nature immunology

Article

https://doi.org/10.1038/s41590-023-01599-7

Intrinsically disordered domain of transcription factor TCF-1 is required for T cell developmental fidelity

Received: 11 October 2022

Accepted: 20 July 2023

Published online: 17 August 2023

Check for updates

Naomi Goldman[1,2,3], Aditi Chandra[1,2,3,15], Isabelle Johnson[1,2,3,15], Matthew A. Sullivan[2,4,5], Abhijeet R. Patil[1,2,3], Ashley Vanderbeck[2,6], Atishay Jay[1,2,3,7], Yeqiao Zhou[2,3,8], Emily K. Ferrari[1,2,3], Leland Mayne[9], Jennifer Aguilan[10], Hai-Hui Xue[11,12], Robert B. Faryabi[2,3,8], E. John Wherry[2,4,13,14], Simone Sidoli[10], Ivan Maillard[2,6,14] & Golnaz Vahedi[1,2,3,13,14]

In development, pioneer transcription factors access silent chromatin to reveal lineage-specific gene programs. The structured DNA-binding domains of pioneer factors have been well characterized, but whether and how intrinsically disordered regions affect chromatin and control cell fate is unclear. Here, we report that deletion of an intrinsically disordered region of the pioneer factor TCF-1 (termed L1) leads to an early developmental block in T cells. The few T cells that develop from progenitors expressing TCF-1 lacking L1 exhibit lineage infidelity distinct from the lineage diversion of TCF-1-deficient cells. Mechanistically, L1 is required for activation of T cell genes and repression of GATA2-driven genes, normally reserved to the mast cell and dendritic cell lineages. Underlying this lineage diversion, L1 mediates binding of TCF-1 to its earliest target genes, which are subject to repression as T cells develop. These data suggest that the intrinsically disordered N terminus of TCF-1 maintains T cell lineage fidelity.

The induction of tissue-specific gene expression programs depends on the reconfiguration of silent chromatin and formation of accessible regulatory elements. Some proteins in the class of transcription factors (TFs) are endowed with the capacity to reprogram silent chromatin and are hence critical for cell fate determination. These TFs, also referred to as pioneer factors[1,2], are thought to target DNA sequences frequently summarized as binding motifs[3] through their DNA-binding domains and can recruit proteins with enzymatic activities to remodel silent

[1]Department of Genetics, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [2]Institute for Immunology and Immune Health, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [3]Epigenetics Institute, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [4]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [5]Graduate Group in Biochemistry and Molecular Biophysics, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [6]Division of Hematology-Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [7]Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [8]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [9]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [10]Department of Biochemistry, Albert Einstein School of Medicine, New York City, NY, USA. [11]Center for Discovery and Innovation, Hackensack University Medical Center, Nutley, NJ, USA. [12]New Jersey Veterans Affairs Health Care System, East Orange, NJ, USA. [13]Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [14]Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Philadelphia, PA, USA. [15]These authors contributed equally: Aditi Chandra, Isabelle Johnson. e-mail: vahedi@pennmedicine.upenn.edu

chromatin. Despite detailed knowledge of TF domains required for DNA-binding specificities, the extent to which non-DNA-binding domains of TFs are critical for lineage determination and chromatin reorganization remains largely understudied.

Here we interrogated whether and how non-DNA-binding domains of a pioneer factor are required to reprogram the chromatin and determine cell fate using T cells as a model. In T cell development, lymphoid progenitors enter the thymus and receive signals from delta-like 4 (DLL4) Notch ligands to initiate the process of commitment to the T cell lineage. The TCF-1 protein, encoded by the transcription factor 7 (*Tcf7*) gene, has been characterized as a lineage-determining TF for T cells since *Tcf7* deletion disrupts T cell development[4]. TCF-1, whose expression rises precipitously as soon as bone marrow (BM)-derived progenitors enter the thymus, has been characterized as a pioneer factor[5,6]. Moreover, TCF-1 can promote long-range interactions across topologically associating domains[7]. Despite these mechanistic insights into the role of TCF-1 in T cell development, whether non-DNA-binding domains of TCF-1 are engaged in shaping the chromatin landscape of T cells is unknown.

Multiple major isoforms of TCF-1 in mouse and human T cells[4] have been characterized including the long isoforms that contain an N-terminal β-catenin-binding domain and respond to Wnt signaling. Both short and long isoforms of TCF-1 are sufficient to initiate and sustain T cell development[8,9]. The high-mobility group (HMG)-box DNA-binding domain of TCF-1 and closely related factor LEF-1 have largely been studied due to the solved crystal structure of the HMG-box[10]. Moreover, an intrinsic histone deacetylase domain within the N terminus of TCF-1 has been linked to the protein's ability to suppress $CD4^+$ lineage genes in $CD8^+$ T cells[11]. Like many eukaryotic TFs whose non-DNA-binding domains are highly disordered[12] and exhibit conformational heterogeneity[13], TCF-1 is predicted to be highly disordered outside the HMG-box DNA-binding domain. The low complexity of TCF-1's disordered domain limits the feasibility of crystallography studies and the predictive power of algorithms like AlphaFold[14]. Disordered regions often harbor TF effector domains whose canonical role is to interact with co-activators or co-repressors to remodel the chromatin[15]. The structure and function of the low complexity disordered regions of many TFs including TCF-1 remains largely understudied.

In this study, we examined the distinct roles of regions within the intrinsically disordered N terminus of TCF-1 in primary developing mouse T cells and a pro-T cell line. An N-terminal region (termed L1) was necessary for efficient transition between early T lineage progenitors in the double-negative 1 (DN1) and DN2 subsets. The L1 region was required for development and lineage fidelity. Cells that developed without the L1 region of TCF-1 (ΔL1) expressed mast cell genes and exhibited epigenetic reprogramming downstream of *Gata2* de-repression. The L1 domain was principally required for the binding of TCF-1 to its earliest target genes, which were subject to repression as T cells develop. Additionally, the L1 domain could be functionally replaced with a heterologous disordered domain of B cell pioneer factor, the early B cell factor 1 (EBF1), to rescue both early binding and T cell development. However, the L1 region was no longer required once T cells reached a post-commitment stage. These studies suggest the functional relevance of TF effector domains and the importance of careful dissection of protein function through mutational approaches at multiple stages of development.

Results

The N terminus of TCF-1 is intrinsically disordered

We constructed an alignment of the long isoform of murine TCF-1 (P45), referred to as 'wild-type TCF-1', with 150 vertebrate homologs and plotted the evolutionary conservation score of each amino acid position[16]. The most conserved positions across species fell in the HMG-box DNA-binding domain; however, a large non-DNA-binding domain within the N terminus demonstrated moderate conservation (Fig. 1a). Relying on a quantitative method to predict TCF-1 structure, we utilized the predictor of natural disordered regions (VSL2 in PONDR[17]) and plotted the disorder score at each residue (Fig. 1a). The DNA-binding domain had a low PONDR score indicative of ordered residues, while mostly disordered amino acids spanned the surrounding sequence.

To interrogate the protein's secondary structure in vitro, we expressed and purified recombinant TCF-1 protein from *Escherichia coli* (Extended Data Fig. 1a) and used hydrogen–deuterium exchange coupled with mass spectrometry (HX–MS). The peptide bond amide ^{1}H ('H') of each amino acid, except proline, undergoes exchange in aqueous environments with solvent-derived hydrogen at variable rates that depend on the pH, temperature and flanking amino acid side chains. In proteins, the chemical exchange rate is slowed by hydrogen bonded structure[18]. Low structural stability or highly dynamic regions of proteins exhibit less protection from exchange, and thus faster exchange rates, than regions with stable secondary structure. HX–MS measures this exchange over time in deuterium (^{2}H, 'D')-containing buffer. We observed very rapid exchange for all measured N-terminal peptides, with nearly complete exchange within 4 s of deuterium oxide addition at a measured sample pH (pH_{meas}) of 7.0 and temperature of ~4 °C, whereas peptides within the HMG-box domain underwent much slower exchange (Fig. 1b, Extended Data Fig. 1b–d and Supplementary Table 1). We repeated HX measurements at a lower pH_{meas} of 6.0 and 5.0, where the H-to-D exchange rate is 10-fold and 100-fold lower, respectively (Extended Data Fig. 1b). The exchange versus time relationship for N-terminal peptides very closely approximated that predicted for each respective peptide sequence if residues were dynamically disordered random coil and not subject to any protection[19,20] (Extended Data Fig. 1e), whereas HMG-box peptides were protected relative to this prediction across all pH_{meas} values (Fig. 1c). Collectively, the N-terminal region of TCF-1 lacks stable secondary structure, consistent with an intrinsically disordered region (IDR).

To examine if the N terminus of TCF-1 plays any role in T cell development, we utilized a collection of mutant TCF-1 constructs[11] in which internal deletions were made tiling the protein from the N terminus to the DNA-binding domain and labeled them sequentially as ΔL1 to ΔL7 (Fig. 1d). We also deleted the DNA-binding domain of TCF-1 to generate

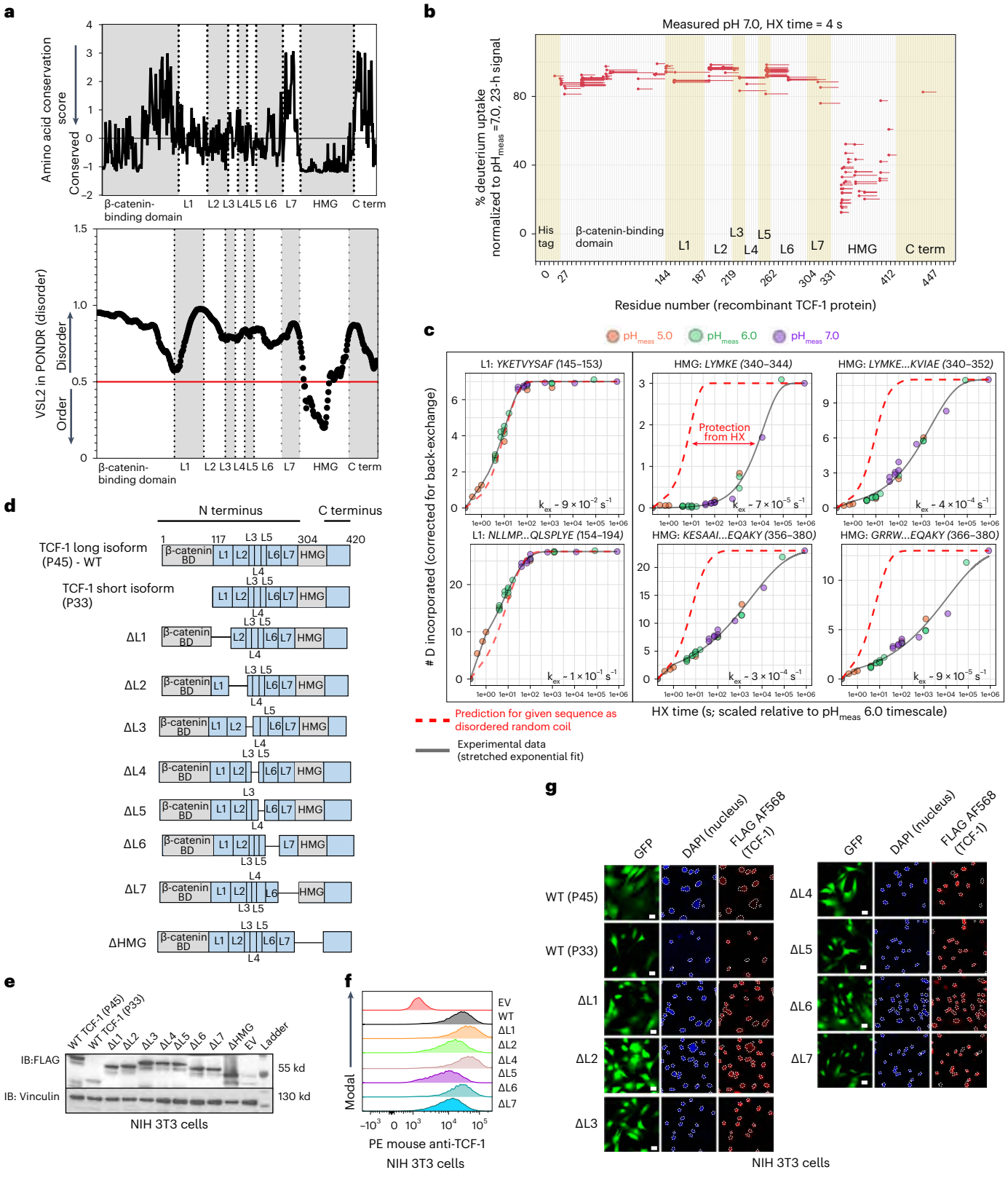
Fig. 1 | The N terminus of TCF-1 is intrinsically disordered. a, Profile of amino acid conservation score across residues of mouse TCF-1 protein utilizing ConSurf-DB and MAFFT alignment of 150 vertebrate homologous sequences (top). Profile of VSL2 score across residues in mouse TCF-1 utilizing the predictor of natural disorder regions (PONDR)[46] (bottom). **b**, Percentage of deuterium uptake at 4 s and measured sample pH of 7.0 for exchange (normalized to measured deuterium content after 23 h of H-to-D exchange) for each TCF-1 peptide observation (different peptide charge states treated as separate observations). Line represents mean value of $n = 2$ technical replicates. **c**, Number of incorporated deuterium (D) atoms (corrected for back exchange) versus H-to-D exchange (HX) time for each indicated peptide observation as representative examples of the time-dependent HX behavior of L1 and HMG-box domains. HX time for pH_{meas} 5.0 and 7.0 was scaled by a factor of 10 relative to a pH_{meas} 6.0 timescale to directly compare all data. Solid line corresponds to fit of data to stretched exponential function used for estimating approximate experimental peptide-level HX rate constant k_{ex} (Methods). The red dashed line corresponds to the predicted behavior for each indicated peptide sequence as random coil (calculated as described in refs. 19,20,47). Time, 0-s value is assumed as 0 D. **d**, Schematic of wild-type (WT) isoforms of TCF-1 (P33 and P45) and internal deletions. **e**, Immunoblot (IB) analysis of NIH 3T3 cells transduced with FLAG-tagged WT TCF-1 and mutant TCF-1 constructs with internal deletions (ΔL1–ΔL7). Vinculin was used as a loading control. **f**, Representative histogram of flow cytometry depicts TCF-1 expression with intracellular anti-mouse TCF-1 staining in NIH 3T3 cells transduced with EV, WT TCF-1 (WT) and mutant TCF-1 constructs with internal deletions (ΔL1–ΔL7). **g**, Representative immunofluorescence depicts nuclear localization of FLAG-tagged WT and mutant TCF-1 with internal deletions. A nuclear mask is indicated with a dotted line in DAPI images and superimposed to the FLAG AF568 channel to indicate nuclear localization of FLAG-tagged WT and mutant TCF-1. Scale bar, 10 μm.

https://doi.org/10.1038/s41590-023-01599-7

a ΔHMG mutant construct. To confirm that these internal deletions did not disrupt protein localization, stability or expression, we visualized the individual mutants with immunofluorescence, immunoblotting and intracellular flow cytometry to detect FLAG-tagged nuclear constructs in transduced National Institutes of Health (NIH) 3T3 cells (Fig. 1e–g). Thus, this series of truncation mutants of TCF-1 enabled us to test the role of TCF-1 domains in T cell development.

