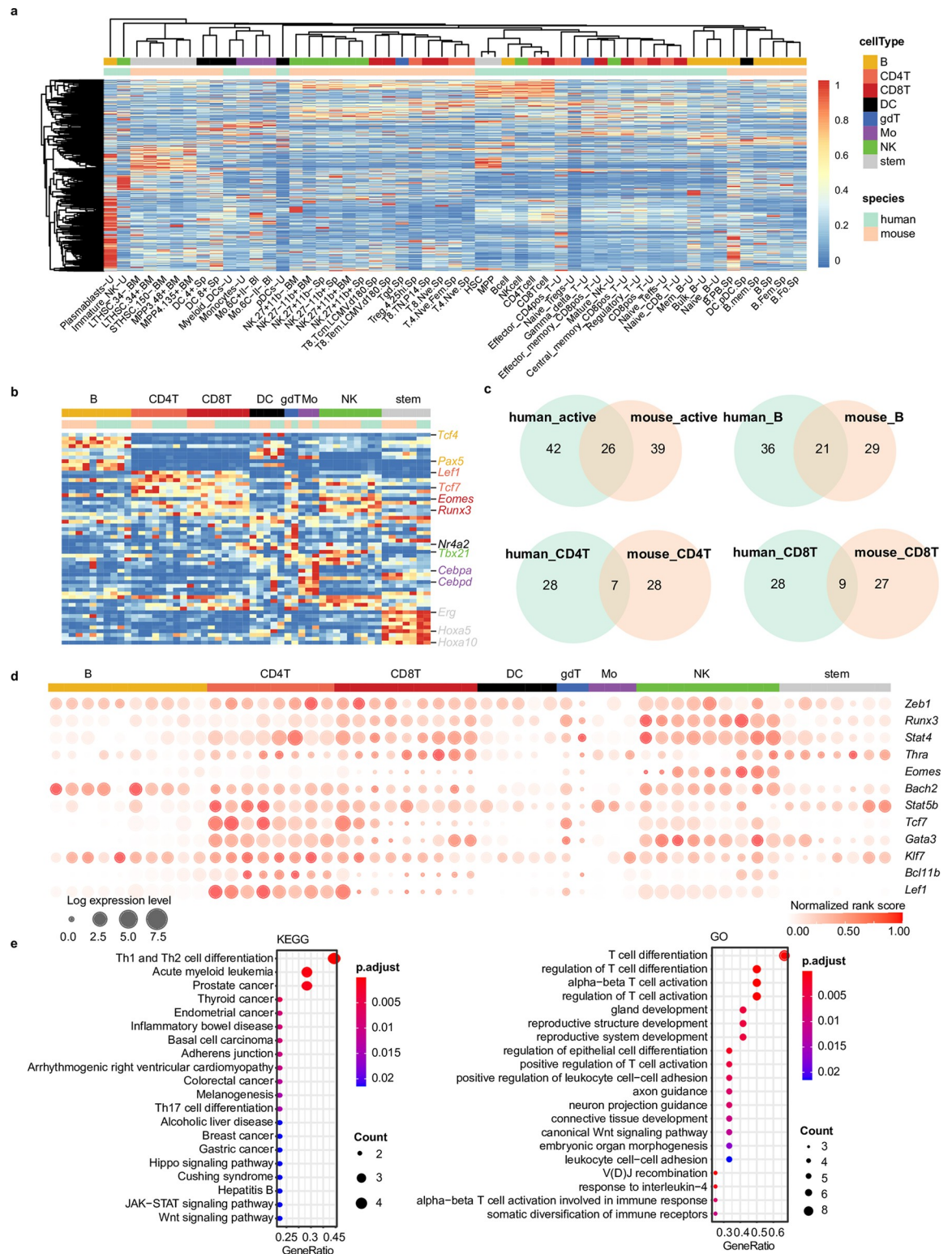


Fig 7. Comparison between human and mouse immune system. (A) Hierarchical clustering of 53 cell types from both human and mouse systems based on row-wise normalized PageRank scores of 733 TFs. TFs in rows, cell types in columns, and color of the cell in the matrix indicate the normalized PageRank scores with red displaying high scores. (B) PageRank scores heatmap of top10 cell lineage-specific TFs. Selected gene names are colored based on their cell lineage specificity. Legends and columns are the same as in (A). (C) Comparison of active TFs and 3 comparisons of specific TFs between human and mouse systems. “Active” means the constitutively

active TFs, whose average PageRank scores across all the cell types were ranked top 10% with CVs<0.5. (D) Twelve TFs important in several cell lineages in both mouse and human. Circle size represents the logarithm of gene expression, and the color represents the normalized PageRank score. Column order is the same as (B). (E) Enriched KEGG pathways and GO terms of 12 putative driver TFs identified in (D).

<https://doi.org/10.1371/journal.pcbi.1010116.g007>

organisms including *Tcf4* [37–40], *Pax5* [78–80], *Bhlhe41* [81–83], and several *Irf* family TFs (*Irf4*, *Irf5*, *Irf8*, [23,84,85]). This relatively large overlapping percentage is consistent with the clustering of B cells across species shown in Fig 7A. Similarly, the overlapping percent of constitutively active TFs is relatively high (24%), showing that these house-keeping genes are required for fundamental biological functions across a variety of cell types and species. On the other hand, the overlap percentages for CD4⁺ and CD8⁺ T cells are relatively small, 11% and 14%, respectively. The conserved TFs like *Eomes* [86–88], *Bcl11b* [89–91], *Lef1* and *Tcf7* [26,92,93] are well-known for their important roles in T cell differentiation in both systems. We also identified species-specific TFs including *Msc*, *Maz*, *Klf5* and *Snai3*, *Sp4*, and *Foxo4* in human and mouse CD8⁺ T cells, respectively.

We performed Fisher's exact test to check if the intersection list of human and mouse specific TFs is significant. All 9 comparisons including 8 cell lineage-specific TFs and active TFs passed the test with p-value as 0.05 except the stem-specific TFs (p-value as 0.2). It's possibly due to the relatively large identity differences between human and mouse stem cells, which is consistent with our clustering result in Fig 7A where the human and mouse stem cells are not clustered together. As for the experimentally validated TFs, we were unable to perform similar identification of tissue-resident-specific TFs in the human system as the data is unavailable—the human dataset does not include Trm samples. We can identify the memory-specific TFs in the human system and the intersection list includes 3 TFs (*Ikzf3*, *Klf2*, *Snai3*) with p-value of 0.04. The full test results along with the specific TF lists are summarized in S9 Table.

The heatmap in Fig 7B displays the PageRank scores of cell-lineage specific TFs shared between mouse and human. Some of them are well-known in both organisms, like *Cebpd*, *Cebpa* as monocytes-specific TFs [94,95]; *Erg* [96,97], several *Hoxa* [98–100] family TFs as stem cell-specific TFs. Some of these specific TFs can be very important in several cell lineages in both mouse and human systems, like *Tcf7*, *Lef1*, *Bcl11b*, *Gata3* shown in Fig 7D, suggesting the potential sophisticated regulatory roles in closely related cell lineages like CD4⁺ and CD8⁺ T cells. The functional analysis of TFs shows strong connection with regulation of T cell differentiation (Fig 7E). This comparative analysis between human and mouse reveals some common regulatory code across species, which would be very valuable for elucidating the functions of unknown TFs in the human immune system.

Discussion

The mouse immune system provides an ideal setting for implementing system-level analyses of cellular differentiation as well-defined cell states have been established and the differentiation stages from progenitor to mature cells have been thoroughly characterized [101]. Considerable efforts have been made in mapping the regulatory networks driving immune cell differentiation, yet identification of the critical TFs that regulate cell fate decisions at key differentiation branch points has remained a challenge. Here, we have systematically identified key TFs in 9 immune cell lineages that encompass a range of key lineages in the mouse and human immune systems based on the Taiji framework. Taiji integrates gene expression and epigenomic information (ATAC-seq or histone modification) to construct regulatory networks, thus better delineating the cell identity than single-omics data. Furthermore, Taiji calculates PageRank scores for each regulator by considering the impacts of its upstream regulators and

downstream regulatees including the feedback from its regulatees. Importantly, Taiji considers not only the expression levels of a gene's regulators and regulatees but also the strength of each regulatory interaction indicated by the extent of open chromatin, motif match, and expression of the TF itself. Therefore, the PageRank score of a gene in Taiji represents its predicted global importance in the genetic network. The "master regulators" would have the highest PageRank scores as they regulate many TFs that in turn regulate highly expressed genes. Note that the majority of the TFs have PageRank scores well correlated with their expression levels but there are a substantial portion that show weak or no correlation. These TFs have moderate expression themselves but regulate many highly expressed genes and their regulatory activities are thus better characterized by PageRank scores, which is a particularly powerful benefit of this approach: Taiji assesses functional regulation of target genes rather than TF expression dynamics.