Efficient DN1-to-DN2 transition requires the L1 domain of TCF-1

Our study focused on the murine pro-T cell program that is induced in developing progenitors divided by cell surface markers into early T cell precursors from the DN1, DN2 and DN3 subsets. We next tested the ability of each TCF-1 mutant to rescue T cell development in primary TCF-1-deficient T cell progenitors by modeling T cell development in vitro[21–27]. Differentiation of primary mouse progenitor cells into T



https://doi.org/10.1038/s41590-023-01599-7

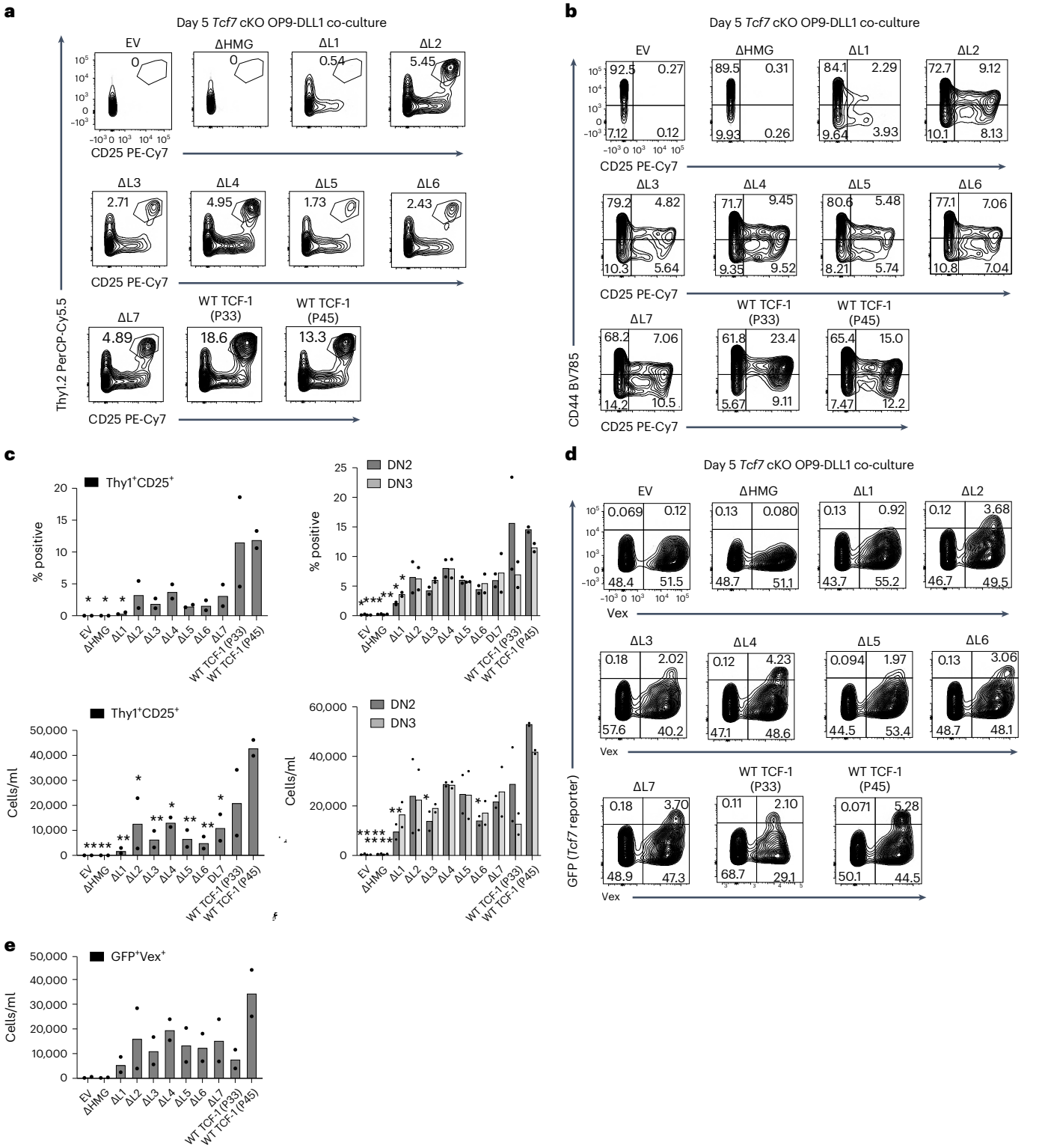


Fig. 2 | Loss of TCF-1's L1 domain limits DN1-to-DN2 transition. a, Identification of Thy1$^+$CD25$^+$ cells in OP9-DLL1 co-cultures of *Tcf7* cKO cells transduced with EV, WT TCF-1 or mutant TCF-1 (ΔL1–ΔL7 and ΔHMG) on day 5 after in vitro differentiation. Data are representative of at least three independent experiments. All cells were pre-gated on SSC-A/FSC-A, singlets, live cell (viability$^-$), CD45$^+$, transduced (vex$^+$). **b**, Detection of DN1, DN2 and DN3 cells by CD44 and CD25 surface expression in co-cultures described in **a**. Data are representative of at least three independent experiments. All cells were pre-gated on SSC-A/FSC-A, singlets, live cell (viability$^-$), CD45$^+$, transduced (vex$^+$). **c**, Quantification of frequency and number of Thy1$^+$CD25$^+$ cells (left), CD44$^+$CD25$^+$ DN2s, and CD44$^-$CD25$^+$ DN3 cells (right) from *Tcf7* cKO cells on day 5 after in vitro differentiation on OP9-DLL1 cells. Bars represent the mean from $n = 2$ independent animals. Individual replicates are represented by data points. *P* values were calculated by one-way analysis of variance (ANOVA) followed by Dunnett's multiple-comparison test with WT TCF-1 (P45) as a control. *$P \le 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.001$. **d**, Representative flow cytometric analysis identifying transduced (vex$^+$) GFP$^+$ cells (*Tcf7* eGFP reporter) of *Tcf7* cKO cells on day 5 after in vitro differentiation on OP9-DLL1 co-cultures. **e**, Quantification of frequency of vex$^+$GFP$^+$ (*Tcf7* eGFP reporter) cells in OP9-DLL1 co-cultures on day 5 as described in **d**. All cells were pre-gated on SSC-A/FSC-A, singlets, live cell (viability$^-$) and CD45$^+$. Bars represent the mean from $n = 2$ independent animals. Individual replicates are represented by data points.

https://doi.org/10.1038/s41590-023-01599-7

lymphocytes can be achieved in vitro using a BM-derived stromal cell line that ectopically expresses the Notch ligand (OP9-DLL1)[28,29] and closely mimics T cell development in vivo[24,26]. We first generated TCF-1 conditional knockout mice (*Tcf7* cKO) by breeding Vav-iCre mice[30] with *Tcf7* eGFP reporter mice[31] in which exon 2 of *Tcf7* is floxed and an eGFP reporter cassette is inserted into the first intron. Cre$^+$ *Tcf7* cKO mice displayed altered T cell development in the thymus and lacked thymic expression of all TCF-1 protein isoforms[8,32]. To accommodate the eGFP reporter in *Tcf7* cKO mice, we cloned all mutant TCF-1 constructs onto a backbone with violet-excited fluorescent protein (MSCV-IRES-Vex). Expectedly[32], neither Lin$^-$Sca1$^+$ckit$^+$ (LSK) BM cells nor ckit$^+$ BM progenitors from *Tcf7* cKO mice generated any Thy1$^+$CD25$^+$ T cells after co-culture on OP9-DLL1 cells for 5 or 13 d (Fig. 2a–c and Extended Data Fig. 2a–c).

We characterized early T cells between the DN2 and DN3 stages as Thy1$^+$CD25$^+$ and resolved DN2s and DN3s as CD25$^+$CD44$^+$ and CD25$^+$CD44$^-$, respectively. Rescue of T cell development in *Tcf7* cKO progenitors with retroviral transduction of the P45 or P33 isoforms of TCF-1 was evident after co-culture on OP9-DLL1 for 5 d (Fig. 2a–c). Because T cell development is accelerated in wild-type progenitors transduced with TCF-1 (ref. 32), we also transduced wild-type ckit$^+$ BM progenitors with control empty vector (EV) or wild-type TCF-1 and characterized the extent of T cell development in these cultures. Co-cultures led to robust T cell development among wild-type TCF-1-transduced cells at both days 5 and 13, while development of untransduced (GFP$^-$) progenitors resulted in fewer Thy1$^+$CD25$^+$ cells (Extended Data Fig. 2d,e). OP9–DLL4 co-cultures recapitulated OP9-DLL1 results (Extended Data Fig. 2f), while in the absence of Notch ligand the OP9-control co-cultures failed to give rise to developing T cells despite overexpression of wild-type TCF-1 at either day 5 or day 13 (Extended Data Fig. 2g). Together, we established a system to evaluate the functional relevance of TCF-1 domains.

We aimed to test how retroviral transduction of mutant TCF-1 constructs in parallel with wild-type TCF-1, EV and ΔHMG controls in wild-type and *Tcf7* cKO progenitors affected T cell development. We ensured that the levels of transduction of wild-type and mutant TCF-1 were comparable to those detected in wild-type pro-T cells (Extended Data Fig. 2h). Although most TCF-1 mutants restricted T cell development to various degrees, *Tcf7* cKO progenitors transduced with ΔL1 demonstrated a major defect in progression toward the DN2 and DN3 stages at both days 5 and 13 (Fig. 2a–c and Extended Data Fig. 2a–c). Corroborating this finding, wild-type progenitors transduced with ΔL1 showed no substantial increase in DN2 proportions over that seen in untransduced cells, further demonstrating that the defect was cell intrinsic and not dominant negative (Extended Data Fig. 2d,e). Furthermore, ΔL1 co-cultures did not generate an increased proportion of alternative lineage B220$^+$ cells but showed an increased percentage of CD11b$^+$ cells (Extended Data Fig. 2i,j). We next exploited the eGFP fluorescent reporter in *Tcf7* cKO mice to assess the ability of ΔL1 to transactivate the endogenous *Tcf7* locus. At day 5 in wild-type TCF-1-transduced co-cultures, the presence of Vex$^+$ GFP reporter-positive cells suggested the activation of endogenous *Tcf7* transcriptional activity with transduction of either full-length isoform of TCF-1 (P45 or P33; Fig. 2d,e). Intriguingly, ΔL1 co-cultures showed very few GFP reporter-positive Vex$^+$ cells akin to levels seen in EV and ΔHMG control co-cultures lacking TCF-1, suggesting limited transcriptional activity at the endogenous *Tcf7* locus (Fig. 2d,e). Together, the L1 region of TCF-1 is necessary for efficient transition from DN1 to DN2 stages.

L1 is required for expression of T cell identity genes

We profiled the transcriptomes of DN1s and DN2s using bulk RNA sequencing (RNA-seq) and compared these populations to wild-type, EV-, ΔL7- and ΔHMG-transduced cells (Fig. 3a and Supplementary Table 1). Dimensionality reduction of RNA-seq data using principal component analysis (PCA) separated DN1s and DN2s of all conditions along the first principal component (PC1; Fig. 3b). EV- and ΔHMG-expressing DN1s clustered closely together and were separated from other conditions (Fig. 3b). Rescue of development with wild-type TCF-1-, ΔL1- or ΔL7- compared to EV-transduced cells led to modest differences between DN1s across conditions (Fig. 3c). Intriguingly, we observed a significant de-repression of over 600 genes and a reduction in expression of around 130 genes in ΔL1-expressing DN2s compared to wild-type TCF-1-transduced counterparts (Fig. 3d and Supplementary Table 1). ΔL7-transduced DN2s showed much fewer differential genes compared to wild-type TCF-1-expressing DN2s (121 genes up and 38 genes down; Fig. 3d). Among the significantly downregulated genes in ΔL1-transduced DN2s compared to wild-type TCF-1-expressing DN2s, we identified numerous T cell identity genes including *Gata3*, *Bcl11b*, *Lck*, *Lef1*, *Thy1*, *Il2rb*, *Rag2*, *CD3g* and *Cd3d* (Fig. 3e,f). Hence, transcriptional divergence between ΔL1- and wild-type TCF-1-expressing T cell progenitors occurs after the DN1 stage as cells enter the DN2 stage and that ΔL1-expressing DN2s have significantly reduced expression of T cell identity genes.

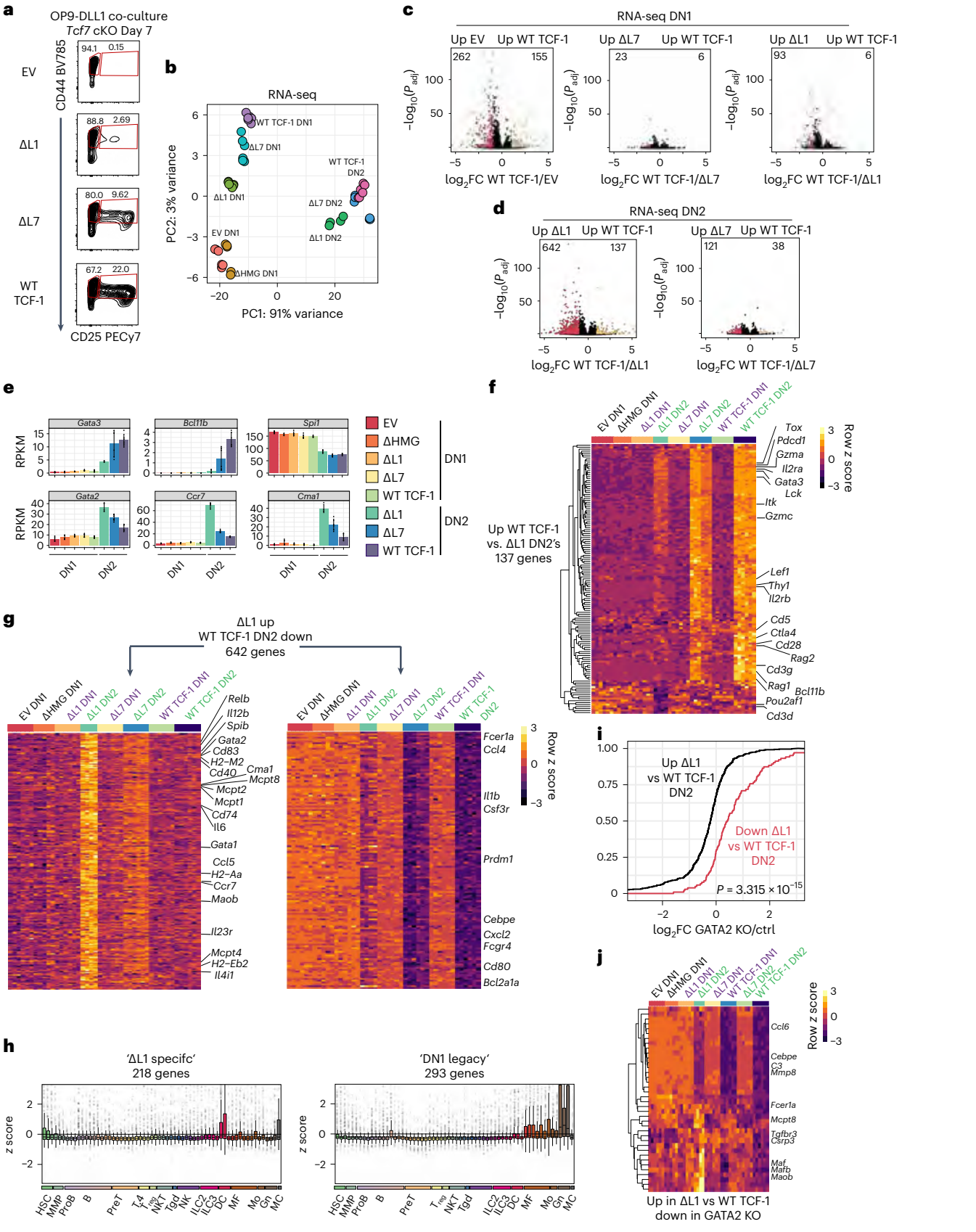
Loss of L1 identifies a hidden gene signature

Among the genes with significantly upregulated expression in ΔL1-expressing DN2s compared to wild-type TCF-1-expressing counterparts, we found genes enriched for inflammation, chemotaxis and cytokine production ontologies (Extended Data Fig. 3a). One group of genes, which we called 'ΔL1-specific genes', showed upregulation in ΔL1-expressing DN2s uniquely compared to all other DN1s and DN2s (left, Fig. 3g). The second group constituted a set of genes, which we called 'DN1 legacy genes', that were expressed in DN1s across conditions and were downregulated in ΔL7 and wild-type TCF-1-expressing DN2s but only modestly reduced in ΔL1-transduced DN2s (right, Fig. 3g). To gain insight into the identity of the two groups of genes with upregulated expression in the ΔL1-expressing cells, we plotted the normalized expression of each gene set across 62 immune cell populations[33] (Fig. 3h).

Fig. 3 | GATA2-driven mast cell gene signature is identified in developing T cells lacking L1. a, Identification and sorting strategy for DN1 and DN2 cells in *Tcf7* cKO co-cultures after in vitro differentiation on OP9-DLL1 cells for 7 d. **b**, PCA of RNA-seq on cell populations depicted in **a**. RNA-seq for each population was performed in 2–3 technical replicates for $n = 2$ independent animals. **c**, Volcano plots demonstrating significantly differential genes as calculated by DESeq2 between EV versus WT TCF-1, ΔL7 versus WT TCF-1 and ΔL1 versus WT TCF-1 transduced *Tcf7* cKO DN1s. (Adjusted $P < 0.05$ and $|\log_2\text{fold change}| > 1$) P values were calculated by the Wald test and adjusted using the Benjamini and Hochberg method. **d**, Volcano plot demonstrating significantly differential genes as calculated by DESeq2 between ΔL1 versus WT TCF-1 and ΔL7 versus WT TCF-1 transduced *Tcf7* cKO DN2s. (Adjusted $P < 0.05$ and $|\log_2\text{fold change}| > 1$). Significance was calculated as in **c**. **e**, Bar plot of expression values (in RPKM) of select genes in DN1 and DN2s. Bars represent the mean expression values ± s.d., and individual data points are overlaid. **f**, Heat map depicting genes ($n = 137$) significantly upregulated in WT versus ΔL1-transduced DN2s (Adjusted $P < 0.05$ and $|\log_2\text{fold change}| > 1$). Significance was calculated as in **c**. **g**, Heat map of two sets of genes ('ΔL1 specific' and 'DN1 legacy') significantly upregulated in ΔL1 versus WT TCF-1 transduced *Tcf7* cKO DN2s. (Adjusted $P < 0.05$ and $|\log_2\text{fold change}| > 1$) Significance was calculated as in **c**. **h**, Box plots of normalized expression of gene sets ('ΔL1 specific' and 'DN1 legacy') depicted in **g** in 62 immune cell populations from ImmGen[33]. The center line of box plots represents the median, the bounds of the box represent the 1st and 3rd quartiles, whiskers represent maximum and minimum values and data points represent outlier values. **i**, Cumulative distribution plot of corresponding fold change in GATA2 KO dendritic cell progenitors (GATA2KO/control)[36] of genes differentially upregulated and downregulated between ΔL1 and WT transduced DN2s. P value was calculated by two-sample two-sided Kolmogorov–Smirnov test on $\log_2$fold change values derived from RNA-seq on $n = 2$ independent animals, with 2–3 technical replicates each. **j**, Heat map depicting genes significantly upregulated in ΔL1 versus WT transduced DN2s that also were downregulated between GATA2 knockout and control dendritic cell progenitors. DC, dendritic cell.

https://doi.org/10.1038/s41590-023-01599-7

While the DN1 legacy gene set was specifically expressed in macrophages, monocytes and granulocytes across multiple tissues, surprisingly the ΔL1-specific gene set was distinctly upregulated in splenic dendritic cell populations and peritoneal cavity mast cells (Fig. 3h). However, canonical T cell genes such as *Bcl11b*, *Gata3*, *Il2ra* and *Lck* that were downregulated compared to wild-type TCF-1-expressing DN2s were still more highly expressed compared to all DN1 populations (Fig. 3f). The majority of the nearly 7,500 genes differentially



expressed between wild-type TCF-1-expressing DN1s and DN2s showed a similar expression pattern in ΔL1-expressing progenitors (Extended Data Fig. 3b). The expected downregulation of PU.1 (*Spi1*) from DN1 to DN2 was intact in ΔL1-expressing progenitors (Fig. 3e). Of note, the L1 domain is conserved in the human TCF-1 protein and the L1-dependent modulation of early T cell-associated genes was also recapitulated in a human cell line using RNA-seq (Extended Data Fig. 3c). We next characterized the factors that could orchestrate expression of de-repressed dendritic cell-specific and mast cell-specific genes. Notably, *Gata2* was coexpressed with other ΔL1-specific genes and had higher expression than other TFs that were differentially expressed between ΔL1 and wild-type TCF-1-expressing DN2s (Extended Data Fig. 3d). Since GATA2 is expressed in mast cells and has been reported to regulate dendritic cell differentiation[34–37], we tested whether the de-repression of genes in ΔL1-transduced DN2s corresponded with increased activation of GATA2 target genes. We reanalyzed publicly available transcriptome profiling data in GATA2-deficient dendritic cell progenitors[36] and found a significantly higher proportion of the genes upregulated in ΔL1-expressing DN2s to be downregulated in *Gata2*-null dendritic cell progenitors, suggesting that the de-repressed genes in ΔL1 DN2s are positive targets of GATA2 (Fig. 3i). Notable genes that were responsive to GATA2 in dendritic cell progenitors and found to have upregulated expression in ΔL1-expressing DN2s included: *Mcpt8*, *Maf*, *Ccl6*, *Cebp3* and *Fcer1a* (Fig. 3j and Extended Data Figs. 3e and 4). These data support the partial functionality of mutant TCF-1 lacking the L1 region and reflect a precise defect in TCF-1-dependent transcriptional repression.

L1 is dispensable for chromatin accessibility in early T cells

Previous reports characterized TCF-1 as a pioneer TF that is able to establish de novo chromatin accessibility[5,6]. We therefore hypothesized that the ability for TCF-1 to affect changes in local chromatin accessibility could be endowed by the L1 domain, and the developmental block that we observed in ΔL1-expressing progenitors may represent a downstream consequence of this failure. We profiled chromatin accessibility in DN1s and DN2s using the assay for transposase-accessible chromatin with sequencing (ATAC-seq). Notably, at the DN1 stage, TCF-1-dependent chromatin opening in ΔL1-expressing progenitors was intact (Fig. 4a,b and Supplementary Table 1). TCF-1's cognate motif appeared as the most significantly enriched motif in genomic regions demonstrating increased chromatin accessibility in both wild-type TCF-1-expressing and ΔL1-expressing DN1s compared to TCF-1-deficient EV-transduced DN1s (Extended Data Fig. 5a). Together, early chromatin opening by TCF-1 is not dependent on the L1 domain.

In DN2s, the chromatin landscape of wild-type TCF-1-expressing and ΔL1-expressing cells diverged extensively. We measured loss of chromatin accessibility in ΔL1-expressing DN2s compared with wild-type TCF-1-expressing counterparts in ~3,000 genomic regions, while an extensive gain in chromatin accessibility was measured in ~2,800 genomic regions (Fig. 4b). Motif enrichment at genomic loci that lost accessibility in ΔL1-expressing DN2s compared with wild-type DN2s showed an enrichment for RUNX1, STAT2, ETV4 and TCF motifs (Fig. 4c). We mapped chromatin accessibility levels along with binding intensity of relevant TFs including RUNX1, GATA2 and GATA3 at the lost regions. We observed that the majority of these sites were accessible in DN1s and required wild-type TCF-1 to maintain accessibility in DN2s (Fig. 4d). A smaller number of sites showed L1-dependent de novo opening in DN2s (cluster 2; Fig. 4d). In particular, these sites were correspondingly bound by GATA3 and RUNX1 in DN1s (chromatin immunoprecipitation (ChIP) columns, Fig. 4d). These data suggest a requirement for the L1 domain to maintain accessibility at regions co-bound by RUNX1.

Sites that gained accessibility in ΔL1-expressing DN2s compared with wild-type counterparts showed an enrichment for GATA, AP1 and NFAT motifs, but the TCF motif did not appear to be enriched at these sites (Fig. 4e). We mapped chromatin accessibility along with TF binding at the ~2,800 regions that gained accessibility in ΔL1-expressing DN2s compared with wild-type counterparts (Fig. 4f). These de novo accessible sites in ΔL1 DN2 were inaccessible across DN1s and in wild-type TCF-1-expressing DN2s. Correspondingly, we observed robust binding of these same loci by GATA2 in mast cells and no substantial binding of GATA3 or RUNX1 in DN1s (Fig. 4f). Altogether, the L1 domain of TCF-1 is dispensable for early changes to chromatin accessibility in DN1s. Divergence in the accessibility landscape occurs as development progressed to the DN2 stage, a stage at which the L1 domain is required to repress GATA2-induced chromatin accessibility.

L1 is required for TCF-1 binding in early T cell development

We next reasoned that the early defect in progenitors expressing ΔL1 may instead be attributed to a requirement for the L1 domain in the initial step of targeting chromatin at genomic regions. We mapped genome-wide binding profiles of wild-type and ΔL1 TCF-1 in DN1s and DN2s using CUT&RUN (Supplementary Table 1). Remarkably, we observed a 90% reduction in global binding of ΔL1 TCF-1 in DN1s compared with the wild-type counterpart. Only 4,163 binding events were detected for ΔL1 TCF-1 compared to 39,867 binding events for wild-type TCF-1 (Fig. 4g, Extended Data Fig. 5b). All sites bound by ΔL1 TCF-1 overlapped with sites bound by wild-type TCF-1 in DN1s (Extended Data Fig. 5b). In DN2s, the divergence in binding profiles narrowed, where wild-type TCF-1 bound at 65,576 sites compared to 24,082 sites bound by ΔL1 TCF-1 (Fig. 4g). The majority of binding events in DN2s were shared between wild type and ΔL1 mutant; however, 4,006 sites were uniquely bound by ΔL1 TCF-1 (Extended Data Fig. 5b). Together, ΔL1 TCF-1 has a major defect in binding DNA in DN1s.

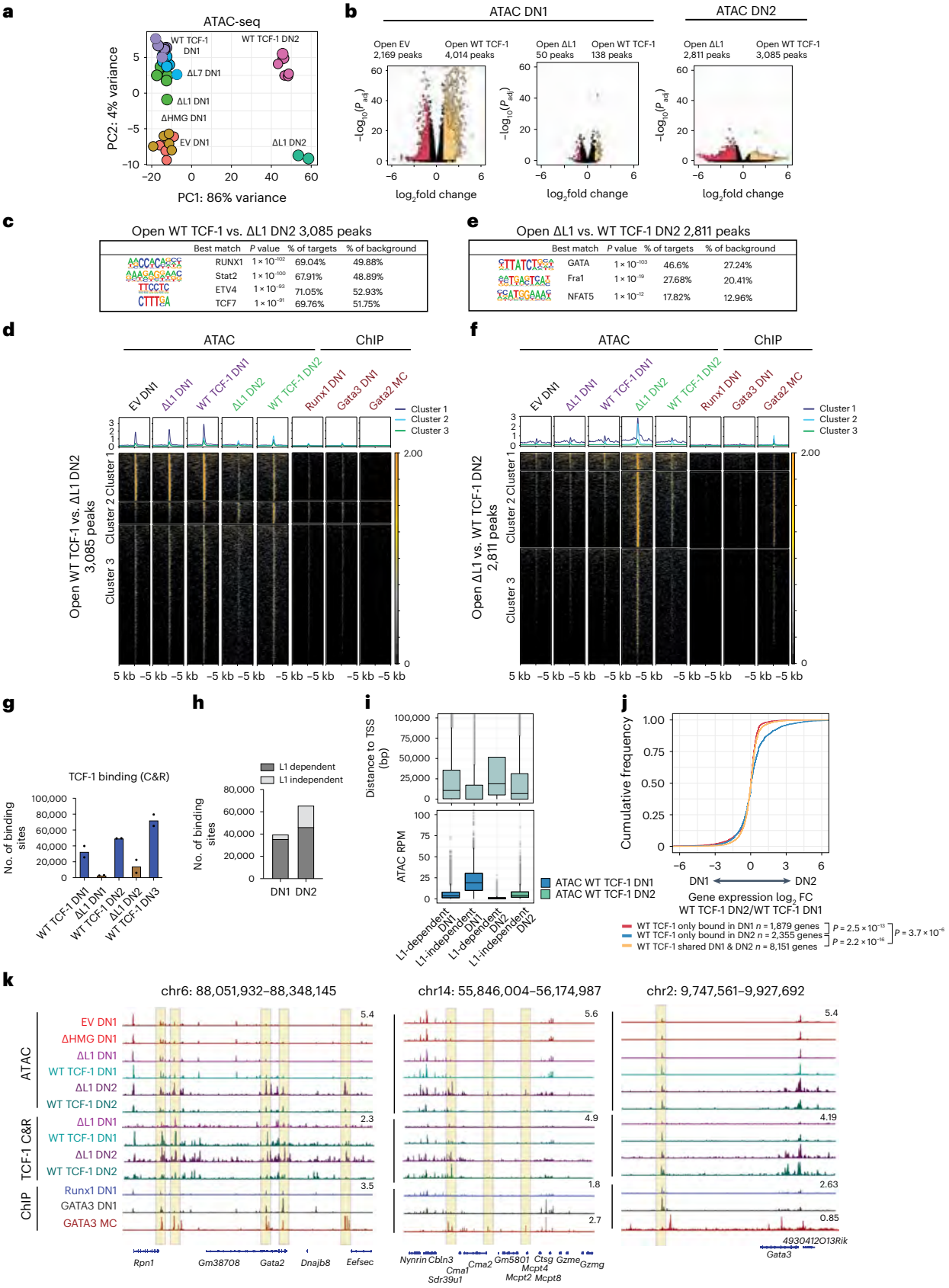
We utilized a dimensionality reduction strategy for genomic regions demonstrating TCF-1 binding and chromatin accessibility across conditions using PCA (Extended Data Fig. 5c). Wild-type TCF-1 binding did not colocalize with accessibility measurements in cells at the corresponding stage. Additionally, ΔL1 TCF-1 binding did not overlap with wild-type TCF-1 binding in DN2s and instead clustered more closely with accessibility measurements in DN2s (Extended Data

Fig. 4 | L1 modulates binding and transcriptional outcomes in early T cell development. a, PCA of ATAC-seq in WT and mutant TCF-1 DN1/DN2. ATAC-seq was performed in 1–3 technical replicates for $n = 2$ independent animals. **b**, Volcano plots demonstrating differentially accessible peaks EV versus WT, EV versus ΔL1, ΔL1 versus WT DN1 (left) and ΔL1 versus WT DN2 (right) (adjusted $P < 0.05$ and $|\log_2 \text{fold change}| > 1$). *P* values were calculated using the Wald test and adjusted using the Benjamini and Hochberg method. **c**, SeqLogo depicting enriched motifs from de novo HOMER analysis on differentially accessible peaks open in WT TCF-1 versus ΔL1-transduced DN2 with non-differential peaks as background. *P* values were calculated using a hypergeometric test. **d**, Heat map depicting chromatin accessibility in DN1 and DN2 with binding of GATA2 in mast cells and GATA3 and RUNX1 in DN1 (refs. 21,35,43) at differentially accessible peaks between WT versus ΔL1 DN2. **e**, As in **c**, motif analysis on differential peaks closed in WT TCF-1 versus ΔL1-transduced DN2. *P* values calculated as in **c**. **f**, As in **d**, depicting differentially accessible peaks closed in WT compared to ΔL1 DN2. **g**, Number of WT and ΔL1 binding sites profiled by TCF-1 CUT&RUN in DN1, DN2 and DN3 cells; $n = 2$ independent animals. Bars represent the mean number of binding sites, and individual replicate data points are overlaid. **h**, L1-dependent and L1-independent TCF-1-binding sites in DN1 and DN2 cells. **i**, Box plot representing distance to the TSS (bp) and read normalized ATAC coverage for groups of binding sites described in **h**. The center line of box plots represents the median, the bounds of the box represent the 1st and 3rd quartiles, whiskers represent the maximum and minimum values, and data points represent outliers. **j**, Cumulative distribution of genes within 1,000 bp of a WT TCF-1-binding site shared or unique to DN1/DN2 and change in expression between DN1 and DN2 ($\log_2$ fold change). *P* values were calculated by two-sample two-sided Kolmogorov–Smirnov test on $n = 2$ independent animals, with 2–3 technical replicates. WT TCF-1 only bound in DN1 versus shared DN1 and DN2, $P = 3.7 \times 10^{-6}$; WT TCF-1 only bound in DN2 versus shared DN1 and DN2, $P = 2.2 \times 10^{-16}$; WT TCF-1 only bound in DN1 versus only bound DN2, $P = 2.53 \times 10^{-13}$. **k**, Genome browser view of *Gata2*, *Mcpt* and *Gata3*.

https://doi.org/10.1038/s41590-023-01599-7

Fig. 5c). Hence, the binding of wild-type TCF-1 at distinct stages is not dictated by chromatin accessibility, consistent with previous reports of TCF-1's ability to bind to nucleosome-occupied DNA[5]. Furthermore, this intrinsic property of TCF-1 is endowed by the L1 domain.