Other similar methods that identify important TFs have been reported before. Specifically, SCENIC developed by Aibar et al 2017 [9] aims to analyze single cell RNA-seq data without considering ATAC-seq. This method only uses gene expression data and constructs the genetic network by predicting each gene's expression using the TFs' expression levels and finding the most predictive TFs as the regulators. SCENIC cannot be applied to the bulk mouse immune data [16] as Taiji did and does not have enough samples to select the regulator TFs. Furthermore, the motif analysis in SCENIC only considers sequences around TSS while Taiji includes both promoters and enhancers.

Importantly, Taiji integrates RNA-seq and ATAC-seq data to construct a genetic network so it can capture the global impact of TFs in the genetic network. Taiji considers not only the effect on TF's direct target genes, but also all the descendants in the whole network. Due to the iterative calculation of PageRank scores, Taiji takes into account the impact of its upstream regulators and downstream regulatees including the feedback from its regulatees, which makes Taiji capable of assessing the global influence for individual TFs. This global view is what distinguishes Taiji from other methods. It is worth noting that we have previously shown that Taiji is very robust and resistant to noise [13] (Fig 2 in Zhang et al., 2019). The PageRank scores can correctly predict TF's importance even when 80% of the edges in the network are disturbed.

As for the comparison with the Maslova et al. 2020 [10] analysis, Taiji integrates RNA-seq and ATAC-seq to construct a genetic network and assess the global importance of TFs while AI-TAC focuses exclusively on predicting ATAC-seq signals using sequences and extracting the motifs predictive of cell type specification. Maslova et al. did not use gene expression information in their analysis and did not build a genetic network. Identifying enriched motifs in specific cell types is equivalent to considering the direct targets of TFs when evaluating the TF importance. While such an approach is useful, we have demonstrated in our previous work [13,14] that the PageRank scores performed better by considering not only the direct target genes of TFs but also their own regulators and the descendant genes of their direct targets. In this sense, our study is the first systems biology approach that integrates RNA-seq and ATAC-seq to systematically identify the key regulators during mouse immune system development.

Furthermore, if a motif is weak, the PageRank score of the TF is decided by whether these motifs occur in the open chromatin regions (measured by the peak intensity of the ATAC-seq data), the TF expression and its target expression levels. We also compare TFs between different cell types to identify cell-type/lineage specific regulators. The relative difference between the PageRank scores of TFs also helps to uncover important TFs with weak motifs. As such, our method identifies *Gata3* as an important TF in several T cell subsets (see S2 Table), which is a CD4⁺ T cell master regulator with weakly defined sites. Therefore, integration of motif strength with open chromatin strength, expressions of TF and its target genes is superior to

only considering motifs that are present or absent in open chromatin as in Maslova et al. study.

Another advantage of Taiji is its capability to construct genetic networks in individual cell types and thus allows identifying cell-type specific regulators. By comparing PageRank scores of TFs across cell types, we have successfully identified cell lineage-specific putative driver TFs, thus constructing a comprehensive catalog of key regulators in various immune cell types. In addition to identifying dozens of known regulators, for instance, *Tcf7* [31–35], *Lef1* [26,32,33], *Stat4* [102], *Gata3* [5,103] as T cell-specific TFs; *Bcl11a* [20,21], *Tcf4* [37–40], *Ebf1* [79,104,105], *Pax5* [78–80] etc. as B cell-specific TFs, we have also found several TFs with previously undefined roles in regulation of immune cell lineage differentiation and development, for instance, *Pbx4* may have important role in T cell development; *Meis1*, *Jdp2* may be critical regulators in ILC lineage development. Overall, the identified novel cell type-specific TFs provide insights for the future experimental validation to fully reveal the underlying regulatory mechanisms.

We also provided here a comprehensive temporal transcriptomic and epigenomic dynamics in a diverse set of cell types, revealing “transcriptional waves” in which combinations of TFs are particularly active in specific cell types during specific developmental stages. The observed transcriptional waves patterns suggest the possible mechanisms of how the TF activities are coordinated to orchestrate the cell type differentiation. We found different types of transcriptional waves: constitutively active waves (Fig 5A and 5B), cell lineage-specific waves (Fig 5C and 5D), and stage-specific waves (Fig 5E–5H). The cell lineage-specific waves are composed of coordinately highly ranked TFs across multiple cell types in the same cell lineage. This indicates that certain TFs are “multi-taskers” [106] which play critical roles in multiple or successive developmental stages along one lineage. The stage-specific waves, on the other hand, show TF combinations having exclusively important roles during a very specific developmental stage. For example, three *Duxbl* family members display high rank in preT.DN3.Th (Fig 5F), consistent with that *Duxbl* was shown to have high recombination activity before β -selection and result in a developmental block at the DN3-to-DN4 transition due to enhanced apoptosis and reduced proliferation of pre-T cells in thymus [57]; noticeably, there are additional TFs in the same cluster, such as *Lcor*, *Tcf4* and *Runx2*, suggesting their possible roles in the development of preT.DN3.Th. Another example is C10, which is specifically important in T8.Tem.LCMV.d180.Sp and most of the 19 TFs have not been reported to function in T_{EM}, such as *Rfx6*, *Lhx4* and *Atf6*. Identification of these transcriptional waves with defined regulatory patterns and predicted regulatory interactions not only uncover new regulators but also would greatly facilitate mechanistic studies.

To reveal the conserved regulatory code across organisms, we also constructed the regulatory networks in matched human immune cells using the Taiji framework. The systematic comparison of constitutively active TFs and cell lineage-specific TFs showed that some TFs undertake similar crucial roles in the development and differentiation of immune cells in both human and mouse systems, like *Stat1*, *Stat3* [107–109], *Sp2* and *Sp3* [110–112] as common constitutively active TFs; *Eomes* [86–88], *Bcl11b* [89–91], *Lef1* and *Tcf7* [26,92,93] as CD4⁺ and CD8⁺ T cell-important TFs. This systematic analysis that spans all the immune cell lineages identifies the conserved TFs across species and shows the potential of elucidating the molecular functions of unknown TFs in the human immune system.

To demonstrate the power of our analysis, we experimentally validated 4 regulators that have no reported functions in circulating and tissue-resident CD8⁺ T cell memory. T cells responding to infection engage a transcriptional network that directs differentiation of progeny that adopt a range of cellular states; thus, creating a heterogeneous population of effector T cells composed of cells able to eliminate the present infection and others capable of providing

robust long-lasting immunological memory [106,113]. A large fraction of effector T cells are terminally differentiated and after providing immediate effector function to clear the pathogen from the host will undergo apoptosis. Alternatively, other cells will be maintained in a 'stem-like' state capable of seeding the memory T cell population that provides protection from reinfection [106,113]. The T cells within the memory population also exist in a spectrum of phenotypes. T_{EM} and T_{RM} reside in the circulation and peripheral tissues, respectively, and provide an immediate effector response while T_{CM} remains less differentiated, able to undergo extensive proliferation to generate a new burst of effector T cells [106]. While key transcriptional regulators that program these cellular fates have been identified [114], a comprehensive network of TFs has not been completed, particularly when the subsets represented within the memory T cell population are considered. Our analysis uniquely compared several memory B and T cell populations as a whole to effector populations and identified a number of novel putative important TFs of memory states, including *Elf1* and *Prdm9*. *Elf1* has been associated with NKT cell development [115] and *Prdm9* with regulation of cell cycle, a process closely linked with cellular differentiation [116,117]. Our validation of their role in $CD8^+$ T cell differentiation following acute infection clearly demonstrated essential roles for these TFs in promoting the terminally differentiated state of $CD8^+$ effector T cells. This suggests that *Elf1* and *Prdm9* control the terminal differentiation transcriptional program directly or inhibit the program that promotes formation of the memory cell state. While our analysis predicts relative regulatory weight of the TF at the given cell-fate branch point, it cannot determine direction of regulation (i.e., whether it promotes or inhibits). Thus, additional hits from our analysis may be vital for regulating differentiation of the memory T cell subsets as suggested by the identification of the well-known TF *Foxo1* that promotes and sustains the $CD8^+$ T cell memory population [63–66].