To characterize the mechanism through which the L1 domain might affect TCF-1 binding, we delineated binding events in both DN1s and DN2s for which binding was dependent or independent of the L1 domain (Fig. 4h). We observed in each of these stages binding events



https://doi.org/10.1038/s41590-023-01599-7

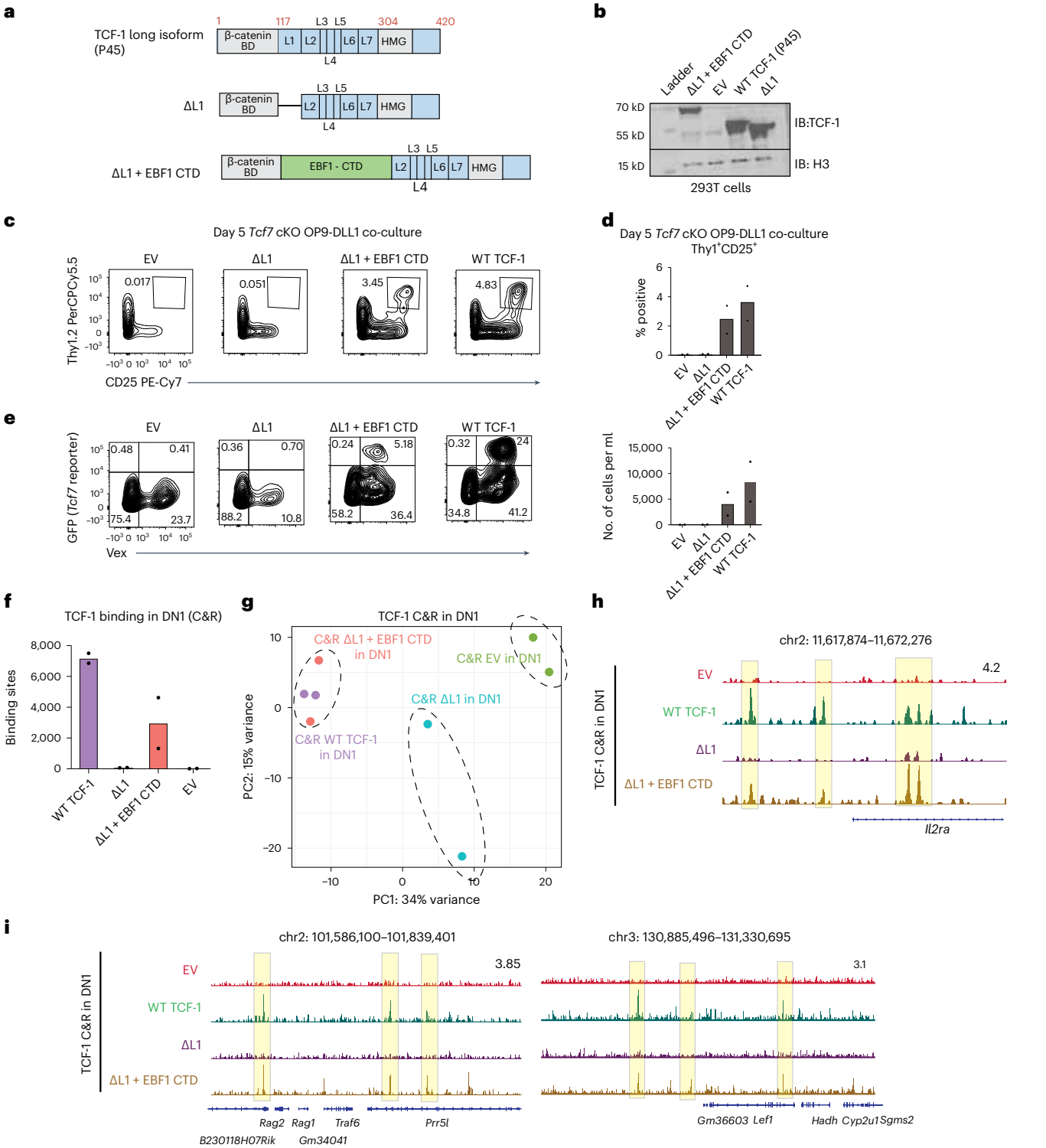


Fig. 5 | L1 can be functionally substituted with another unstructured domain. **a**, Schematic of WT isoform of TCF-1 (P45), mutant lacking the L1 domain (ΔL1) and mutant in which the L1 domain is replaced with the C terminus of EBF1 (ΔL1 + EBF1 CTD). **b**, Immunoblot analysis of 293T cells transfected with WT TCF-1 and mutant TCF-1 constructs; ΔL1 and ΔL1 + EBF1 CTD. Immunoblot was probed with TCF-1 antibody and H3 as a loading control. **c**, Identification of Thy1$^+$CD25$^+$ cells in OP9-DLL1 co-cultures of *Tcf7* cKO cells transduced with EV, WT TCF-1 or mutant TCF-1 (ΔL1 and ΔL1 + EBF1 CTD) on day 5 after in vitro differentiation. All cells were pre-gated on SSC-A/FSC-A, singlets, live cell (viability$^-$), CD45$^+$, transduced (vex$^+$). **d**, Quantification of frequency (left) and numbers (right) of Thy1$^+$CD25$^+$ cells from *Tcf7* cKO cells on day 5 after in vitro differentiation on OP9-DLL1 cells (**c**). Bars represent mean values from $n = 2$ independent animals. Individual replicate data points are shown. **e**, Representative flow cytometric analysis identifying transduced (vex$^+$) GFP$^+$ cells (*Tcf7* eGFP reporter). All cells were pre-gated on SSC-A/FSC-A, singlets, live cell (viability$^-$), CD45$^+$. **f**, Quantification of binding sites identified by TCF-1 CUT&RUN in DN1s transduced with WT TCF-1, ΔL1, ΔL1 + EBF1 CTD and EV. CUT&RUN experiments for each population were performed in at least two biological replicates. Bars represent the mean number of binding sites from $n = 2$ independent animals. Individual data points are shown. **g**, PCA of TCF-1 CUT&RUN in *Tcf7* cKO DN1s and DN2s transduced with WT TCF-1, ΔL1, ΔL1 + EBF1 CTD and EV. CUT&RUN experiments for each population were performed in at least two biological replicates. **h**,**i**, Genome browser view of *Il2ra*, *Rag1/Rag2* and *Lef1* loci visualizing CUT&RUN profiles of DN1s and DN2s from OP9-DLL1 co-culture of *Tcf7* cKO cells at day 7.

https://doi.org/10.1038/s41590-023-01599-7

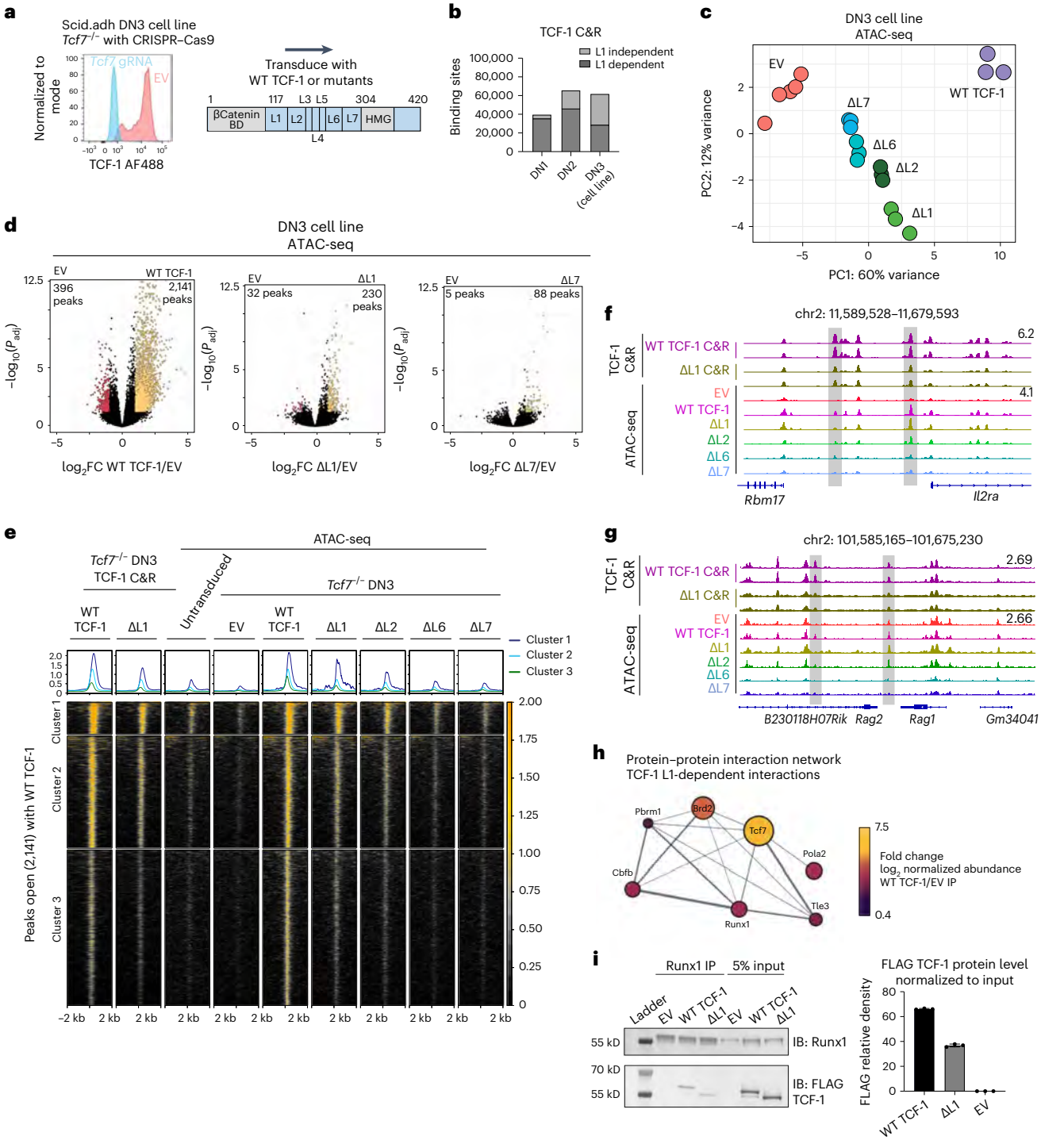


Fig. 6 | Loss of L1 has limited effect on chromatin accessibility in committed T cells. a, Experimental design of gene replacement strategy using retroviral transduction of WT (P45 isoform) or mutant TCF-1 in Scid.adh (DN3) cells after CRISPR–Cas9 disruption of endogenous TCF-1. **b**, Quantification of L1-dependent and L1-independent binding sites detected with TCF-1 CUT&RUN in DN1, DN2 and Scid.adh DN3 cells. Bars represent the total number of binding sites. **c**, PCA depicting ATAC-seq in *Tcf7*$^{-/-}$ DN3 cells transduced with WT, EV or mutant TCF-1 (ΔL1, ΔL2, ΔL6 or ΔL7). **d**, Volcano plots demonstrating differential accessibility of ATAC-seq peaks between EV-transduced and WT TCF-1, ΔL1 and ΔL7 transduced *Tcf7*$^{-/-}$ DN3 cells. (adjusted $P < 0.05$ and $|\log_2\text{fold change}| > 1$). *P* values were calculated by the Wald test and adjusted using the Benjamini and Hochberg method. **e**, Heat map demonstrating TCF-1 and ΔL1 binding measured by CUT&RUN and chromatin accessibility in WT and mutant TCF-1 (ΔL1, ΔL2, ΔL6 and ΔL7) transduced cells at peaks significantly open in WT TCF-1 compared to EV-transduced *Tcf7*$^{-/-}$ DN3 cells. (adjusted $P < 0.05$ and $|\log_2\text{fold change}| > 1$). *P* values were calculated by the Wald test and adjusted using the Benjamini and Hochberg method. **f**,**g**, Genome browser views of *Il2ra* and *Rag1/Rag2* loci depicting TCF-1 CUT&RUN and ATAC-seq profiles in WT TCF-1, EV and mutant TCF-1 (ΔL1, ΔL2, ΔL6 and ΔL7) transduced *Tcf7*$^{-/-}$ DN3. **h**, Depiction of L1-dependent TCF-1 protein–protein interaction network identified by mass spectrometry (MS) of TCF-1 immunoprecipitation in DN3 cells. All interactions were filtered and ranked (Methods) to identify proteins with enrichment in WT TCF-1 versus EV and ΔL1 immunoprecipitations. Network was filtered based on first neighbor nodes of *Tcf7*. Node size and color indicate fold change in normalized abundance between DN3 cells expressing WT TCF-1 and EV. **i**, RUNX1 co-immunoprecipitation with separate parallel immunoblotting for RUNX1 and TCF-1. Bar plot depicts mean FLAG protein level quantification normalized to 5% input. Data points indicate three quantifications per condition. Error bars represent the s.d. Results are representative of $n = 2$ biologically independent samples.

that depended on the presence of the L1 domain were more distant from promoters and showed lower chromatin accessibility than sites that were bound independently of the L1 domain (Fig. 4i). We performed de novo motif analysis and observed an enrichment of ETS and RUNX motifs, but not the TCF-1's cognate motif, at L1-dependent sites bound by TCF-1 in DN1s compared to DN2s (Extended Data Fig. 5d). Despite a requirement for the L1 domain in binding of TCF-1 to early DN1 targets, a corresponding L1 dependency in creating de novo chromatin accessibility in DN1s or DN2s was not detected (Extended Data Fig. 5e). Together, the L1 domain is required for binding of TCF-1 at distal regions with a low level of chromatin accessibility and low enrichment for TCF-1's cognate motif at early stages of T cell development.

L1 is required for stage-dependent transcriptional outcomes

To determine the consequences of L1-dependent binding at distinct stages, we linked TCF-1 binding with the transcriptional regulation of its target genes. We selected TCF-1 target genes based on detection of wild-type TCF-1 binding events within 1,000 bp of genes' transcriptional start sites (TSSs) and evaluated gene expression differences in DN1s and DN2s in three classes defined by shared and unique binding of TCF-1 in DN1s and DN2s (Extended Data Fig. 5b). Genes bound by TCF-1 in both DN1s and DN2s were moderately expressed in DN1s and showed no increase in expression in DN2 (Fig. 4j). In contrast, genes bound by TCF-1 specifically in DN2s were biased to DN2-specific gene expression (Fig. 4j). Notably, genes bound by TCF-1 in DN1s did not coincide with increased expression subsequently in DN2s (Fig. 4j), suggesting an early role of transient TCF-1 occupancy in preemptive gene repression. With these data, we reasoned that the effects of reduced binding by ΔL1 in DN1s preferentially impacted suppression of alternative lineage genes including *Gata2* and other mast cell genes (Fig. 4k) at which we observed a corresponding decrease in DN1 ΔL1 occupancy and an increase in subsequent lineage inappropriate chromatin accessibility. The binding disparity in DN2s may underlie inefficient T cell gene activation including at *Gata3* (Fig. 4k). Together, these data provided a model through which TCF-1 orchestrates transcriptional control to allow T cell developmental competence. This model postulates a transient early wave of L1-dependent TCF-1 binding, as we observed at the *Gata2* locus, to regions with low TCF motif enrichment, low chromatin accessibility and frequently enriched for RUNX1 binding at target genes whose expression is inhibited in T cells. A second wave of binding of TCF-1 in DN2s occurs at regions enriched for TCF-1 motifs and promotes T cell-specific gene activation, as illustrated at the *Gata3* locus.

L1 can be functionally substituted with another unstructured domain

The interchangeable nature of the IDRs of proteins has been described previously[38]. To determine if the L1 domain of TCF-1 could be functionally replaced with another previously characterized IDR, we designed a construct in which the L1 region was replaced with the C-terminal domain (CTD) of EBF1 (refs. 39,40; Fig. 5a,b). We referred to this construct as ΔL1 + EBF1 CTD. Surprisingly, we observed a significant rescue in both the absolute number and percentage of $Thy1^+CD25^+$ cells when progenitors were transduced with ΔL1 + EBF1 CTD, unlike the progenitors transduced with ΔL1 (Fig. 5c,d). The expression of ΔL1 + EBF1 CTD also rescued the defect in the induction of the GFP reporter in transduced cells (Fig. 5e). We next evaluated if this unrelated IDR could also rescue the defect in the ability of ΔL1 to target chromatin in DN1s. We mapped the global binding events of ΔL1 + EBF1 CTD TCF-1 in DN1s using CUT&RUN (Fig. 5f). Remarkably, ΔL1 + EBF1 CTD TCF-1 showed binding to a substantially increased number of genomic sites compared to ΔL1 TCF-1 including *Rag1*/*Rag2*, *Il2ra* and *Lef1* loci, although it did not completely recapitulate the binding profile of wild-type TCF-1 (Fig. 5f–i). Together, the T cell developmental defect associated with the loss of L1 domain was linked to the ability of TCF-1 to access its full range of binding sites in DN1s and this defect could be rescued with another TF's IDR.

Limited effect of L1 on chromatin accessibility in committed T cells

We next sought to analyze how deletion of L1 or other domains in the N terminus of TCF-1 affects chromatin accessibility and gene regulation at a post-commitment stage of T cell development in which cells cannot adopt alternative fates to T cells. Hence, we utilized a gene-replacement strategy in a T cell post-commitment *Tcf7*$^{-/-}$ pro-T cell line, DN3-like Scid.adh cells, abbreviated as DN3 (refs. 7,21; Fig. 6a). We first ablated endogenous TCF-1 with CRISPR–Cas9 in DN3s[7] and then 'replaced' expression with wild-type or mutant TCF-1. We measured transcriptional outputs and found ΔL1-expressing DN3s clustered more closely with cells expressing wild-type TCF-1, while ΔL7-expressing cells were closer to TCF-1-deficient EV-transduced cells (Extended Data Fig. 6a). Surprisingly, deletion of the L7 region of TCF-1 was more detrimental to TCF-1-dependent gene regulation at DN3s than deletion of the L1 region (Extended Data Fig. 6b–d).