While distinct regulatory networks drive the differentiation and maintenance of effector and memory $CD8^+$ T cells cell states, additional transcriptional programs define the trafficking patterns and location of T cells within the host. To identify TFs important for directing the residency of T cells within peripheral tissues, we compared several lymphocyte populations localized to the gut to those found in the circulation. We identified numerous known and also unknown regulators including *Kdm2b* and *Tet3*. We validated the function of *Kdm2b* and *Tet3* in the formation of $CD8^+$ T_{RM} across several tissues. $CD8^+$ T_{RM} failed to accumulate the following infection in the absence of *Kdm2b*. *Kdm2b* has been shown to promote not only cell proliferation but also migration of mouse embryonic fibroblasts [118] and analogous mechanisms may be in play in tissue-resident lymphocyte populations. Our validation experiments also revealed that the differentiation of $CD103^+$ T_{RM} cells were delayed with the loss of *Tet3*. Interestingly, *Tet3* has been linked to the regulation of multiple key TGF pathway genes [119], a signaling pathway that is critical for the generation and maintenance of $CD103^+$ $CD8^+$ T_{RM} [120].

Taken together, our analyses have identified critical TFs that regulate two important T cell differentiation pathways—determination of cell state/fate and cellular trafficking/localization. The literature evidence of well-known TFs and the validation experimental results of novel regulators suggest the power of our study in guiding mechanistic investigations in the future. With the increasing understanding of $CD8^+$ T cell community, our approach will also be applicable to further identify key regulators in the finer definition of $CD8^+$ T cells, thus facilitating the novel discovery of $CD8^+$ T cell diversities and functions. Importantly, our approach is generally applicable to any cell type. Given the fast accumulating transcriptomic and epigenomic data generated from projects such as Human Cell Atlas, Taiji provides a powerful tool to profile the landscape of regulators that define cell-type and cell-state specificity, which would

facilitate understanding the underlying mechanisms and elucidate manipulations such as over-expression of the key regulators to manipulate cell fate.

Materials and methods

Dataset acquisition

The mouse dataset was downloaded from Gene Expression Omnibus (GEO) (GSE100738 and GSE109125 corresponding to ATAC-seq and RNA-seq, respectively [16]). We matched the ATAC-seq and RNA-seq by sample name and 115 samples in 73 cell types were used as input to Taiji. The cell types span nine cell lineages: stem cells, dendritic cells (DC), macrophages/neutrophils (MF/GN), monocytes (Mo), B cells, innate lymphoid cells (ILC), $\alpha\beta$ T cells (abT), $\gamma\delta$ T cells (gdT), and activated T cells (Act_T).

The human datasets analyzed in this study were retrieved from GEO (GSE118189 and GSE118165, GSE74912 and GSE74246) [76,77]. In total, 226 samples distributed in 61 immune cell types were chosen as input. The cell types span nine cell lineages: B, CD4⁺T, CD8⁺T, $\gamma\delta$ T, monocytes, dendritic cells, natural killer (NK), and stem cells. All of the sample names and division methods were consistent with the original paper.

TF regulatory networks construction and visualization

Taiji v1.1.0 with default parameters was used for the integrative analysis of RNA-seq and ATAC-seq data (<https://github.com/Taiji-pipeline/Taiji>). The motif file was downloaded directly from the CIS-BP database. There were 815 mouse motifs and 1112 human motifs. The majority of these proteins are TFs. The PageRank scores for mouse and human are shown in S1 Table.

The network shown in Fig 3C is a subset of the whole network for PBX4 in T8.IEL.LCMV.d7.SI. We first removed the edges with weight less than 1 and then retained 500 out of 1437 regulatees via random sampling while keeping the ratio between non-TF nodes and TF-nodes constant. Next, we constructed the regulatory network between TF-nodes and non-TF nodes within 500 regulatees and thus expanded the network to 4213 edges. For better visualization we further downsampled 25% of the total edges and used it as input. Cytoscape v3.8.0 was used for network visualization. The size of the node is proportional to its PageRank score.

TF regulatory networks weighting scheme

As described in the original Taiji paper [13], a personalized PageRank algorithm is applied to calculate the ranking scores for TFs. We first initialized the edge weights and node weights in the network. The node weight was calculated as e^z , where z_i is the gene's relative expression level in cell type i , which is computed by applying the z score transformation to its absolute expression levels. The edge weight is determined by $e_{ij} = \sqrt{g \sum_{k=1}^n p_k * m_k}$, where p is the peak intensity, calculated as $\frac{1}{1+e^{-(x-b)}}$, where x is $-\log_{10}(p)$, represented by the p-value of the ATAC-seq peak at the predicted TF binding site, rescaled to [0, 1] by a sigmoid function; m is the motif binding affinity, represented by the p-value of the motif binding score, rescaled to [0, 1] by a sigmoid function; g is the TF expression value; n is the number of binding sites linked to gene j . Let s be the vector containing node weights and W be the edge weight matrix. The personalized PageRank score vector v was calculated by solving a system of linear equations $v = (1 - d)s + dWv$, where d is the damping factor (default to 0.85). The above equation can be solved in an iterative fashion, i.e., setting $v_{t+1} = (1 - d)s + dWv_t$.

If the TFs in the same protein family share the same motifs, their PageRank scores are distinguished by their own expression levels because their motifs and the target genes are the

same. If a motif is weak, the PageRank score of the TF is decided by whether these motifs occur in the open chromatin regions (measured by the peak intensity of the ATAC-seq data), the TF expression and its target expression levels. The relative difference between the PageRank scores of TFs also helps to uncover important TFs with weak motifs.

Identification of putative driver TFs in cell lineage specification

To identify lineage-specific TFs, we divided the samples into two groups: target group and background group. Target group included all the samples belonging to the cell type/lineage of interest and the background group comprised the remaining samples (S1 Table). We then performed the normality test using Shapiro-Wilk's method to determine if each TF is log-normally distributed across two sets of samples in each test. And on average, most of the TFs (90%) follow the log-normality distribution (p-value of 0.05). Based on log-normality assumption, an unpaired t-test with Welch's correction was used to calculate the P-value. As for the very small target group (size smaller than 3) like Mo, two-sample unpaired t-test is no longer available. So, we used one sample t-test to test whether the mean of background population is equal to some value. A P-value cutoff of 0.001 and \log_2 fold change cutoff of 1 were used for calling lineage-specific TFs (see the full result in S2 Fig).

Identification of memory cell-specific and tissue-resident-specific TFs

To identify memory cell-specific TFs, we grouped three memory cells, i.e. B.mem.Sp, T8.Tcm.LCMV.d180.Sp, T8.Tem.LCMV.d180.Sp, into target group while leaving other 70 samples except for T8.MP.LCMV.d7.Sp as background, because T8.MP.LCMV.d7.Sp is a memory precursor cell. The normality test showed that most samples were log-normal distributed so that we performed unpaired t-test with Welch's correction based on log-normality assumption. A P-value cutoff of 0.1 was used to identify the memory cell-specific TFs. The list was filtered by removing the digestion-related genes. We identified 30 memory cell-specific TFs.

This partition best reflects the biological identities. However, the unbalanced number of two partitions needs to be considered. That is why we applied Welch's correction to t-test, which accounts for the unequal variance brought by the unbalanced number. To further reduce the effect of confounding factors, we randomly sampled three cell types from the non-memory group to make the t-test balanced and applied the same criteria as in the main text (p-value of 0.1) to select specific TFs in each test. We repeated the balanced Welch's t-test for 100, 500, and 1000 times respectively and ranked the TF based on the selected frequency among all the tests. We compared the overlap set between the memory TF list in the main text and the ranked list using the permutation test. More than half of the memory TFs are recovered including Prdm9 when compared with top 30 TFs and 80% of the memory TFs are recovered including both Prdm9 and Elf1 when compared with top 60 TFs. This permutation test shows that our memory TF list is reliable. The permutation test result is summarized in S8 Table.

To identify tissue-resident-specific TFs, we grouped six lymphocytes resident in small intestine (ILC3.NKp46+.SI, ILC3.NKp46-CCR6-.SI, ILC3.CCR+.SI, ILC2.SI, Treg.4.FP3+.Nrpl.Co, T8.IEL.LCMV.d7.SI) as target group and eight similar cells in circulation as background (T8.Tcm.LCMV.d180.Sp, T8.Tem.LCMV.d180.Sp, T8.MP.LCMV.d7.Sp, T8.TE.LCMV.d7.Sp, Treg.4.25hi.Sp, NK.27+11b-.Sp, NK.27+11b+.Sp, NK.27-11b+.Sp). Based on the log-normality assumption, we performed the unpaired t-test with Welch's correction and used double cutoff (P-value < 0.1 and \log_2 fold change > 0.5) for calling the putative driver TFs. The list was filtered by removing the digestion-related genes. A list of 54 tissue-residency-specific TFs were thus identified.