In DN3s, wild-type TCF-1 bound 62,046 sites, while ΔL1 TCF-1 bound to 36,448 sites as measured by CUT&RUN (Extended Data Fig. 6e). The comparison of TCF-1 binding data across stages suggested that the progression of cells between DN1 and DN3 coincided with a reduction in the percentage of sites that depended on the L1 domain for binding (Fig. 6b and Extended Data Fig. 6e,f). We next analyzed chromatin accessibility measured by ATAC-seq in mutant TCF-1-expressing DN3s. PCA of the chromatin accessibility in mutant TCF-1-expressing DN3s displayed a distinct epigenetic state compared to either EV or wild-type TCF-1-expressing controls. Moreover, the ΔL1-replaced cells exhibited a closer relationship to wild-type TCF-1-transduced cells, while the ΔL7-replaced cells were closer to TCF-1-deficient EV-transduced cells (Fig. 6c). Wild-type TCF-1 transduction led to a significant gain in chromatin accessibility, while the ΔL1 and ΔL7 TCF-1 established only 230 and 88 accessible regions, respectively, at which accessibility was gained compared to EV, with the greatest defect observed in ΔL7-replaced cells (Fig. 6d). We performed *k*-means clustering on chromatin accessibility and visualized data using heat maps across mutant and wild-type TCF-1-replaced DN3s at the 2,141 genomic sites significantly more open with wild-type TCF-1 compared to EV-transduced DN3s (Fig. 6e). Both ΔL6 and ΔL7 showed a greater reduction in creating open chromatin regions compared to the ΔL1 and ΔL2 relative to wild-type TCF-1 as illustrated at the *Il2ra* and *Rag1*/*Rag2* loci (Fig. 6e–g). These findings suggest the importance of the L7 region for the functionality of TCF-1 after T cell commitment.

L1-dependent interaction between RUNX1 and TCF-1

Recent studies on protein interactions mediated by TF IDRs highlighted the formation of biomolecular condensates or foci representing high local concentrations of TFs and transcriptional machinery[15,40]. We generated constructs in which wild-type TCF-1 and ΔL1 were fused with GFP, transduced DN3s with GFP fusion constructs and visualized cells with confocal microscopy. The GFP signal in both wild-type and ΔL1 TCF-1 localized to the nucleus with distinct granular morphology compared to an EV control in which GFP alone is expressed homogenously in both the cytoplasm and nucleus (Extended Data Fig. 7a). This morphology was not consistent with discrete foci; however, we found the GFP signal associated with both wild-type and ΔL1 TCF-1 to be more granular within the nucleus than GFP alone (Extended Data Fig. 7a). Hence, a local partitioning of TCF-1 in the nucleus does not depend on the L1 domain.

To identify proteins that could interact with TCF-1 in an L1-dependent manner in DN3s, we performed affinity purification of FLAG-tagged wild-type and ΔL1 TCF-1 followed by liquid chromatography with tandem mass spectrometry (LC–MS/MS). Identified interactors were scored by enrichment in the immunoprecipitation of wild-type TCF-1 compared to ΔL1 and EV control (Fig. 6h). We constructed a network of top L1-dependent putative protein–protein interactions based on the extent of enrichment between wild-type TCF-1 and EV immunoprecipitations (Fig. 6h, Extended Data Fig. 7b,c

and Supplementary Table 1). UniProt keywords 'acetylation', 'phosphoprotein' and 'nucleus' were significantly enriched in the network (Extended Data Fig. 7d). Notably, we identified RUNX1 (with cofactor CBFB[41]) and Tle3 as L1-dependent interactors (Fig. 6h). The TCF-1 and Tle3 interactions have been described previously to partition Tle3 between TCF-1 and RUNX1/RUNX3 in $CD8^+$ T cell lineage specification[42]. We validated the L1-mediated association of TCF-1 with RUNX1 by co-immunoprecipitation (Fig. 6i). Together, the L1-dependent interaction of RUNX1 and chromatin-associated proteins with TCF-1 enables TCF-1-dependent gene regulation. Furthermore, the interaction between TCF-1 and RUNX1 mediated by the L1 domain can occur at early stages of T cell development as well as in post-commitment DN3s and likely has functional significance.

Discussion

Pioneering work over 30 years ago identified TCF-1 as an exquisitely tissue-specific factor that binds DNA in the minor groove to distort and bend the double helix[10]. In the intervening years, the molecular mechanisms of TCF-1's function in development and disease have come to light. Despite these advances, key questions remain about the role of non-DNA-binding domains of TCF-1. In this study, we showed that distinct regions within the N terminus of TCF-1 have integral roles in orchestrating T cell development. We uncovered L1, an IDR within the N terminus of TCF-1 that was required for efficient early T cell development. BM progenitors that lacked L1 were unable to upregulate T cell identity genes and showed a marked de-repression of GATA2 target genes normally restricted to mast cell and dendritic cell lineages. The L1 region of TCF-1 facilitated early binding to inaccessible loci lacking the TCF motifs, which corresponded to genes repressed later in T cell development. This impact on early binding was linked to the inability of ΔL1-expressing cells to progress developmentally. Rescue of early binding and development was achieved by substituting the L1 domain for a heterologous disordered domain. We additionally identified L7, a region flanking the DNA-binding domain of TCF-1 that contributed to TCF-1-dependent chromatin opening and gene regulation in a T cell committed DN3 cell line, but whose loss did not contribute to a developmental block in primary early T cells.

TCF-1 is one of the earliest mediator of T cell-specific gene control and as such is positioned to reshape cell fate. Pioneer TFs can interface with repressed chromatin and shape cell identity, while other TFs are limited to sites within already accessible chromatin[2]. Pioneer factors can engage with compacted chromatin but may still require recruitment of other factors to affect sustained changes[2]. The L1 domain was required not only for binding of TCF-1 in DN1s, but also for an association with RUNX1 and its obligate cofactor CBFB. Whether this interaction is direct or whether the L1 domain enables TCF-1 to bind at RUNX1 co-occupied regions remains unclear. Recent reports have described dynamic genomic occupancy and transcriptional control by RUNX1/RUNX3 during T cell development, enabling distinct associations with cofactors[24]. In one example, early expression of PU.1 can lead to a redistribution of RUNX1 binding[43]. In such cases, TCF-1-mediated repression of PU.1 may facilitate the L1-dependent co-binding of RUNX1 and TCF-1, thereby promoting T cell development.

A reductionist view of TFs separates DNA-binding and effector functions into modular distinct domains. However, a large body of work demonstrates that non-DNA-binding domains often enable TFs to bind compacted chromatin and initiate chromatin opening[25,40,44,45]. The function of non-DNA-binding domains intersects the sequential process through which TFs function and interact with chromatin. Here, the deletion of the L1 domain had a distinct impact on TCF-1 binding and chromatin opening. The lack of binding stability, separate from chromatin opening, suggests a regulatory mechanism where continuous occupancy is not essential. Instead, a transient 'hit and run' binding event may initiate accessibility in this early context, allowing other partner factors to bind and sustain accessibility. In later stages of T cell development, TCF-1 binding was less dependent on the L1 domain. This highlights the specific requirement for non-DNA-binding domains early in developmental trajectories before cell specification when the chromatin landscape has not been extensively acted upon by other factors.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-023-01599-7.

References

1. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
2. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
3. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
4. Verbeek, S. et al. An HMG-box-containing T cell factor required for thymocyte differentiation. *Nature* **374**, 70–74 (1995).
5. Johnson, J. L. et al. Lineage-determining transcription factor TCF-1 initiates the epigenetic identity of T cells. *Immunity* **48**, 243–257 (2018).
6. Emmanuel, A. O. et al. TCF-1 and HEB cooperate to establish the epigenetic and transcription profiles of $CD4^+CD8^+$ thymocytes. *Nat. Immunol.* **19**, 1366–1378 (2018).
7. Wang, W. et al. TCF-1 promotes chromatin interactions across topologically associating domains in T cell progenitors. *Nat. Immunol.* **23**, 1052–1062 (2022).
8. Xu, Z. et al. Cutting edge: beta-catenin-interacting Tcf1 isoforms are essential for thymocyte survival but dispensable for thymic maturation transitions. *J. Immunol.* **198**, 3404–3409 (2017).
9. Zhao, X., Shan, Q. & Xue, H. H. TCF1 in T cell immunity: a broadened frontier. *Nat. Rev. Immunol.* **22**, 147–157 (2022).
10. Love, J. J. et al. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795 (1995).
11. Xing, S. et al. Tcf1 and Lef1 transcription factors establish $CD8^+$ T cell identity through intrinsic HDAC activity. *Nat. Immunol.* **17**, 695–703 (2016).
12. Wang, C., Uversky, V. N. & Kurgan, L. Disordered nucleiome: abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. *Proteomics* **16**, 1486–1498 (2016).
13. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* https://doi.org/10.1126/science.aaf4382 (2017).
14. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
15. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 (2018).
16. Ben Chorin, A. et al. ConSurf-DB: an accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci.* **29**, 258–267 (2020).
17. Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. & Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208 (2006).
18. Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E. & Englander, S. W. Protein hydrogen exchange: testing current models. *Protein Sci.* **21**, 987–995 (2012).

19. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
20. Connelly, G. P., Bai, Y., Jeng, M. F. & Englander, S. W. Isotope effects in peptide group hydrogen exchange. *Proteins* **17**, 87–92 (1993).
21. Hosokawa, H. et al. Stage-specific action of Runx1 and GATA3 controls silencing of PU.1 expression in mouse pro-T cells. *J. Exp. Med.* https://doi.org/10.1084/jem.20202648 (2021).
22. Hu, G. et al. Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. *Immunity* **48**, 227–242 (2018).
23. Rothenberg, E. V., Hosokawa, H. & Ungerback, J. Mechanisms of action of hematopoietic transcription factor PU.1 in initiation of T cell development. *Front Immunol.* **10**, 228 (2019).
24. Shin, B. et al. Runx1 and Runx3 drive progenitor to T-lineage transcriptome conversion in mouse T cell commitment via dynamic genomic site switching. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.2019655118 (2021).
25. Ungerback, J. et al. Pioneering, chromatin remodeling, and epigenetic constraint in early T cell gene regulation by SPI1 (PU.1). *Genome Res.* **28**, 1508–1519 (2018).
26. Zhou, W., Gao, F., Romero-Wolf, M., Jo, S. & Rothenberg, E. V. Single-cell deletion analyses show control of pro-T cell developmental speed and pathways by Tcf7, Spi1, Gata3, Bcl11a, Erg, and Bcl11b. *Sci. Immunol.* **7**, eabm1920 (2022).
27. Zhou, W. et al. Single-cell analysis reveals regulatory gene expression dynamics leading to lineage commitment in early T cell development. *Cell Syst.* **9**, 321–337 (2019).
28. Schmitt, T. M. & Zuniga-Pflucker, J. C. T cell development, doing it in a dish. *Immunol. Rev.* **209**, 95–102 (2006).
29. Schmitt, T. M. & Zuniga-Pflucker, J. C. Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity* **17**, 749–756 (2002).
30. de Boer, J. et al. Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33**, 314–325 (2003).
31. Yang, Q. et al. TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nat. Immunol.* **16**, 1044–1050 (2015).
32. Weber, B. N. et al. A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* **476**, 63–68 (2011).
33. Yoshida, H. et al. The *cis*-regulatory atlas of the mouse immune system. *Cell* **176**, 897–912 (2019).
34. Ling, K. W. et al. GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J. Exp. Med.* **200**, 871–882 (2004).
35. Li, Y. et al. GATA2 regulates mast cell identity and responsiveness to antigenic stimulation by promoting chromatin remodeling at super-enhancers. *Nat. Commun.* **12**, 494 (2021).
36. Onodera, K. et al. GATA2 regulates dendritic cell differentiation. *Blood* **128**, 508–518 (2016).
37. Tsai, F. Y. & Orkin, S. H. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* **89**, 3636–3643 (1997).
38. Jin, W. et al. Critical POU domain residues confer Oct4 uniqueness in somatic cell reprogramming. *Sci. Rep.* **6**, 20818 (2016).
39. Boller, S. et al. Pioneering activity of the C-terminal domain of EBF1 shapes the chromatin landscape for B cell programming. *Immunity* **44**, 527–541 (2016).
40. Wang, Y. et al. A prion-like domain in transcription factor EBF1 promotes phase separation and enables B cell programming of progenitor chromatin. *Immunity* **53**, 1151–1167 (2020).
41. de Bruijn, M. F. & Speck, N. A. Core-binding factors in hematopoiesis and immune function. *Oncogene* **23**, 4238–4248 (2004).
42. Xing, S. et al. Tle corepressors are differentially partitioned to instruct $CD8^+$ T cell lineage choice and identity. *J. Exp. Med.* **215**, 2211–2226 (2018).
43. Hosokawa, H. et al. Transcription factor PU.1 represses and activates gene expression in early T cells by redirecting partner transcription factor binding. *Immunity* **48**, 1119–1134 (2018).
44. Frederick, M. A. et al. A pioneer factor locally opens compacted chromatin to enable targeted ATP-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* **30**, 31–37 (2023).
45. Minderjahn, J. et al. Mechanisms governing the pioneering and redistribution capabilities of the non-classical pioneer PU.1. *Nat. Commun.* **11**, 402 (2020).
46. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **1804**, 996–1010 (2010).
47. Nguyen, D., Mayne, L., Phillips, M. C. & Walter Englander, S. Reference parameters for protein hydrogen exchange rates. *J. Am. Soc. Mass. Spectrom.* **29**, 1936–1939 (2018).

Methods

Cell culture

Scid.adh cell line, a pro-T cell line derived from spontaneous thymic lymphomas[48], was a kind gift from W. Pear's laboratory at the University of Pennsylvania. These cells were grown in RPMI 1640 medium (Invitrogen), supplemented with 10% FBS (Fisher Scientific), 1 mM sodium pyruvate (Gibco), 1% non-essential amino acids (Gibco), 2 mM L-glutamine (Lonza), 1% penicillin–streptomycin and 0.1% 2-mercaptoethanol (Gibco). OP9-ctrl, OP9-DLL1 and OP9–DLL4 cells were a kind gift from the laboratory of I.M. at the University of Pennsylvania. These cells were maintained in αMEM (Invitrogen), supplemented with 20% FBS and 1% penicillin–streptomycin. HEK 293T cells were purchased from the American Type Culture Collection (ATCC; CRL-3216; RRID: CVCL_0063). HEK 293T cells were maintained in high-glucose DMEM medium 1× with L-glutamine (Invitrogen), supplemented with 100 U ml^{-1} penicillin and 100 mg ml^{-1} streptomycin (Gibco) with 10% FBS. NIH 3T3 cells were purchased from the ATCC (CRL-1658 RRID: CVCL_0594). NIH 3T3 cells were maintained in high-glucose DMEM medium 1× with L-glutamine (Invitrogen), supplemented with 100 U ml^{-1} penicillin and 100 mg ml^{-1} streptomycin (Gibco) with 10% bovine serum, heat inactivated (Thermo). Cells were maintained at a low passage number (<12), at 70–80% confluency. All cells were grown at 37 °C and 5% CO_2. Cell lines were not authenticated. Mycoplasma contamination was tested periodically in all cell lines, and no mycoplasma contamination was detected. Commonly misidentified cell lines were not used.

Mice

All wild-type mice used in our study were on a C57BL/6J background. Female and male breeder Vav-iCre transgenic mice (008610)[30,49,50] and *Tcf7*^{eGFP} mice (030909)[31] were purchased from Jackson Laboratory. '*Tcf7*$^{-/-}$ cKO' mice were generated by breeding *Tcf7*^{eGFP} mice, in which two *loxP* sites are inserted on either side of exon 2 of the *Tcf7* gene, with Vav-iCre mice. The F1 generation was backcrossed to *Tcf7*^{eGFP} mice to reach homozygous floxed Cre+ experimental mice (*Tcf7* cKO). All mice were bred and housed in an American Association for the Accreditation of Laboratory Animal Care accredited vivarium at the University of Pennsylvania. All husbandry and experimental procedures were performed according to the protocol reviewed and approved by the Institutional Animal Care and Use Committee. Mice were fed with 5010 - Laboratory Autoclavable Rodent Diet (LabDiet), and were maintained at a 12-h light/12-h dark cycle, between 18–23 °C and 40–60% humidity. Experimental and control mice were 6–10 weeks old of either sex. At least two biological replicate mice of matching age and sex were used for each experiment.