The cutoff selection for p-values and the fold changes depends on different biological comparisons and different sample effect sizes. When choosing the cutoff values, we consider both the interpretation significance and the output TF list size. We apply more stringent cutoff values in the identification of lineage-specific TFs and sub-cell lineage-specific TFs compared to that in the tissue-resident-specific and memory-specific TFs because different lineages are more distinct from each other than cell types within the same lineage due to both biological differentiation and sample size. The guideline for the cutoff choice is to keep a reasonable size of output TF list as 30–50, which is an empirical range considering both biological significance and interpretation convenience. As for the relatively loose cutoffs in the identification of tissue-resident- and memory-specific TFs, it's mainly restricted by the small effect sizes. That's also part of the reason to experimentally validate the TFs with relatively low significance (*Tet3*, *Elf1* with p-value as 0.1) in tissue-resident-T cells and memory T cells formation.

GO and pathway enrichment analysis

R package clusterProfiler 4.0 (<https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>) was used to perform the enrichment analysis of the identified TFs, in which KEGG Pathway (http://www.genome.jp/kegg/catalog/org_list.html) and DAVID (via RDAVIDWebService <https://bioconductor.org/about/removed-packages/>) databases are used to query the functional terms. A cutoff of P-value < 0.01 was used to select the significantly enriched pathways.

Mouse and cell lines

All mouse strains were bred and housed in specific pathogen-free conditions in accordance with the Institutional Animal Care and Use Guidelines of the University of California San Diego. C57Bl/6J, P14 (with transgenic expression of H-2D^b-restricted TCR specific for LCMV glycoprotein gp₃₃), CD45.1⁺ and CD45.1.2⁺ congenic mice were bred in house. Both male and female mice were used throughout the study, with sex- and age-matched T cell donors and recipients.

Infections studies

shRNAmirs targeting mouse *Elf1*, *Prdm9*, *Kdm2b*, *Tet3* or *Cd19* in a pLMPd-Amt vector were used to generate retroviral supernatant as previously described [14]. Naive P14 T cells from the spleen and LN were enriched via negative selection using MACS magnetic beads following the manufacturer's protocol (Miltenyi Biotec) then activated at 2x10⁶ cells/well in six-well plates coated with 100µg/mL goat anti-hamster IgG (H+L, ThermoFisher), 1µg/mL anti-CD3 (145-2C11), and 1µg/mL anti-CD28 (37.51) (eBioscience) for 18 h. Retroviral supernatant supplemented with 50µM beta-mercaptoethanol and 8µg/mL polybrene (Millipore) replaced the culture media and cells were centrifuged for 60 min at 2,000 rpm, 37°C. After 24 h, P14 T cells transduced with gene targeting or *Cd19* control shRNAmir encoding-retroviruses were mixed at 1:1 ratio and a total of 5x10⁵ cells were transferred to congenically distinct recipient mice that were infected with 2x10⁵ plaque-forming unit (PFU) LCMV-Armstrong i.p. A portion of the P14 CD8⁺ T cells transduced with gene targeting or *Cd19* control shRNAmir encoding-retroviruses were cultured *in vitro* for another 24h in media supplemented with 50 U/ml of human IL-2 and sorted CD8α⁺Ametrine⁺ cells were evaluated by qPCR for knock-down efficiency using the following primer pairs: *Elf1* F:CCAAGTATTAGGACTATACAGG, *Elf1* R:GGCTATAACCGTTGTGAGTG; *Prdm9* F:ATATGGAATGGAATCATCGC, *Prdm9* R:GTGCTG GGAAAGGTTGTT; *Kdm2b* F:AGCAGACAGAAGCCACCAAT, *Kdm2b* R:

AGGTGCCTCCAAAG TCAATG; *Tet3* F: TGCGATTGTGTCGAACAAATAGT, *Tet3* R: TCCATACCGATCCTC CATGAG.

Preparation of single cell suspensions

Single-cell suspensions were prepared from spleen by mechanical disruption. For small intestine preparations, Peyer's patches were excised, and luminal contents were removed. The tissue was cut longitudinally then into 1cm pieces then incubated in 10% HBSS/HEPES bicarbonate solution containing 15.4mg/100mL of dithioerythritol (EMD Millipore) on a stir plate with rotating stir bar at 37°C for 30 min to extract intraepithelial lymphocytes (IEL). Gut pieces were further treated with 100U/ml type I collagenase (Worthington Biochemical) in RPMI-1640 containing 5% bovine growth serum, 2 mM MgCl₂ and 2 mM CaCl₂ at 37°C for 30 min. Salivary gland and kidney were cut with scissors into fine pieces then incubated while shaking with 100U/ml type I collagenase at 37°C for 30 min. Lymphocytes from all tissues but spleen were purified on a 44%/67% Percoll density gradient.

Flow cytometry and cell sorting

The following antibodies (from eBioscience and Biolegend) were used for surface staining: CD8 α (53–6.7), CD8 β (eBioH35-17.2), CD45.1 (A20-1.7), CD45.2 (104), CD62L (MEL-14), CD69 (H1.2F3), CD103 (2E7), CD127 (A7R34), KLRG1 (2F1). Cells were incubated for 30 min at 4°C in PBS supplemented with 2% bovine growth serum and 0.1% sodium azide. To identify tissue-resident CD8⁺ T cells, 3 mg of CD8 α (53–6.7) conjugated to APC eFlour780 was injected into mice three minutes prior to sacrifice and organ excision. CD8 β cells without CD8 α labeling were considered to be localized within non-lymphoid tissues. Stained cells were analyzed using LSRFortessa or LSRFortessa X-20 (BD) and FlowJo software (TreeStar). Cell sorting was performed on BD FACSAria Fusion instruments.

Quantification and statistical analysis

Statistical analysis was performed using GraphPad Prism software. A one-way paired ANOVA or two-tailed ratio paired t-test was used for comparisons between groups as indicated in figure legends. P values of <0.05 were considered significant.

Supporting information

S1 Fig. Topological properties of genetic networks. (A) The number of edges (top) and nodes (bottom) in each genetic network. (B) The distribution of the number of regulators per gene for each genetic network. (C) The distribution of the number of regulatees per TF for each genetic network.

(PDF)

S2 Fig. PageRank scores (color shade) and expression levels (circle size) of 48 constitutively active TFs across all 73 cell types.

(PDF)

S3 Fig. Identification of important TFs in 9 immune cell lineages. (A) $\alpha\beta$ T cell. (B) Act₂T cell. (C) $\gamma\delta$ T cell. (D) ILC. (E) DC cell. (F) Mo. (G) MF/GN. (H) Stem cell. (I) B cell. The circle size and color scale are the same as Fig 4.

(PDF)

S4 Fig. The cumulative proportion of variance explained against the number of principal components (PCs). The first 30 PCs were kept according to the “elbow” method.

(PDF)

S5 Fig. Selecting the best distance metric and number of clusters according to the Silhouette metric. The Pearson correlation was chosen and $k = 40$ was the ideal cluster number.

(PDF)

S6 Fig. Forty transcriptional waves with clusters of TFs direct cell lineage differentiation in the mouse immune system.

(PDF)

S7 Fig. Validation of novel TFs’ roles in CD8⁺T memory cell formation and tissue residency. (A) qPCR analysis to assess knockdown efficiency of indicated gene in P14 cells prior to transfer. (B,C) Representative expression of CD69 and CD103 on *Kdm2b* KD and control P14 cells recovered from indicated tissue on day 7 and 21 of infection (left). Quantification of the frequency of populations (right). (D,E) Representative expression of CD69 and CD103 on *Tet3* KD and control P14 cells recovered from indicated tissue on day 7 and 21 of infection (left). Quantification of the frequency of populations (right). Numbers in graphs indicate percent of cells in the corresponding gate. Data are cumulative of 2 (B,C) or 3 (D,E) independent experiments with $n = 3-4$. Graphs show mean \pm SEM. A two-tailed ratio paired t-test was used to determine statistical significance; * $p < 0.05$, ** $p < 0.01$.

(PDF)

S1 Table. PageRank scores and expression values for all TFs across different immune cell types in mouse and human systems.

(XLSX)

S2 Table. Cell types in target/background groups in the identification of putative driver TFs.

(XLSX)

S3 Table. P-values and log₂ fold change in the identification of important TFs in 9 immune cell lineages.

(XLSX)

S4 Table. Predicted putative driver TFs in the five cell lineages.

(DOCX)

S5 Table. Predicted putative driver TFs in memory cells and tissue resident cells.

(DOCX)

S6 Table. GO terms and KEGG pathway enrichment statistics for the identified putative driver TFs in Fig 3 and Fig 4.

(XLSX)

S7 Table. GO terms and KEGG pathway enrichment statistics for the regulatees of 4 known TFs and 4 experimentally validated TFs.

(XLSX)

S8 Table. Permutation test of memory TF identification.

(XLSX)

S9 Table. Overlap of cell type-specific TF lists between human and mouse.

(XLSX)

Author Contributions

Conceptualization: Cong Liu, Ananda W. Goldrath, Wei Wang.

Data curation: Cong Liu, Kyla Omilusik.

Formal analysis: Cong Liu, Kyla Omilusik, Clara Toma, Nadia S. Kurd.

Funding acquisition: John T. Chang, Ananda W. Goldrath, Wei Wang.

Investigation: Kyla Omilusik, Clara Toma, Nadia S. Kurd.

Methodology: Cong Liu, Kyla Omilusik.

Project administration: Ananda W. Goldrath, Wei Wang.

Resources: Cong Liu, Kyla Omilusik, Clara Toma, Nadia S. Kurd.

Software: Cong Liu.

Supervision: John T. Chang, Ananda W. Goldrath, Wei Wang.

Visualization: Cong Liu, Kyla Omilusik, Clara Toma, Nadia S. Kurd.

Writing – original draft: Cong Liu, Kyla Omilusik.

Writing – review & editing: Cong Liu, Kyla Omilusik, Ananda W. Goldrath, Wei Wang.

References

1. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol*. 2010; 125: S3–23. <https://doi.org/10.1016/j.jaci.2009.12.980> PMID: 20176265
2. Brodin P, Davis MM. Human immune system variation. *Nat Rev Immunol*. 2017; 17: 21–29. <https://doi.org/10.1038/nri.2016.125> PMID: 27916977
3. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. *Nat Immunol*. 2017; 18: 725–732. <https://doi.org/10.1038/ni.3768> PMID: 28632713
4. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*. 2012. pp. 613–626. <https://doi.org/10.1038/nrg3207> PMID: 22868264
5. Hosokawa H, Rothenberg EV. How transcription factors drive choice of the T cell fate. *Nat Rev Immunol*. 2020; 21: 162–176. <https://doi.org/10.1038/s41577-020-00426-6> PMID: 32918063
6. Monticelli S, Natoli G. Transcriptional determination and functional specificity of myeloid cells: making sense of diversity. *Nat Rev Immunol*. 2017; 17: 595–607. <https://doi.org/10.1038/nri.2017.51> PMID: 28580958
7. Schacht T, Oswald M, Eils R, Eichmüller SB, König R. Estimating the activity of transcription factors by the effect on their target genes. *Bioinformatics*. 2014; 30: i401–7. <https://doi.org/10.1093/bioinformatics/btu446> PMID: 25161226
8. Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, Shuster B, et al. An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol Syst Biol*. 2015; 11: 839. <https://doi.org/10.1525/msb.20156236> PMID: 26577401
9. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*. 2017; 14. <https://doi.org/10.1038/nmeth.4463> PMID: 28991892
10. Maslova A, Ramirez RN, Ma K, Schmutz H, Wang C, Fox C, et al. Deep learning of immune cell differentiation. *Proc Natl Acad Sci U S A*. 2020; 117: 25655–25666. <https://doi.org/10.1073/pnas.2011795117> PMID: 32978299
11. Zotenko E, Mestre J, O’Leary DP, Przytycka TM. Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. *PLoS Computational Biology*. 2008. p. e1000140. <https://doi.org/10.1371/journal.pcbi.1000140> PMID: 18670624
12. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nature Reviews Genetics*. 2007. pp. 437–449. <https://doi.org/10.1038/nrg2085> PMID: 17510664

13. Zhang K, Wang M, Zhao Y, Wang W. Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci Adv.* 2019; 5: eaav3262. <https://doi.org/10.1126/sciadv.aav3262> PMID: 30944857
14. Milner JJ, Toma C, Yu B, Zhang K, Omilusik K, Phan AT, et al. Runx3 programs CD8 T cell residency in non-lymphoid tissues and tumours. *Nature.* 2017; 552: 253–257. <https://doi.org/10.1038/nature24993> PMID: 29211713
15. Yu B, Zhang K, Milner JJ, Toma C, Chen R, Scott-Browne JP, et al. Epigenetic landscapes reveal transcription factors that regulate CD8 T cell differentiation. *Nat Immunol.* 2017; 18: 573–582. <https://doi.org/10.1038/ni.3706> PMID: 28288100
16. Yoshida H, Lareau CA, Ramirez RN, Rose SA, Maier B, Wroblewska A, et al. The cis-Regulatory Atlas of the Mouse Immune System. *Cell.* 2019. pp. 897–912.e20. <https://doi.org/10.1016/j.cell.2018.12.036> PMID: 30686579
17. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell.* 2014. pp. 1431–1443. <https://doi.org/10.1016/j.cell.2014.08.009> PMID: 25215497
18. Zhu Y, Chen Z, Zhang K, Wang M, Medovoy D, Whitaker JW, et al. Constructing 3D interaction maps from 1D epigenomes. *Nature Communications.* 2016. <https://doi.org/10.1038/ncomms10812> PMID: 26960733
19. Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. 1999 [cited 28 Jun 2021]. Available: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>
20. Yu Y, Wang J, Khaled W, Burke S, Li P, Chen X, et al. Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med.* 2012; 209: 2467. <https://doi.org/10.1084/jem.20121846> PMID: 23230003
21. Lee B-S, Lee B-K, Iyer VR, Sleckman BP, Shaffer AL, III, et al. Corrected and Republished from: BCL11A Is a Critical Component of a Transcriptional Network That Activates Recombinase Activating Gene Expression and V(D)J Recombination. *Mol Cell Biol.* 2018; 38. <https://doi.org/10.1128/MCB.00362-17> PMID: 29038163
22. Hagman J. Critical Functions of IRF4 in B and T Lymphocytes. *The Journal of Immunology.* 2017; 199: 3715–3716. <https://doi.org/10.4049/jimmunol.1701385> PMID: 29158346
23. Vipul S, Runqing L. IRF4 and IRF8: Governing the virtues of B Lymphocytes. *Front Biol.* 2014; 9: 269.
24. Ubieta K, Garcia M, Grötsch B, Uebe S, Weber GF, Stein M, et al. Fra-2 regulates B cell development by enhancing IRF4 and Foxo1 transcription. *J Exp Med.* 2017; 214: 2059–2071. <https://doi.org/10.1084/jem.20160514> PMID: 28566276
25. Wang Y, Hu J, Li Y, Xiao M, Wang H, Tian Q, et al. The Transcription Factor TCF1 Preserves the Effector Function of Exhausted CD8 T Cells During Chronic Viral Infection. *Front Immunol.* 2019; 10. <https://doi.org/10.3389/fimmu.2019.00169> PMID: 30814995
26. Xing S, Gai K, Li X, Shao P, Zeng Z, Zhao X, et al. Tcf1 and Lef1 are required for the immunosuppressive function of regulatory T cells. *J Exp Med.* 2019; 216: 847–866. <https://doi.org/10.1084/jem.20182010> PMID: 30837262
27. Maxime D, Vijaykumar C, Joana Gomes S, SiegertWerner H. Suppression of Tcf1 by Inflammatory Cytokines Facilitates Effector CD8 T Cell Differentiation. *Cell Rep.* 2018; 22: 2107–2117. <https://doi.org/10.1016/j.celrep.2018.01.072> PMID: 29466737
28. Henze L, Schwinge D, Schramm C. The Effects of Androgens on T Cells: Clues to Female Predominance in Autoimmune Liver Diseases? *Front Immunol.* 2020; 11. <https://doi.org/10.3389/fimmu.2020.01567> PMID: 32849531
29. Yordy JS, Li R, Sementchenko VI, Pei H, Muise-Helmericks RC, Watson DK. SP100 expression modulates ETS1 transcriptional activity and inhibits cell invasion. *Oncogene.* 2004; 23: 6654–6665. <https://doi.org/10.1038/sj.onc.1207891> PMID: 15247905
30. Suzuki J, Maruyama S, Tamauchi H, Kuwahara M, Horiuchi M, Mizuki M, et al. Gfi1, a transcriptional repressor, inhibits the induction of the T helper type 1 programme in activated CD4 T cells. *Immunology.* 2016. pp. 476–487. <https://doi.org/10.1111/imm.12580> PMID: 26749286
31. Amania A, Sheikh JRG. Transcription tipping points for T follicular helper cell and T-helper 1 cell fate commitment. *Cellular and Molecular Immunology.*: 1.
32. Chen Y, Zander R, Khatun A, Schauder DM, Cui W. Transcriptional and Epigenetic Regulation of Effector and Memory CD8 T Cell Differentiation. *Front Immunol.* 2018; 9. <https://doi.org/10.3389/fimmu.2018.02826> PMID: 30581433
33. Xinyuan Z, Haihui X. Generation of memory precursors and functional memory CD8+ T cells depends on TCF-1 and LEF-1. *J Immunol.* 2012; 189: 2722.