Cell preparation

Single-cell suspensions were prepared from the BM removed from the femurs and tibiae of 6- to 8-week-old C57BL/6J or *Tcf7* cKO mice. Ckit$^+$ BM cells were enriched for with EasySep Mouse CD117 (cKIT) Positive Selection kit (StemCell, 187757) according to the manufacturer's instructions. Enriched cells were co-cultured on OP9 monolayers or stained for LSK sorting. For LSK sorting, cells were stained with LD Aqua (Invitrogen, L34957), a combination of lineage antibodies (Ter119 (BioLegend, 116211), CD3 (BioLegend, 100311), NK1.1 (BioLegend, 108709), GR1 (BioLegend, 108411), TCRgd (BioLegend, 108411), TCRb (BioLegend, 109211), Cd11c (BioLegend, 117309), Cd19 (BioLegend, 152410) B220 (BioLegend, 103211), CD11b (BioLegend, 101211); all diluted at 1:200), Sca1 (BioLegend, 122513, dilution of 1:200) and Ckit (BioLegend, 105807, dilution of 1:300) and were sorted for viability (Thermo Scientific, L34966), Lin$^-$, Ckit$^+$, Sca1$^+$. Ckit$^+$ or sorted LSK cells were activated in IMDM medium supplemented with 20% FBS, 1% penicillin–streptomycin, SCF (100 ng ml^{-1}), interleukin (IL)-6 (5 ng ml^{-1}) and IL-3 (10 ng ml^{-1}) overnight. Transduced cells were plated the following day on OP9 monolayers in OP9 medium supplemented with 5 ng ml^{-1} Flt-3L and 1 ng ml^{-1} IL-7 for 5, 7 or 13 d. Co-cultures were passaged by gently disrupting cells, passaged through a 40-µm cell strainer (Falcon) and transferred onto new OP9 monolayers every 4–5 d. Cells from co-cultures were stained with L/D APCef780 (Invitrogen, 65-0865-14, dilution of 1:4,000) and fluorescence antibodies to B220 (BioLegend, 103211, dilution of 1:300), CD44 (BioLegend, 103041, dilution of 1:400), CD45 (BioLegend, 103151, dilution of 1:400), Thy1.2 (BioLegend, 105338, dilution of 1:300), Ckit (BioLegend, 105807, dilution of 1:300), CD25 (BioLegend, 105338, dilution of 1:350) and CD11b (BioLegend, 101211, dilution of 1:200). Sorting was performed on a BD FACSAria after 7 d to isolate DN1 (CD45$^+$c-Kit^{hi}CD44^{hi}CD25^{lo}), DN2 (CD45$^+$c-Kit^{lo}CD44^{lo}CD25^{hi}) and DN3 (CD45$^+$ Ckit^{lo}CD44^{lo}CD25^{lo}) cells.

Cloning/generation of TCF-1 mutants

FLAG-tagged MSCV GFP-TCF-1 constructs for the long (P45) and short (P33) isoforms of TCF-1 as well as mutants ΔL1–ΔL5 were a kind gift from H.-H.X. To create ΔL6 and ΔL7 mutants, deletion flanking primers were used with Q5 site-directed mutagenesis kit (NEB, E0554S) according to the manufacturer's instructions. TCF-1 P45 Vex MSCV constructs[5] were utilized with deletion flanking primers and Q5 site-directed mutagenesis kit to create all mutants on a Vex MCSV backbone. Mutant TCF-1 ΔL1 + EBF1 CTD was constructed with HIFI NEBuilder HiFi Assembly Kit (NEB, E5520S) and PCR-based cloning with primers designed to amplify a 489-bp region encoding EBF1's CTD region with overlaps flanking the L1 domain of TCF-1 on the Vex MSCV backbone. Constructs in which wild-type TCF-1 and ΔL1 were fused to GFP along with an EV GFP construct were created with HIFI NEBuilder HiFi Assembly Kit and PCR-based cloning into a custom pMSCV-derived plasmid containing an EGFP variant (with monomerizing p.Ala206Lys mutation) ('mEGFP') downstream of the mouse PGK1 promoter. Human ΔL1 TCF-1 was created using the Q5 site-directed mutagenesis kit according to the manufacturer's instructions, both human wild-type TCF-1 and corresponding mutant ΔL1 were cloned into lentiviral LRG2.1 downstream of the U6 promoter using HIFI NEBuilder HiFi Assembly Kit and PCR-based cloning. All constructs were confirmed by Sanger sequencing.

Transduction for *Tcf7*$^{-/-}$ cells

CRISPR–Cas9 was used to delete TCF-1 in *Scid.adh* cells as described previously[7]. Transduction of *Tcf7* KO scid.adh cells was accomplished by addition of retroviral supernatants to culture medium supplemented with polybrene (8 mg ml^{-1}) and spinfected at 700*g* for 25 min. At 72 h after transduction, live transduced cells were sorted for downstream experiments. Retroviral transduction of Ckit$^+$ BM and LSK cells was performed by spinfection of cells with equal volumes of viral supernatants for 90 min at 1,300*g* at room temperature (RT), after 4-h virions were diluted with fresh IMDM medium and cells were returned to the incubator overnight, cells were plated on OP9 monolayers the following day.

Retroviral packaging

For retroviral packaging of mutant TCF-1 plasmids (GFP MSCV or Vex MSCV backbone), 4×10^6 293T cells were plated in 4 ml DMEM medium in 10-cm dishes on the day before transfection. Immediately before transfection, chloroquine was added to the medium to a final concentration of 25 mM. The retroviral construct/empty vector and the pCL-Eco plasmid were transiently co-transfected using Lipofectamine 3000 (Invitrogen). The cells were returned to the incubator for 6 h. Subsequently, the medium was replaced with fresh medium. Virions were collected 24 and 48 h after transfection, snap frozen, and stored at −80 °C for future use.

Immunoblot

Immunoblotting was performed on whole-cell lysates from transduced 3T3 and DN3 cells, and transfected 293T cells. Cells were lysed with 1× RIPA buffer supplemented with proteinase inhibitor cocktail. Equal numbers of cells for each condition were utilized and equal volumes of lysate were loaded on a NuPAGE 4–12% Bis-Tris gel and transferred

using the iBlot 2 Gel Transfer Device. Membranes were blocked with 5% non-fat dry milk in 1× TBST buffer followed by incubation with primary anti-mouse M2 FLAG antibody (MilliporeSigma, F1804; dilution of 1:1,000), mouse anti-RUNX1 (Santa Cruz, sc-365644; dilution of 1:200) and rabbit anti-mouse vinculin (Santa Cruz, sc-25336; dilution of 1:200) and finally probed with horseradish peroxidase-conjugated anti-rabbit IgG (CST, 7074, dilution of 1:2,000) or anti-mouse IgG (CST, 7076, dilution of 1:2,000) secondary antibodies. Blots were visualized with SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) on the ChemiDoc Imaging system (Bio-Rad).

Co-immunoprecipitation assays for immunoblot

Co-immunoprecipitation assays were performed as described[51]. Antibodies were conjugated to protein G beads including FLAG antibody (6 µg, Sigma, F1804), anti-TCF7 (6 µg, Cell Signaling Technology, C63D9) or anti-RUNX1 antibody (6 µg, Abcam, Ab23980 RUNX1). Beads were washed in blocking buffer three times and clarified lysate was incubated with the antibody-conjugated beads rotating overnight at 4 °C. The mixture was washed with immunoprecipitation buffer without supplements three times and eluted by boiling in NuPAGE loading dye (Invitrogen) at 95 °C for 5 min. Samples were analyzed by immunoblotting. Immunoblots were quantified using Fiji[52] (ImageJ2 version 2.9.0) to assess protein density. FLAG immunoprecipitation protein density was normalized to input protein density. Quantification of band density was performed three times for each condition. Error bars represent the standard deviation of these three quantifications.

Co-immunoprecipitation assays for mass spectrometry

For samples analyzed by mass spectrometry (MS), the following modifications were made to the co-immunoprecipitation protocol. Cells for 5% input were lysed separately with a non-detergent lysis buffer (6 M urea, 2 M thiourea in 50 mM ammonium bicarbonate (pH 8)). After the overnight incubation, the beads were washed once with immunoprecipitation buffer, and then twice with non-detergent immunoprecipitation buffer (20 mM Tris, pH 7.5, 137 mM sodium chloride, 1 mM magnesium chloride ($MgCl_2$) and 1 mM calcium chloride ($CaCl_2$)). On-bead digestion of protein was performed by incubating the beads in 50 mM triethylammonium bicarbonate and 5 mM dithiothreitol (DTT) at RT for 60 min, with shaking at 1,200 r.p.m. Iodoacetamide was added to the mixture at a concentration of 20 mM, and continued shaking at 1,200 r.p.m. in the dark for 60 min. Trypsin was added to the mixture and incubated overnight with shaking at 900 r.p.m. The samples were frozen at −80 °C and then analyzed with MS.

Sample desalting

Before MS analysis, samples were desalted using a 96-well plate filter (Orochem) packed with 1 mg of Oasis HLB C-18 resin (Waters). Briefly, the samples were resuspended in 100 µl of 0.1% trifluoroacetic acid (TFA) and loaded onto the HLB resin, which was previously equilibrated using 100 µl of the same buffer. After washing with 100 µl of 0.1% TFA, the samples were eluted with a buffer containing 70 µl of 60% acetonitrile (ACN) and 0.1% TFA and then dried in a vacuum centrifuge.

LC–MS/MS acquisition and analysis

Samples were resuspended in 10 µl of 0.1% TFA and loaded onto a Dionex RSLC Ultimate 300 (Thermo Scientific), coupled online with an Orbitrap Fusion Lumos (Thermo Scientific). Chromatographic separation was performed with a two-column system, consisting of a C-18 trap cartridge (internal diameter of 300 µm, length of 5 mm) and a PicoFrit analytical column (internal diameter of 75 µm, length of 25 cm) packed in-house with reversed-phase Repro-Sil Pur C18-AQ 3 µm resin. To analyze the proteome, peptides were separated using a 60-min gradient from 4–30% buffer B (buffer A, 0.1% formic acid; buffer B, 80% ACN + 0.1% formic acid) at a flow rate of 300 nl min^{-1}. The mass spectrometer was set to acquire spectra in data-dependent acquisition mode. Briefly, the full MS scan was set to 300–1,200 *m/z* in the orbitrap with a resolution of 120,000 (at 200 *m/z*) and an automatic gain control target of 5×10^5. MS/MS was performed in the ion trap using the top speed mode (2 s), an automatic gain control target of 1×10^4 and an HCD collision energy of 35.

Proteome raw files were searched using Proteome Discoverer software (v2.5, Thermo Scientific) using SEQUEST search engine and the SwissProt mouse database (updated January 2023). The search for total proteome included variable modification of N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. Trypsin was specified as the digestive enzyme with up to two missed cleavages allowed. Mass tolerance was set to 10 pm for precursor ions and 0.2 Da for product ions. The peptide and protein false discovery rate was set to 1%. Following the search, data were processed as described[53]. Briefly, proteins were $\log_2$ transformed, normalized by the average value of each sample and missing values were imputed using data from a normal distribution that were two standard deviations lower than the mean. Statistical regulation was assessed using a heteroscedastic *t*-test (if *P* value < 0.05). Data distribution was assumed to be normal, but this was not formally tested. To prioritize proteins of interest that were enriched in wild-type TCF-1 immunoprecipitation compared to both ΔL1 and EV, proteins were ranked using an enrichment score calculated for each comparison (wild-type TCF-1 immunoprecipitation versus EV immunoprecipitation and wild-type TCF-1 immunoprecipitation versus ΔL1 immunoprecipitation) using the product of the fold change and −log of the *P* value. Proteins were then filtered for non-differential enrichment in input samples. Proteins with the top 100 enrichment scores were plotted using Cytoscape to create a network of L1-dependent protein–protein interactions. The stringApp was utilized with the tool STRING: protein query for visualization of entire network or network of first neighbor proteins to *Tcf7*.

Immunofluorescence

TCF-1 wild-type and mutant transduced NIH 3T3 cells were plated on poly-L-lysine-treated glass slides and allowed to adhere for 2 h in a humidified chamber and then flooded with medium and returned to the incubator overnight. Wild-type, ΔL1 and EV GFP fusion constructs were transduced into Scid.adh DN3 cells. Cells were collected after 48 h and sorted according to the same level of GFP expression. Cells were fixed on slides for 10 min with 4% formaldehyde at RT, followed by permeabilization with 0.5% Triton X-100 in PBS for 15 min at RT. Slides were blocked for 1 h with 10% BSA in 1× PBS, and stained overnight with primary antibody (monoclonal anti-Flag M2 antibody, Sigma, F1804) at a 1:1,000 dilution. Slides were washed and stained with an AF568-conjugated goat anti-mouse secondary antibody (Invitrogen, A-11004, 1:200 dilution) for 2 h. Slides were stained with DAPI at a 1:10,000 dilution and mounted with Slowfade, Gold anti-fade reagent mounting media (Invitrogen by Thermo Fisher Scientific, cat no. S36936). Imaging was carried out on a Leica Multiphoton Confocal using a ×63 oil immersion objective with a 2.0 zoom factor, a pixel size of 58.77 nm × 58.77 nm, and *z*-stack sizes of 15 µm with a *z*-step size of 300 nm.

Flow cytometry

Single-cell suspensions were stained following standard protocols. The fluorochrome-conjugated, anti-mouse antibodies were as follows: CD44-BV785 (BioLegend, 103041, dilution of 1:400), CD45-BV650 (BioLegend, 103151, dilution of 1:400), Thy1.2 PerCPCy5.5 (BioLegend, 105338, dilution of 1:300), Ckit PE (BioLegend, 105807, dilution of 1:300), CD25 PECy7 (BioLegend, 102015, dilution of 1:350), B220-APC (BioLegend, 102015, dilution of 1:300), CD11b-BV421 (BioLegend, 101235, dilution of 1:200), Sca1-PECy7 (BioLegend, 122513, dilution of 1:200), Ter119-APC (BioLegend, 116211, dilution of 1:200), CD3-APC (BioLegend, 100311, dilution of 1:200), NK1.1-APC (BioLegend, 108709, dilution of 1:200), GR1-APC (BioLegend, 108411, dilution of 1:200),

TCRgd-APC (BioLegend, 118115, dilution of 1:200), TCRb-APC (BioLegend, 109211, dilution of 1:200), Cd11c-APC (BioLegend, 117309, dilution of 1:200), Cd19-APC (BioLegend, 152410, dilution of 1:200) and CD11b-APC (BioLegend, 101211, dilution of 1:200). Cells were stained with LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Invitrogen, L34957, dilution of 1:500) or Invitrogen eBioscience Fixable Viability Dye eFluor 780 (Invitrogen, 65-0865-14, dilution of 1:4,000) for discrimination of live cells. Resuspended cells were supplemented with 123count eBeads (Thermo Fisher Scientific, 01–1234) following the manufacturer's recommendations for cell counting. For intracellular flow cytometry of TCF-1, data were collected on an LSR II running DIVA software (BD Biosciences) and were analyzed with FlowJo v10.6.1.

RNA-seq

Cells were washed once with 1× PBS before resuspending pellet in 350 µl Buffer RLT Plus (QIAGEN) with 1% 2-mercaptoethanol (Sigma), vortexed briefly, and stored at −80 °C. Subsequently, total RNA was isolated using the RNeasy Plus Micro Kit (QIAGEN, 74034). RNA integrity numbers were evaluated using a TapeStation 2200 (Agilent), and all samples used for RNA-seq library preparation had RIN numbers greater than 9. Libraries were prepared using the SMARTer Stranded Total RNA-seq Kit v2- Pico Input Mammalian kit (Takara, 634411). Two to three biological replicates were generated for each experiment. Two separate aliquots of cells for each condition were used as technical replicates for each biological replicate. Libraries were validated for quality and size distribution using a TapeStation 2200 (Agilent). Libraries were paired-end sequenced (38 bp + 38 bp) on a NextSeq 550 (Illumina).

ATAC-seq

ATAC-seq was performed as previously described with minor modifications[54,55]. Fifty thousand cells were pelleted at 550*g* and washed with 50 µl ice-cold 1× PBS, followed by treatment with 50 µl lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM sodium chloride, 3 mM magnesium chloride and 0.1% IGEPAL CA-630). Nuclei pellets were resuspended in 50 µl transposition reaction containing 2.5 µl Tn5 transposase (FC-121-1030; Illumina). The reaction was incubated in a 37 °C heat block for 45 min. Tagmented DNA was purified using a MinElute Reaction Cleanup Kit (QIAGEN, 28204) and amplified with varying cycles, depending on the side reaction results. Libraries were purified using a QIAQuick PCR Purification Kit (QIAGEN, 28104). Libraries were validated for quality and size distribution using a TapeStation 2200 (Agilent). Libraries were paired-end sequenced (38 bp + 38 bp) on a NextSeq 550 (Illumina).