34. Jeannet G, Boudousqu   C, Gardiol N, Kang J, Huelsken J, Held W. Essential role of the Wnt pathway effector Tcf-1 for the establishment of functional CD8 T cell memory. *Proc Natl Acad Sci U S A*. 2010; 107. <https://doi.org/10.1073/pnas.0914127107> PMID: 20457902
35. Zhou X, Yu S, Zhao DM, Harty JT, Badovinac VP, Xue HH. Differentiation and persistence of memory CD8(+) T cells depend on T cell factor 1. *Immunity*. 2010;33. <https://doi.org/10.1016/j.immuni.2010.08.002> PMID: 20727791
36. Hosokawa H, Romero-Wolf M, Yui MA, Ungerb  ck J, Quiloan MLG, Matsumoto M, et al. Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing Id2 and Zbtb16. *Nat Immunol*. 2018; 19: 1427–1440. <https://doi.org/10.1038/s41590-018-0238-4> PMID: 30374131
37. Jain N, Hartert K, Tadros S, Fiskus W, Havranek O, Ma MCJ, et al. Targetable genetic alterations of TCF4 (E2-2) drive immunoglobulin expression in diffuse large B cell lymphoma. *Sci Transl Med*. 2019; 11. <https://doi.org/10.1126/scitranslmed.aav5599> PMID: 31217338
38. Laidlaw BJ, Cyster JG. Transcriptional regulation of memory B cell differentiation. *Nat Rev Immunol*. 2020; 21: 209–220. <https://doi.org/10.1038/s41577-020-00446-2> PMID: 33024284
39. Kwon K, Hutter C, Sun Q, Bilic I, Cobaleda C, Malin S, et al. Instructive Role of the Transcription Factor E2A in Early B Lymphopoiesis and Germinal Center B Cell Development. *Immunity*. 2008. pp. 751–762. <https://doi.org/10.1016/j.immuni.2008.04.014> PMID: 18538592
40. W  hner M, Tagoh H, Bilic I, Jaritz M, Poliakova DK, Fischer M, et al. Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. *J Exp Med*. 2016; 213: 1201–1221. <https://doi.org/10.1084/jem.20152002> PMID: 27261530
41. Luc S, Huang J, McEldoon JL, Somuncular E, Li D, Rhodes C, et al. Bcl11a Deficiency Leads to Hematopoietic Stem Cell Defects with an Aging-like Phenotype. *Cell Rep*. 2016; 16: 3181–3194. <https://doi.org/10.1016/j.celrep.2016.08.064> PMID: 27653684
42. Demirci S, Zeng J, Wu Y, Uchida N, Shen AH, Pellin D, et al. BCL11A enhancer-edited hematopoietic stem cells persist in rhesus monkeys without toxicity. *J Clin Invest*. 2020; 130. <https://doi.org/10.1172/JCI140189> PMID: 32897878
43. Ippolito GC, Dekker JD, Wang Y-H, Lee B-K, Shaffer AL 3rd, Lin J, et al. Dendritic cell fate is determined by BCL11A. *Proc Natl Acad Sci U S A*. 2014; 111: E998–1006. <https://doi.org/10.1073/pnas.1319228111> PMID: 24591644
44. Jackson JT, Ng AP, Shields BJ, Haupt S, Haupt Y, McCormack MP. Hhex induces promyelocyte self-renewal and cooperates with growth factor independence to cause myeloid leukemia in mice. *Blood Advances*. 2018; 2: 347. <https://doi.org/10.1182/bloodadvances.2017013243> PMID: 29453249
45. Kueh HY, Champhekar A, Nutt SL, Elowitz MB, Rothenberg EV. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science*. 2013; 341: 670. <https://doi.org/10.1126/science.1240831> PMID: 23868921
46. Li K, Wu Y, Li Y, Yu Q, Tian Z, Wei H, et al. Landscape and dynamics of the transcriptional regulatory network during natural killer cell differentiation. <https://doi.org/10.1016/j.gpb.2020.12.003> PMID: 33385611
47. Levanon D, Negreanu V, Lotem J, Bone KR, Brenner O, Leshkowitz D, et al. Transcription Factor Runx3 Regulates Interleukin-15-Dependent Natural Killer Cell Activation. *Mol Cell Biol*. 2014; 34: 1158. <https://doi.org/10.1128/MCB.01202-13> PMID: 24421391
48. Dandan Wang SM. Transcriptional Regulation of Natural Killer Cell Development and Functions. *Cancers*. 2020; 12. <https://doi.org/10.3390/cancers12061591> PMID: 32560225
49. Vellingiri B, Iyer M, Devi Subramaniam M, Jayaramayya K, Siama Z, Giridharan B, et al. Understanding the Role of the Transcription Factor Sp1 in Ovarian Cancer: from Theory to Practice. *Int J Mol Sci*. 2020; 21. <https://doi.org/10.3390/ijms21031153> PMID: 32050495
50. Smith-Raska MR, Arenzana TL, D’Cruz LM, Khodadadi-Jamayran A, Tsirigos A, Goldrath AW, et al. The Transcription Factor Zfx Regulates Peripheral T Cell Self-Renewal and Proliferation. *Front Immunol*. 2018; 9. <https://doi.org/10.3389/fimmu.2018.01482> PMID: 30022979
51. Zheng Q, Fu Q, Xu J, Gu X, Zhou H, Zhi C. Transcription factor E2F4 is an indicator of poor prognosis and is related to immune infiltration in hepatocellular carcinoma. *J Cancer*. 2021; 12: 1792–1803. <https://doi.org/10.7150/jca.51616> PMID: 33613768
52. Jackson SH, Yu CR, Mahdi RM, Ebong S, Ekwuagu CE. Dendritic cell maturation requires STAT1 and is under feedback regulation by suppressors of cytokine signaling. *J Immunol*. 2004; 172. <https://doi.org/10.4049/jimmunol.172.4.2307> PMID: 14764699
53. Webster B, Werneke SW, Zafirova B, This S, Col  n S, D  cembre E, et al. Plasmacytoid dendritic cells control dengue and Chikungunya virus infections via IRF7-regulated interferon responses. 2018 [cited 8 Jun 2021]. <https://doi.org/10.7554/eLife.34273> PMID: 29914621

54. Tussiwand R, Lee W-L, Murphy TL, Mashayekhi M, Wumesh KC, Albring JC, et al. Compensatory dendritic cell development mediated by BATF-IRF interactions. *Nature*. 2012; 490: 502. <https://doi.org/10.1038/nature11531> PMID: 22992524
55. Bhatlekar S, Fields JZ, Boman BM. Role of HOX Genes in Stem Cell Differentiation and Cancer. *Stem Cells Int*. 2018; 2018. <https://doi.org/10.1155/2018/3569493> PMID: 30154863
56. Abby Sarkar KH. The Sox Family of Transcription Factors: Versatile Regulators of Stem and Progenitor Cell Fate. *Cell Stem Cell*. 2013; 12: 15. <https://doi.org/10.1016/j.stem.2012.12.007> PMID: 23290134
57. Klein F, Mitrovic M, Roux J, Engdahl C, von Muenchow L, Alberti-Servera L, et al. The transcription factor Duxbl mediates elimination of pre-T cells that fail β -selection. *J Exp Med*. 2019; 216: 638. <https://doi.org/10.1084/jem.20181444> PMID: 30765463
58. Frison H, Giono G, Paméla Thébault, Fournier M, Labrecque N, Bijl JJ. Hoxb4 Overexpression in CD4 Memory Phenotype T Cells Increases the Central Memory Population upon Homeostatic Proliferation. *PLoS One*. 2013; 8: e81573. <https://doi.org/10.1371/journal.pone.0081573> PMID: 24324706
59. Szabo PA, Miron M, Farber DL. Location, location, location: Tissue resident memory T cells in mice and humans. *Science Immunology*. 2019; 4. <https://doi.org/10.1126/sciimmunol.aas9673> PMID: 30952804
60. Chong T, Luo MOL. Foxo Transcription Factors in T Cell Biology and Tumor Immunity. *Semin Cancer Biol*. 2018; 50: 13. <https://doi.org/10.1016/j.semcancer.2018.04.006> PMID: 29684436
61. Kim MV, Ouyang W, Liao W, Zhang MQ, Li MO. The Transcription Factor Foxo1 Controls Central-Memory CD8 T Cell Responses to Infection. *Immunity*. 2013. pp. 286–297. <https://doi.org/10.1016/j.immuni.2013.07.013> PMID: 23932570
62. Rao RR, Li Q, Mr GB, Shrikant PA. Transcription factor Foxo1 represses T-bet-mediated effector functions and promotes memory CD8(+) T cell differentiation. *Immunity*. 2012; 36. <https://doi.org/10.1016/j.immuni.2012.01.015> PMID: 22425248
63. Delpoux A, Michelini RH, Verma S, Lai CY, Omilusik KD, Utschneider DT, et al. Continuous activity of Foxo1 is required to prevent anergy and maintain the memory state of CD8 + T cells. *J Exp Med*. 2018; 215. <https://doi.org/10.1084/jem.20170697> PMID: 29282254
64. Hess MR, Doedens AL, Goldrath AW, Hedrick SM. Differentiation of CD8 memory T cells depends on Foxo1. *J Exp Med*. 2013; 210. <https://doi.org/10.1084/jem.20130392> PMID: 23712431
65. Delpoux A, Lai CY, Hedrick SM, Doedens AL. FOXO1 opposition of CD8 + T cell effector programming confers early memory properties and phenotypic diversity. *Proc Natl Acad Sci U S A*. 2017; 114. <https://doi.org/10.1073/pnas.1618916114> PMID: 28973925
66. Utschneider DT, Delpoux A, Wieland D, Huang X, Lai CY, Hofmann M, et al. Active Maintenance of T Cell Memory in Acute and Chronic Viral Infection Depends on Continuous Expression of FOXO1. *Cell Rep*. 2018; 22. <https://doi.org/10.1016/j.celrep.2018.03.020> PMID: 29590615
67. Gray SM, Amezquita RA, Guan T, Kleinstein SH, Kaech SM. Polycomb Repressive Complex 2-Mediated Chromatin Repression Guides Effector CD8 + T Cell Terminal Differentiation and Loss of Multipotency. *Immunity*. 2017; 46. <https://doi.org/10.1016/j.immuni.2017.03.012> PMID: 28410989
68. Quintana FJ, Jin H, Burns EJ, Nadeau M, Yeste A, Kumar D, et al. Aiolos promotes TH17 differentiation by directly silencing Il2 expression. *Nat Immunol*. 13: 770. <https://doi.org/10.1038/ni.2363> PMID: 22751139
69. Clambey ET, Collins B, Young MH, Eberlein J, David A, Kappler JW, et al. The Ikaros Transcription Factor Regulates Responsiveness to IL-12 and Expression of IL-2 Receptor Alpha in Mature, Activated CD8 T Cells. *PLoS One*. 2013; 8: e57435. <https://doi.org/10.1371/journal.pone.0057435> PMID: 23483882
70. Garrett-Sinha LA. Review of Ets1 structure, function, and roles in immunity. *Cell Mol Life Sci*. 2013; 70: 3375. <https://doi.org/10.1007/s00018-012-1243-7> PMID: 23288305
71. Chou C, Li MO. Tissue-Resident Lymphocytes Across Innate and Adaptive Lineages. *Front Immunol*. 2018; 9. <https://doi.org/10.3389/fimmu.2018.02104> PMID: 30298068
72. Bramhall M, Rodrigues G, Christo S, Mackay L, Zaph C. T cell-intrinsic expression of HIC1 links retinoic acid to tissue residency. *The Journal of Immunology*. 2020; 204: 155.3–155.3.
73. Burrows K, Antignano F, Bramhall M, Chenery A, Scheer S, Korinek V, et al. The transcriptional repressor HIC1 regulates intestinal immune homeostasis. *Mucosal Immunol*. 2017; 10: 1518–1528. <https://doi.org/10.1038/mi.2017.17> PMID: 28327618
74. Crowl JT, Heeg M, Ferry A, Milner JJ, Omilusik KD, Toma C, et al. Tissue-resident memory CD8 T cells possess unique transcriptional, epigenetic and functional adaptations to different tissue environments. *Nat Immunol*. 2022; 23: 1121–1131. <https://doi.org/10.1038/s41590-022-01229-8> PMID: 35761084

75. Zaid A, Mackay LK, Rahimpour A, Braun A, Veldhoen M, Carbone FR, et al. Persistence of skin-resident memory T cells within an epidermal niche. *Proc Natl Acad Sci U S A*. 2014; 111: 5307–5312. <https://doi.org/10.1073/pnas.1322292111> PMID: 24706879
76. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet*. 2019; 51: 1494–1505. <https://doi.org/10.1038/s41588-019-0505-9> PMID: 31570894
77. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016; 48: 1193–1203. <https://doi.org/10.1038/ng.3646> PMID: 27526324
78. Cobaleda C, Schebesta A, Delogu A, Busslinger M. Pax5: the guardian of B cell identity and function. *Nat Immunol*. 2007; 8: 463–470. <https://doi.org/10.1038/ni1454> PMID: 17440452
79. Bullerwell CE, Robichaud PP, Deprez PML, Joy AP, Wajnberg G, D'Souza D, et al. EBF1 drives hall-mark B cell gene expression by enabling the interaction of PAX5 with the MLL H3K4 methyltransferase complex. *Sci Rep*. 2021; 11: 1–14.
80. Liu GJ, Jaritz M, Wöhner M, Agerer B, Bergthaler A, Malin SG, et al. Repression of the B cell identity factor Pax5 is not required for plasma cell development. *J Exp Med*. 2020; 217. <https://doi.org/10.1084/jem.20200147> PMID: 32780801
81. Kreslavsky T, Vilagos B, Tagoh H, Poliakova DK, Schwickert TA, Wöhner M, et al. Essential role for the transcription factor Bhlhe41 in regulating the development, self-renewal and BCR repertoire of B-1a cells. *Nat Immunol*. 2017; 18: 442–455. <https://doi.org/10.1038/ni.3694> PMID: 28250425
82. Samten B. Regulation of B-1a cells: another novel function of the basic helix-loop-helix transcriptional regulator BHLHE41. *Cell Mol Immunol*. 2017; 14: 802–804. <https://doi.org/10.1038/cmi.2017.75> PMID: 29026219
83. Postnatal regulation of B-1a cell development and survival by the CIC-PER2-BHLHE41 axis. *Cell Rep*. 2022; 38: 110386. <https://doi.org/10.1016/j.celrep.2022.110386> PMID: 35172136
84. Lien C, Fang C-M, Huso D, Livak F, Lu R, Pitha PM. Critical role of IRF-5 in regulation of B-cell differentiation. *Proc Natl Acad Sci U S A*. 2010; 107: 4664–4668. <https://doi.org/10.1073/pnas.0911193107> PMID: 20176957
85. Fang C-M, Roy S, Nielsen E, Paul M, Maul R, Paun A, et al. Unique contribution of IRF-5-Ikaros axis to the B-cell IgG2a response. *Genes & Immunity*. 2012; 13: 421–430.
86. Dejean AS, Joulia E, Walzer T. The role of Eomes in human CD4 T cell differentiation: A question of context. *European journal of immunology*. 2019. pp. 38–41. <https://doi.org/10.1002/eji.201848000> PMID: 30536524
87. Llaó-Cid L, Roessner PM, Chapaprieta V, Öztürk S, Roeder T, Bordsa M, et al. EOMES is essential for antitumor activity of CD8+ T cells in chronic lymphocytic leukemia. *Leukemia*. 2021; 35: 3152–3162. <https://doi.org/10.1038/s41375-021-01198-1> PMID: 33731848
88. Li J, He Y, Hao J, Ni L, Dong C. High Levels of Eomes Promote Exhaustion of Anti-tumor CD8+ T Cells. *Front Immunol*. 2018; 9. <https://doi.org/10.3389/fimmu.2018.02981> PMID: 30619337
89. Abboud G, Stanfield J, Tahiliani V, Desai P, Hutchinson TE, Lorentsen KJ, et al. Transcription Factor Bcl11b Controls Effector and Memory CD8 T cell Fate Decision and Function during Poxvirus Infection. *Front Immunol*. 2016; 7. <https://doi.org/10.3389/fimmu.2016.00425> PMID: 27790219
90. Sottile R, Panjwani MK, Lau CM, Daniyan AF, Tanaka K, Barker JN, et al. Human cytomegalovirus expands a CD8 T cell population with loss of expression and gain of NK cell identity. *Sci Immunol*. 2021; 6: eabe6968.
91. Kastner P, Chan S, Vogel WK, Zhang L-J, Topark-Ngarm A, Golonzhka O, et al. Bcl11b represses a mature T-cell gene expression program in immature CD4+CD8+ thymocytes. *Eur J Immunol*. 2010; 40: 2143. <https://doi.org/10.1002/eji.200940258> PMID: 20544728
92. Shan Q, Li X, Chen X, Zeng Z, Zhu S, Gai K, et al. Tcf1 and Lef1 provide constant supervision to mature CD8+ T cell identity and function by organizing genomic architecture. *Nat Commun*. 2021; 12: 1–20.
93. Xing S, Li F, Zeng Z, Zhao Y, Yu S, Shan Q, et al. Tcf1 and Lef1 transcription factors establish CD8+ T cell identity through intrinsic HDAC activity. *Nat Immunol*. 2016; 17: 695. <https://doi.org/10.1038/ni.3456> PMID: 27111144
94. Wang W, Xia X, Mao L, Wang S. The CCAAT/Enhancer-Binding Protein Family: Its Roles in MDSC Expansion and Function. *Front Immunol*. 2019; 0. <https://doi.org/10.3389/fimmu.2019.01804> PMID: 31417568
95. Tamura A, Hirai H, Yokota A, Kamio N, Sato A, Shoji T, et al. C/EBP β is required for survival of Ly6C – monocytes. *Blood*. 2017; 130: 1809. <https://doi.org/10.1182/blood-2017-03-772962> PMID: 28807982