CUT&RUN

CUT&RUN was performed using sorted DN1, DN2 and DN3 cells with CUTANA ChIC/CUT&RUN Kit (EpiCypher, 14-1048), according to the manufacturer's recommendation. Briefly, between 20,000 and 200,000 live cells were sorted and nuclei were extracted, washed and allowed to adsorb onto activated Concanavalin A beads. Cells were then resuspended in recommended buffer, 0.5 mg of antibody was added, mixed well, and allowed to incubate at 4 °C overnight on a nutator. Anti-TCF-1 (Cell Signaling Technology, C63D9) was used along with positive and negative controls. Subsequently, the reactions were washed with cell permeabilization buffer and incubated with pAG-MNase, and DNA was isolated for the antibody-bound regions. At least two biological replicates were generated for each experiment. Library preparation was carried out using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) and were paired-end sequenced (38 bp + 38 bp) on a NextSeq 550 (Illumina) or 61 bp + 61 bp on NovaSeq 6000 (Illumina).

RNA-seq data analysis

The FASTQ files of RNA-seq experiments were aligned and further counted using STAR 2.7.7a with parameters '--outSAMtype BAM SortedByCoordinate --outWigType wiggle read1_5p --outWigStrand Stranded --outWigNorm RPM --quantMode GeneCounts'. Next, DESeq2 was performed to identify differentially expressed genes ($|\log_2$ fold change$| > 1$ or 0.5 and adjusted P value < 0.05). Heat maps of differential genes were created using pheatmaps in R with parameters: scale = 'row'.

ATAC-seq data analysis

The FASTQ files of ATAC-seq experiments were aligned to generate the bam file using BWA (version 0.7.17-r1188). In this process, minor chromosomes such as mitochondrial chromosome or chrY were removed using SAMtools (version 1.11). Next, duplicated reads were removed using Picard (version 2.26.7) and then the bam files were indexed using SAMtools. BigWig files were generated using bamCoverage (version 3.3.2) with parameters: 'normalizedUsing = CPM, binsize = 30, smoothLength = 300, p = 5, extendReads = 200'. For peak calling, macs2 (version 2.1.4) was used with following commands: 'macs2 callpeak --t input_file --c control --g mm --n output_path --nomodel -f BAMPE --B --keep-dup all --broad --broad-cutoff 0.25 --q 0.25'. The count data of each peak was then fed to DESeq2 for differential analysis.

CUT&RUN analysis

The FASTQ files of CUT&RUN experiments were aligned to generate the bam file using BWA (version 0.7.17-r1188). In this process, minor chromosomes such as mitochondrial chromosome or chrY were removed using SAMtools (version 1.11). Next, duplicated reads were removed using Picard (version 2.26.7) and then the bam files were indexed using SAMtools. BigWig files were generated using bamCoverage (version 3.3.2) with parameters: 'normalizedUsing = CPM, binsize = 30, smoothLength = 300, p = 5, extendReads = 200'. For peak calling, macs2 (version 2.1.4) was used with following commands: 'macs2 callpeak --t input_file --c control --g mm --n output_path --nomodel --f BAMPE --B --keep-dup all --broad --broad-cutoff 0.1 --q 0.1'. For the background (control), the bam file of IgG CUT&RUN data was used. CUT&RUN peaks from two conditions and both replicates were merged and the number of fragments in each peak were counted with bedtools. The count data of each peak were then fed to DESeq2 for differential analysis.

Deeptools analysis of ATAC-seq data

The differentially gained or lost sites were obtained using DESeq2 ($|\log_2$ fold change$| > 1$ and adjusted P value < 0.05). Next, a deeptools plot was generated with the computeMatrix function using the following parameters: reference point --referencePoint center --a 2000 --b 2000. The heat map was generated with the 'plotHeatmap' function with --kmeans 3.

Motif analysis

Homer de novo motif analysis was performed using findMotifsGenome.pl on differential peak sets identified by DESeq with options --size given --len 6, 8, 10 and background as non-differential peaks or random background.

Re-analysis of GSE82044

Microarray data from GSE82044 were reanalyzed with GEO2R to find differentially expressed genes between Gata2 knockout and control dendritic cells. Probes for Agilent-028005 SurePrint G3 Mouse GE 8x60K Microarray were collapsed to corresponding genes; for genes with multiple probes, the mean fold change and adjusted P value were utilized. Gata2 activated and repressed genes were defined as having a log fold change of greater than or less than 0.5 and −0.5 and adjusted $P < 0.05$. Overlap between differential genes upregulated and downregulated in ΔL1 compared to wild-type TCF-1-expressing DN2 cells and Gata2 activated and repressed gene lists were calculated and eCDF of the Gata2 KO versus control log fold change was plotted in R.

ImmGen analysis of gene sets

Expression values of gene sets were plotted across a curated list of 62 immune cell types. Normalized gene counts were downloaded from

the Immunological Genome Project (ImmGen; GSE109125_Normalized_Gene_count_table.csv). For gene sets of interest, scaled expression values were calculated by subtracting the mean and dividing by the standard deviation of each gene across all cell types.

Gene Ontology analysis

Pre-ranked lists of genes were used by ranking genes using estimated $\log_2$ fold change values in DESeq2 for 293T cells expressing human wild-type TCF-1 versus EV. GSEA v2.2.4 with default parameters was used to perform gene set-enrichment analysis. Metascape (https://metascape.org/gp/index.html#/main/step1) was utilized for Gene Ontology analysis of differential gene sets.

Imaging analysis

Granularity measurements were performed with CellProfiler version 4.2.5 (https://cellprofiler.org/)[56]. Image pre-processing steps were completed using Fiji[52] (ImageJ2 version 2.9.0). The 'IdentifyPrimaryObjects' tool was used to perform segmentation on maximum intensity projections with a minimum and maximum object diameter of 50 and 200 pixels, respectively. Objects outside this range were discarded along with objects that were in contact with the image border. The 'MeasureGranularity' tool was used to report the percentage of the highest-intensity pixels that were subtracted from the image within the iterative range of the granular spectrum specified. Images were subsampled by a factor of 0.25 for granularity measurements, and a subsampling factor of 0.25 was introduced for background reduction, which reduced low-frequency background variations in the image. The radius of the structuring element of interest, referring to the approximate radius of punctate objects, was set at two pixels to represent the effect of subsampling on the original maximum intensity projection images. The two-pixel structuring element radius would therefore correspond to an eight-pixel radius in the unsampled image. The granular spectrum range was specified as 40 iterations and the first iteration percentages were used to compare granularity conditions.

Expression and purification of recombinant TCF-1

cDNA encoding the full-length mouse TCF-1 protein (NCBI sequence ID: EDL33620.1) with an N-terminal 6xHis tag and TEV cleavage site separated by DYDIPTT and GSEF linkers, respectively, was cloned into a pET-derived bacterial expression plasmid (gift from S. McDonald and S. Berger, University of Pennsylvania) via NEB HiFi DNA Assembly. A single sequence-verified clone was transformed into NEB T7 Express *lysY* chemically competent *E. coli* (NEB C3010I) and plated on LB agar + carbenicillin. For this and all subsequent antibiotic selection, 100 µg ml^{-1} carbenicillin (GoldBio) was used. An overnight LB + carbenicillin starter culture was inoculated with isolated colonies of transformed T7 Express *lysY E. coli* and grown at 37 °C with vigorous shaking. Preparative-scale growth cultures were prepared using Terrific Broth (RPI) medium supplemented with 4 ml glycerol/1 l (RPI) and 10 mM magnesium sulfate (Sigma Aldrich), inoculated with starter culture (1:2,000 dilution) and carbenicillin, and grown at 37 °C with vigorous shaking until an optical density (OD) at 600 nm of approximately 0.4–0.6 was achieved. Cultures were subsequently induced with 0.4 mM IPTG (GoldBio) and grown for 12–14 h at 18 °C with vigorous shaking. Bacterial pellets were recovered via centrifugation (> 6,000 relative centrifugal force (r.c.f.), 20 min, 4 °C), resuspended in an adequate volume of Ni Wash/Lysis Buffer (60 mM $NaPO_4H_2/Na_2PO_4H$ pH 8.0, 500 mM sodium chloride (NaCl), 20 mM imidazole pH 8.0, 10% glycerol, +4 mM DTT supplemented with 1× Roche cOmplete Protease Inhibitors EDTA-free), frozen in liquid nitrogen and stored at −80 °C.

Nickel affinity pulldown for purification of recombinant TCF-1

Frozen bacterial pellets were thawed on ice and supplemented with lysozyme (CAS 9001-63-2; MP Biomedicals). Cells were lysed via sonication with ice bath submersion cooling until turbidity and color changes indicative of complete lysis were achieved (approximately 1 min sonication time per 1 l culture equivalent of cell resuspension via cycles of 10 s on, 20 s off at 60% amplitude in increments of 2–3 min total sonication time; Fisher FB505 sonicator, 500 W power, 20 kHz frequency, 0.5-inch solid probe). All subsequent liquid handling, chromatography and other purification procedures were similarly performed at 4 °C or on ice, as appropriate. Lysate was clarified via two sequential rounds of centrifugation (>10,000 r.c.f., 20 min, 4 °C) then mixed for 30 min with Ni^{2+}-NTA agarose resin (GoldBio, 1 ml 50% slurry per 2 l culture equivalent) equilibrated in Ni Wash/Lysis Buffer. Flow through was collected via gravity column and resin was sequentially washed with >15 column volumes each (CVs) of Ni wash/lysis buffer and Ni wash buffer 2 (60 mM $NaPO_4H_2/Na_2PO_4H$ pH 8.0, 300 mM sodium chloride, 20 mM imidazole pH 8.0, 10% glycerol, +4 mM DTT). Bound proteins were eluted in 3 × 5 CVs of Ni Elution Buffer (wash buffer 2 with 200 mM imidazole pH 8.0).

Ion exchange chromatography for purification of recombinant TCF-1

Nickel eluate was diluted with 10 mM HEPES/sodium hydroxide pH 7.8/10% glycerol (+ 5 mM DTT) to approximately equivalent conductivity as IEX Buffer A (20 mM HEPES/sodium hydroxide pH 7.8, 130 mM NaCl, 10% glycerol + 5 mM DTT), then loaded on a buffer A-equilibrated 5 ml HiTrap Heparin Sepharose High Performance (HP) column (Cytiva) via an AKTA Pure 25 sample pump at 2–3 ml min^{-1}. After washing with 5CV buffer A, protein was eluted (1.5 ml min^{-1}) over an 8CV gradient of 0–100% IEX buffer B (20 mM HEPES/sodium hydroxide pH 7.8, 1 M sodium chloride, 10% glycerol + 5 mM DTT), which resolved two partially overlapping major populations of protein by 280 nm absorbance that differed from each other primarily in the relative abundance and size distribution of low-molecular-weight and high-molecular-weight species by SDS–PAGE analysis, but were similarly enriched for the major species of the expression construct. Fractions corresponding to each of the earlier-eluting and later-eluting halves of this major peak ('pool 1' and 'pool 2', respectively) were separately pooled for further purification and chromatographic analysis, although only the later-eluting material was ultimately characterized by HX–MS given its apparently greater capacity for more robust ionic interactions with a DNA-like polymer.

Size exclusion chromatography for purification of recombinant TCF-1

Each heparin pool was separately concentrated via repeated centrifugation (4,000–7,000 r.c.f., 20–30-min increments with mixing in between, 4 °C) in an Amicon Ultra-4 30-kDa molecular-weight-cutoff centrifugal filter. Concentrate was transferred to a new tube and centrifuged at >20,000 r.c.f. (10 min, 4 °C) to ensure absence of any precipitate. This supernatant was loaded via 500-µl injections onto a Superose 6 Increase 10/300 GL column (Cytiva; approximately 24 ml bed volume) equilibrated in 0.2-µm-filtered HGN600 (20 mM HEPES pH 7.8, 600 mM NaCl, 5% glycerol + 5 mM DTT) and eluted over 1.5 CV at 0.5–1.0 ml min^{-1} AKTA Pure 25. Multiple injections and column runs were performed as needed for the total quantity of protein in each heparin pool concentrate. For both heparin pools, a minor void population was similarly separated from two major populations of larger and progressively smaller effective sizes at retention volumes of approximately 10–14 ml and 16–20 ml, respectively. The primary peak of this later-eluting population (hereafter, 'target peak') was enriched for the apparently near-full-length expression construct with only minimal appreciable proteolysis or degradation by SDS–PAGE. A minimal number of equivalent fractions from separate pool 2 Superose 6 runs corresponding to the approximate center of the target peak were combined and dialyzed against 1 l (>1,500-fold excess by volume) of HGN280 (20 mM HEPES pH 7.8, 280 mM NaCl, 5% glycerol + 5 mM DTT) for 16 h (Thermo Scientific Slide-A-Lyzer 2-kDa molecular-weight-cutoff MINI Dialysis Device, approximately 100 µl per device). Before HX–MS analysis, combined

dialyzed material was filtered using 0.22-µm Ultrafree-MC GV Durapore centrifugal filters (Millipore Sigma) pre-equilibrated in HGN280.

To determine whether target peak species were potentially subject to time-dependent aggregation after the initial Superose 6 purification, remaining portions of additional unpooled, undialyzed fractions corresponding to the 10–14 ml peak and a region spanning the target peak (but not used for HX–MS) were separately pooled several days after the conclusion of HX–MS data acquisition. These pools were supplemented with fresh DTT in excess of existing DTT by approximately 5 mM, separately concentrated, 0.1-µm centrifugal-filtered (Ultrafree-MC PVDF, MilliporeSigma), injected in a sample volume of 100 µl onto a Superose 6 Increase 10/300 GL column (GE Healthcare) equilibrated at RT in fresh, 0.1-µm-filtered 20 mM HEPES/sodium hydroxide pH 7.8/600 mM NaCl (+ 10 mM DTT), and analyzed by 280 nm absorbance throughout continuous elution at 0.5 ml min^{-1}. Acquisition of these data was performed at the Johnson Foundation Structural Biology and Biophysics Core at the Perelman School of Medicine with assistance from Core staff.

SDS–PAGE gel electrophoresis for purification of recombinant TCF-1

SDS–PAGE analysis was performed using 4–20% Mini-PROTEAN TGX precast gels (Bio-Rad) with 25 mM Tris/192 mM glycine pH 8.3/0.1% SDS electrophoresis buffer. Gels were stained using either Coomassie G-250 or SYPRO Orange (Thermo Fisher) according to the manufacturer's instructions and imaged on either an Epson document scanner (Coomassie stain) or a GE Typhoon fluorescence imager.

Hydrogen–deuterium exchange mass spectrometry overview

H-to-D exchange (HX) of recombinant full-length, N-terminally 6xHis-tagged mouse TCF-1 protein was queried via electron spray ionization (ESI) MS essentially as described[57,58] using a Thermo Scientific Q Exactive Mass Spectrometer (calibrated every 24 h according to the manufacturer's instructions) at the Perelman School of Medicine Johnson Foundation Structural Biology and Biophysics Core. For liquid chromatography (LC)-based protein digestion and peptide separation, a custom LC system contained within a Peltier cooling chamber set at 0 °C was used that consisted of an injection valve-controlled 50-µl sample loop with downstream pepsin protease column (Thermo Scientific POROS AL 20 µm, 2.1 × 30 mm loaded with Sigma pepsin) in-line with a C8 trap column (TARGA C8 5 µm, 5 × 1.0-mm Piccolo column, Higgins Analytical TP-M501-C085), with isocratic flow of 50 µl min^{-1} 0.1% formic acid + 0.05% TFA; 3 min after initiating flow through the sample loop, flow through the trap column was diverted from waste to a separate path driven by an Eksigent gradient pump to elute peptides from the trap onto an analytic C8 column (TARGA C8 5 µm, 50 × 0.3 µm, Higgins Analytical TS-05M3-C085) via 6 µl min^{-1} of 10% ACN (buffer A, 0.1% formic acid + 0.05% TFA; buffer B, 100% ACN), which was further developed over sequential 15-min and 5-min linear gradients of 10–40% and 40–60% ACN, respectively, with continuous elution onto the ESI path followed by MS peptide separation. For initial identification of the digested peptides obtained under our conditions and their respective retention times, two sequential replicates of high resolution all-^{1}H, tandem MS/MS spectra were acquired in positive-ion mode (Thermo Scientific Q Exactive), with search exclusion of peptides identified from the first MS/MS replicate during the second replicate of MS/MS acquisition. For all subsequent D-containing samples, only single MS positive-ion mode spectra were acquired as previously described[57,58].