96. Knudsen KJ, Rehn M, Hasemann MS, Rapin N, Bagger FO, Ohlsson E, et al. ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev.* 2015; 29: 1915. <https://doi.org/10.1101/gad.268409.115> PMID: 26385962
97. Rothenberg EV. Erg in stem cells: a function emerges. *Nat Immunol.* 2008; 9: 714–716. <https://doi.org/10.1038/ni0708-714> PMID: 18563078
98. Seifert A, Werheid DF, Knapp SM, Tobiasch E. Role of Hox genes in stem cell differentiation. *World J Stem Cells.* 2015; 7: 583. <https://doi.org/10.4252/wjsc.v7.i3.583> PMID: 25914765
99. Dou DR, Calvanese V, Sierra MI, Nguyen AT, Minasian A, Saarikoski P, et al. Medial HOXA genes demarcate haematopoietic stem cell fate during human development. *Nat Cell Biol.* 2016; 18: 595–606. <https://doi.org/10.1038/ncb3354> PMID: 27183470
100. Ng ES, Azzola L, Bruveris FF, Calvanese V, Phipson B, Vlahos K, et al. Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nat Biotechnol.* 2016; 34: 1168–1179. <https://doi.org/10.1038/nbt.3702> PMID: 27748754
101. Cumano A, Berthault C, Ramond C, Petit M, Golub R, Bandeira A, et al. New Molecular Insights into Immune Cell Development. *Annual Review of Immunology.* 2019. pp. 497–519. <https://doi.org/10.1146/annurev-immunol-042718-041319> PMID: 31026413
102. O'Shea JJ, Lahesmaa R, Vahedi G, Laurence A, Kanno Y. Genomic views of STAT function in CD4+ T helper cell differentiation: new technology brings new insights and new questions. *Nat Rev Immunol.* 2011; 11: 239.
103. Ho I-C, Tai T-S, Pai S-Y. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat Rev Immunol.* 2009; 9: 125. <https://doi.org/10.1038/nri2476> PMID: 19151747
104. Nechanitzky R, Akbas D, Scherer S, Györy I, Hoyle T, Ramamoorthy S, et al. Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat Immunol.* 2013; 14: 867–875. <https://doi.org/10.1038/ni.2641> PMID: 23812095
105. Hagman J, Ramírez J, Lukin K. B Lymphocyte Lineage Specification, Commitment and Epigenetic Control of Transcription by Early B Cell Factor 1. *Curr Top Microbiol Immunol.* 2012; 356: 17. https://doi.org/10.1007/82_2011_139 PMID: 21735360
106. Chung HK, McDonald B, Kaech SM. The architectural design of CD8+ T cell responses in acute and chronic infection: Parallel structures with divergent fates. *J Exp Med.* 2021; 218. <https://doi.org/10.1084/jem.20201730> PMID: 33755719
107. Villarino AV, Kanno Y, O'Shea JJ. Mechanisms of Jak/STAT signaling in immunity and disease. *J Immunol.* 2015; 194: 21. <https://doi.org/10.4049/jimmunol.1401867> PMID: 25527793
108. Donnard E, Vangala P, Afik S, McCauley S, Nowosielska A, Kucukural A, et al. Comparative analysis of immune cells reveals a conserved regulatory lexicon. *Cell systems.* 2018; 6: 381. <https://doi.org/10.1016/j.cels.2018.01.002> PMID: 29454939
109. Akira S. Functional Roles of STAT Family Proteins: Lessons from Knockout Mice. *Stem Cells.* 1999; 17: 138–146. <https://doi.org/10.1002/stem.170138> PMID: 10342556
110. The Speckled Protein (SP) Family: Immunity's Chromatin Readers. *Trends Immunol.* 2020; 41: 572–585. <https://doi.org/10.1016/j.it.2020.04.007> PMID: 32386862
111. Baur F, Nau K, Sadic D, Allweiss L, Elsässer H-P, Gillemans N, et al. Specificity Protein 2 (Sp2) Is Essential for Mouse Development and Autonomous Proliferation of Mouse Embryonic Fibroblasts. *PLoS One.* 2010; 5. <https://doi.org/10.1371/journal.pone.0009587> PMID: 20221402
112. Steuernagel L, Meckbach C, Heinrich F, Zeidler S, Schmitt AO, Gültas M. Computational identification of tissue-specific transcription factor cooperation in ten cattle tissues. *PLoS One.* 2019; 14: e0216475. <https://doi.org/10.1371/journal.pone.0216475> PMID: 31095599
113. Omilusik KD, Goldrath AW. Remembering to remember: T cell memory maintenance and plasticity. *Curr Opin Immunol.* 2019; 58. <https://doi.org/10.1016/j.coi.2019.04.009> PMID: 31170601
114. Chang JT, Wherry EJ, Goldrath AW. Molecular regulation of effector and memory T cell differentiation. *Nat Immunol.* 2014; 15. <https://doi.org/10.1038/ni.3031> PMID: 25396352
115. Smith SA, Choi SH, Davis-Harrison R, et al. Polyphosphate exerts differential effects on bloodclotting, depending on polymer size. *Blood.* 2010; 116(20):4353–4359. *Blood.* 2011; 117: 3477–3477. <https://doi.org/10.1182/blood-2010-01-266791> PMID: 20709905
116. Hayashi K, Yoshida K, Matsui Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature.* 2005; 438: 374–378. <https://doi.org/10.1038/nature04112> PMID: 16292313
117. Kinjyo I, Qin J, Tan S-Y, Wellard CJ, Mrass P, Ritchie W, et al. Real-time tracking of cell cycle progression during CD8 + effector and memory T-cell differentiation. *Nat Commun.* 2015; 6: 1–13. <https://doi.org/10.1038/ncomms7301> PMID: 25709008

118. Yan M, Yang X, Wang H, Shao Q. The critical role of histone lysine demethylase KDM2B in cancer. *Am J Transl Res*. 2018; 10. Available: <https://pubmed.ncbi.nlm.nih.gov/30210666/> PMID: 30210666
119. Xu Z, Rao Y, Huang Y, Zhou T, Feng R, Xiong S, et al. Efficient Strategies for Microglia Replacement in the Central Nervous System. *Cell Reports*. 2020. p. 108443. <https://doi.org/10.1016/j.celrep.2020.108443> PMID: 33238120
120. Qiu Z, Chu TH, Sheridan BS. TGF- β : Many Paths to CD103 CD8 T Cell Residency. *Cells*. 2021; 10. <https://doi.org/10.3390/cells10050989> PMID: 33922441