Preparation of HX–MS samples

Dialysis of pooled Superose 6 fractions described above were set up such that the final estimated concentration of TCF-1 protein in each HX sample was approximately 2–3 µM (calculated from A280 using ε280 = 41830). Each HX sample was generated by mixing 10 µl of filtered, dialyzed TCF-1 protein with 2 µl of 60 mM DTT (prepared in DTT-free HGN280), followed by rapid manual addition on ice of 48 µl deuterium oxide dilution buffer (92.5 mM sodium chloride and 5% glycerol prepared in deuterium oxide with one of the following buffer components: for HX at measured pH 7.0, 20 mM HEPES/KOD with measured pH 7.03; for HX at measured pH 6.0, 20 mM MES/KOD with measured pH 5.91; for HX at measured pH 5.0, 20 mM MES/DCl with measured pH 2.44). Before use, each deuterium oxide dilution buffer was freshly supplemented with 2.5 mM TCEP/KOD, prepared in deuterium oxide as a 1 M stock with measured pH 4.62. This setup achieves a final sample deuterium oxide composition of 80% in a background of 20 mM HEPES or MES (with final measured pH of 7.0, 6.0 or 5.0 as above), 130 mM NaCl, 5% glycerol + 2 mM TCEP. Stocks of DCl and KOD used for adjusting the measured pH of each solution were prepared using deuterium oxide. After the specified HX time, this 60 µl mixture was rapidly transferred with mixing to a new tube on ice containing 8.4 µl (pH 7.0 HX) or 5.4 µl (pH 6.0 HX) of 300 mM phosphoric acid (prepared in water), or 3.6 µl of 250 mM phosphoric acid (pH 5.0 HX), to lower the measured pH of each respective sample to 2.44–2.45 and quench H-to-D exchange. Around 50 µl of the quenched sample was immediately loaded into a pre-cooled glass Hamilton syringe and rapidly injected onto the LC sample loop described above. Sufficiently homogeneous mixing of protein with deuterium oxide dilution buffer was achieved via the described pipetting steps, which were chosen to allow for reproducible pipetting with very short HX times < 10 s. Quench conditions were evaluated empirically for each sample series. The measured pH of the sample mixture during both its HX and quenched states was repeatedly verified in advance using scaled larger volume, simulated mixtures of all components (exact lot) and an accupHast pH electrode (Fisher Scientific) freshly calibrated at four points over pH 1.64 to 10.00 with commercial Fisher pH standards. For the all-^{1}H sample used for MS/MS, 10 µl from an undialyzed portion of the size exclusion-purified material in HGN600 was mixed with 2 µl of 30 mM DTT (prepared in DTT-free HGN280), diluted with 48 µl HGN25 (20 mM HEPES pH 7.8, 25 mM sodium chloride, 5% glycerol + 5 mM DTT) to achieve a final sodium chloride concentration of ~130 mM, mixed with 5.4 µl of 300 mM phosphoric acid (prepared in water) to achieve a final pH of approximately 2.1–2.3, then 50 µl of this material was injected immediately as described above.

Analysis of HX–MS data

HX–MS data were analyzed essentially as described using ExMS2 (ref. 59) with the sample pD (given 80% deuterium oxide/20% water) for each condition estimated as the pH_{meas} + 0.4 (ref. 60). The 'preload' data file used by ExMS2 for generating the reference peptide list against which all experimental HX–MS spectra were compared was generated using Proteome Discoverer software (SEQUEST search with default parameters, modified as appropriate, for recombinant TCF-1 protein sequence against a custom database of off-target/decoy protein sequences). To empirically account for the D-to-H back exchange that occurs continuously, even after 'quenching' and transfer onto the LC system, the ExMS2-derived number of incorporated deuterium (D) atoms (versus time) for each condition and peptide observation (sample observation centroid *m/z* – corresponding all-^{1}H centroid *m/z*) was either (i) normalized to the corresponding ExMS2 calculation for a pH_{meas} 7.0 sample with HX time of approximately 23 h to give the percentage of deuterium uptake (Fig. 1b and Extended Data Fig. 1b), or (ii) scaled by the quantity maxD/(pH_{meas} 7.0 23 h centroid *m/z* – corresponding all-^{1}H centroid *m/z*) to give the back exchange-corrected number of incorporated D (Fig. 1c and Extended Data Fig. 1d,e), where maxD is the number of amino acids in an observed peptide – the number of prolines – 2. ExMS2-generated representative examples of the uncorrected number of incorporated D measured from *m/z* differences in centroid distributions are shown in Extended Data Fig. 1c. Given the effectively saturated exchange observed after 20 min under pH_{meas} 7.0 conditions, this 23-h sample is a reasonable estimate of an 'all-D' sample.

Visualizations of the percentage of deuterium uptake across the TCF-1 protein sequence (Fig. 1b and Extended Data Fig. 1b) or exchange versus time (Fig. 1c and Extended Data Fig. 1d,e) were generated using R. For Fig. 1c, the actual time values for pH_{meas} 5.0 and 7.0 samples were multiplied by 0.1 and 10, respectively, to scale the time for all samples relative to a pH_{meas} 6.0 timescale (chemical exchange rates increase tenfold with each pH increase of 1.0, so similar scaling can be applied to protein samples under the assumption that there are no pH-induced structural changes over the pH range of interest, in which case there would be a clear absence of equivalence between time t at pH_{meas} 5.0 and time $0.1 \times t$ at pH_{meas} 6.0, for example). Time-scaled experimental data were then fit using nonlinear least squares regression in R to the stretched exponential function[47] $D(k_{ex}, b, t) = maxD \times (1 - \exp(-(k_{ex} \times t)^b))$, where values of the stretching factor b were not fixed, maxD was defined per peptide as above, and $D(t = 0\ s)$ was forced as 0. This analysis provides an approximate estimate of the effective observed peptide-level HX rate constant k_{ex} for each analyzed peptide, where larger values of k_{ex} generally correspond to less protection from exchange (Fig. 1c and Extended Data Fig. 1d).

For each indicated peptide sequence (in the context of the full-length, unfragmented protein), the random coil-predicted exchange versus time relationships (Fig. 1c and Extended Data Fig. 1e) were calculated as the sum of $(1 - \exp(-k_{i,pred} \times t))$ over all residues within that sequence (except for the first two N-terminal residues of the peptide and proline residues). Here, $k_{i,pred}$ is the predicted single-residue rate constant for −NH exchange (calculated for pD = 6.4 and T = 277.15 K from previously described reference parameters[19,20,59] using publicly available resources at https://hx2.med.upenn.edu/download.html) if that residue were dynamically disordered random coil subject to chemical and steric effects from neighboring residues, but not subject to protection from H-to-D exchange. To extract approximate estimates of the predicted peptide-level HX rate constants k_{pred} for L1–L7 regions (Extended Data Fig. 1e), the predicted number of incorporated D versus time for a given peptide from the sum of $(1 - \exp(-k_{i,pred} \times t))$ above were fit to $D(k_{pred}, b, t) = maxD \times (1 - \exp(-(k_{pred} \times t)^b))$, defined as before for $D(k_{ex}, b, t)$ where the stretching factor b was again not fixed. Approximate peptide-level protection factors can be estimated as k_{pred}/k_{ex}. However, we display only the calculated values of each rate constant in Extended Data Fig. 1e because of minor differences in the experimental versus predicted stretching factors and because the experimental time dimension was scaled based on the expected pH dependence of HX rates. Comparing calculated k_{ex} values between L1–L7 and HMG regions (Extended Data Fig. 1d) quantitatively confirms the differences in exchange behaviors and qualitative extent of protection between these regions, despite experimental uncertainty and possible sample heterogeneity. We emphasize, though, that this rate constant comparison is approximate because of qualitative differences in the shape of many experimental exchange versus time curves between L1–L7 and HMG peptides, which leads to differences in the stretching factors b from above. Because of similar shape differences across HMG peptides between experimental versus predicted curves, we did not extend the quantitative analysis in Extended Data Fig. 1e to HMG peptides.

Statistics and reproducibility

For all experiments, at least two biological replicate mice of matching age and sex were used. All experiments were independently reproduced 2–4 times, except for HX–MS measurements, where 1–3 technical replicates of each condition (Supplementary Table 1) were measured from a single preparation of purified recombinant TCF-1 protein. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[7]. Data distribution was not formally tested. Experimental and control groups were tested for significance in Prism 9 GraphPad software (version 9.2.0 (283), 15 July 2021) using one-way ANOVA followed by Dunnett's multiple-comparison test (NS, not significant, $^*P \le 0.05$, $^{**}P \le 0.01$, $^{***}P \le 0.0005$, $^{****}P \le 0.0001$. Data collection and analysis in this study did not require randomization and blinding. No data were excluded from the analyses.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

ATAC-seq, RNA-seq and CUT&RUN date have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession code GSE213238. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository under dataset identifier PXD043586. Proteome raw files were searched with the SwissProt mouse database (updated January 2023; https://www.uniprot.org/help/downloads). Other publicly available datasets for microarray experiments were accessed in the GEO under accession code GSE82044. Source data are provided with this paper.

Code availability

Code is available upon reasonable request.

References

48. Carleton, M. et al. Signals transduced by CD3ε, but not by surface pre-TCR complexes, are able to induce maturation of an early thymic lymphoma in vitro. *J. Immunol.* **163**, 2576–2585 (1999).
49. Ogilvy, S. et al. Promoter elements of vav drive transgene expression in vivo throughout the hematopoietic compartment. *Blood* **94**, 1855–1863 (1999).
50. Shimshek, D. R. et al. Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* **32**, 19–26 (2002).
51. Dou, Z. et al. Autophagy mediates degradation of nuclear lamina. *Nature* **527**, 105–109 (2015).
52. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
53. Aguilan, J. T., Kulej, K. & Sidoli, S. Guide for protein fold change and *P* value calculation for non-experts in proteomics. *Mol. Omics* **16**, 573–582 (2020).
54. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
55. Fasolino, M. et al. Genetic variation in type 1 diabetes reconfigures the 3D chromatin organization of T cells and alters gene expression. *Immunity* **52**, 257–274 (2020).
56. Stirling, D. R. et al. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).
57. Mayne, L. Hydrogen exchange mass spectrometry. *Methods Enzymol.* **566**, 335–356 (2016).
58. Mayne, L. et al. Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation-mass spectrometry method. *J. Am. Soc. Mass. Spectrom.* **22**, 1898–1905 (2011).
59. Kan, Z. Y., Ye, X., Skinner, J. J., Mayne, L. & Englander, S. W. ExMS2: an integrated solution for hydrogen–deuterium exchange mass spectrometry data analysis. *Anal. Chem.* **91**, 7474–7481 (2019).
60. Glasoe, P. K. & Long, F. A. Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **64**, 188–190 (1960).

Acknowledgements

We thank helpful discussions with members of the laboratory of G.V. in addition to M. Fasolino, K. Zaret and N. Speck. We thank K. Gupta and the Johnson Foundation Structural Biology and Biophysics Core at the Perelman School of Medicine (RRID: SCR_022414) for HX–MS resources. Flow cytometry data were generated in the Penn Cytomics and Cell Sorting Shared Resource Laboratory at the University of Pennsylvania

(RRID: SCR_022376), which is partially supported by P30CA016520 (Abramson Cancer Center). We thank laboratories at the University of Pennsylvania including the Marmorstein laboratory for protein purification equipment, the Berger laboratory for 6xHis-TEV bacterial expression vector, and the Lynch laboratory for Typhoon imager access. The work in this manuscript was supported by F30AI161873 (to A.V.), R01AI091627 (to I.M.), F30AI174776 (to M.A.S.), R01-CA230800 and R01-CA248041 (to R.B.F.), the Burroughs Welcome Fund, the Chan Zuckerberg Initiative Award, W. W. Smith Charitable Trust, Sloan Foundation and the NIH grants R01AI168136, R01AI168240, UC4DK112217, U01DK112217, R01HL145754, U01DK127768 and U01DA052715 (to G.V.).

Author contributions

Concepts and experimental design were provided by N.G. and G.V. with help from A.V., I.M. and M.A.S. Experiments were conducted by N.G., I.J., A.C., M.A.S., J.A. and S.S. Analysis of genomics data was done by N.G. with help from A.R.P. Imaging and analysis were done by A.J., N.G. and E.K.F. HX experiments and analysis were done by M.A.S. and N.G. with support from L.M. Experimental support and discussion were provided by Y.Z., R.B.F., E.J.W. and H.-H.X. The paper was written by N.G. and G.V.

Competing interests

E.J.W. is a member of the Parker Institute for Cancer Immunotherapy that supports research in the laboratory of E.J.W. E.J.W. is an advisor for Danger Bio, Marengo, Janssen, NewLimit, Pluto Immunotherapeutics Related Sciences, Santa Ana Bio, Synthekine and Surface Oncology. E.J.W. is a founder of and holds stock in Surface Oncology, Danger Bio and Arsenal Biosciences. The other authors declare no competing interests.

Additional information

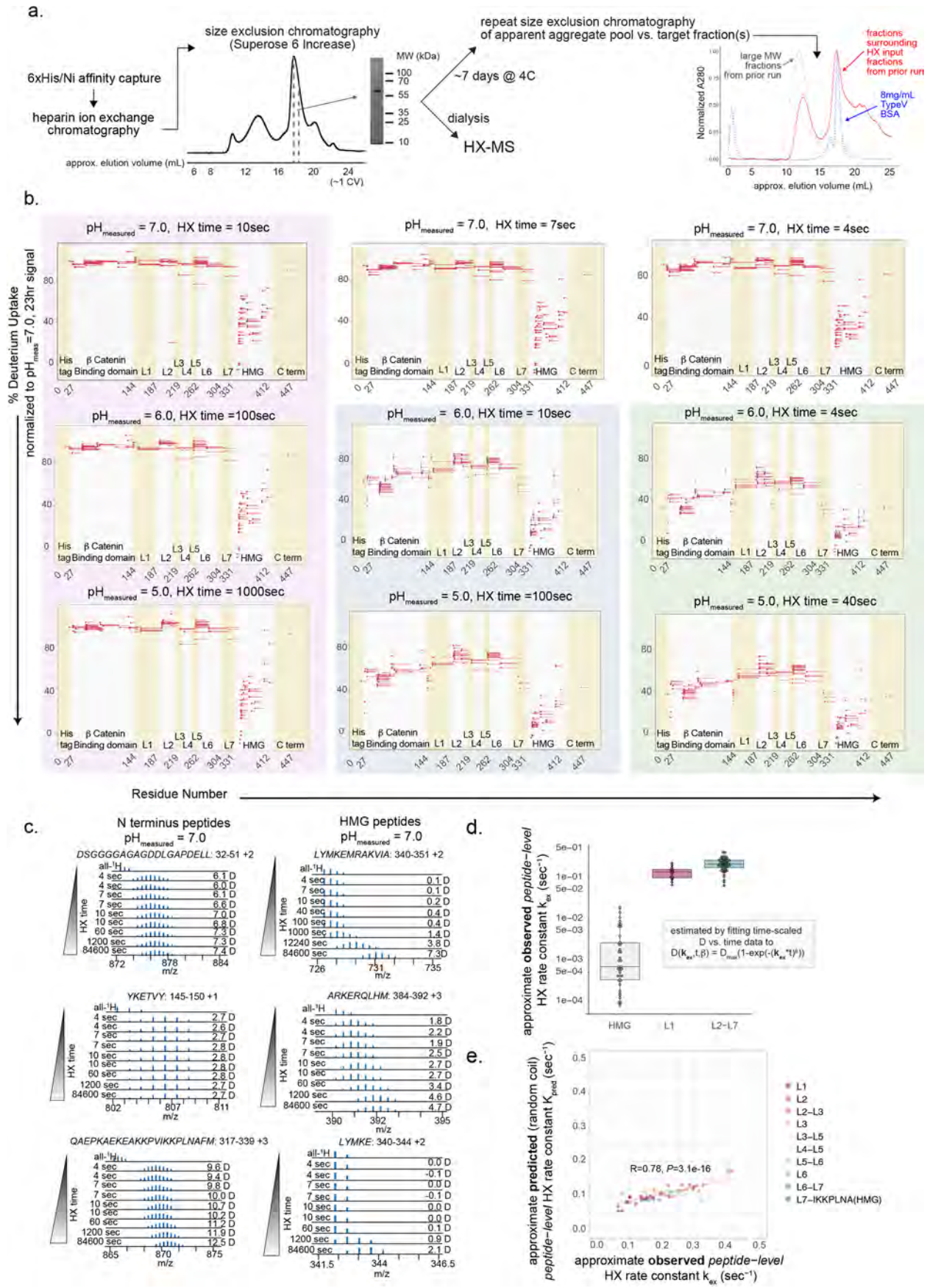
Extended data is available for this paper at https://doi.org/10.1038/s41590-023-01599-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-023-01599-7.

Correspondence and requests for materials should be addressed to Golnaz Vahedi.

Peer review information *Nature Immunology* thanks David Weiss and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: N. Bernard, in collaboration with the *Nature Immunology* team.

https://doi.org/10.1038/s41590-023-01599-7

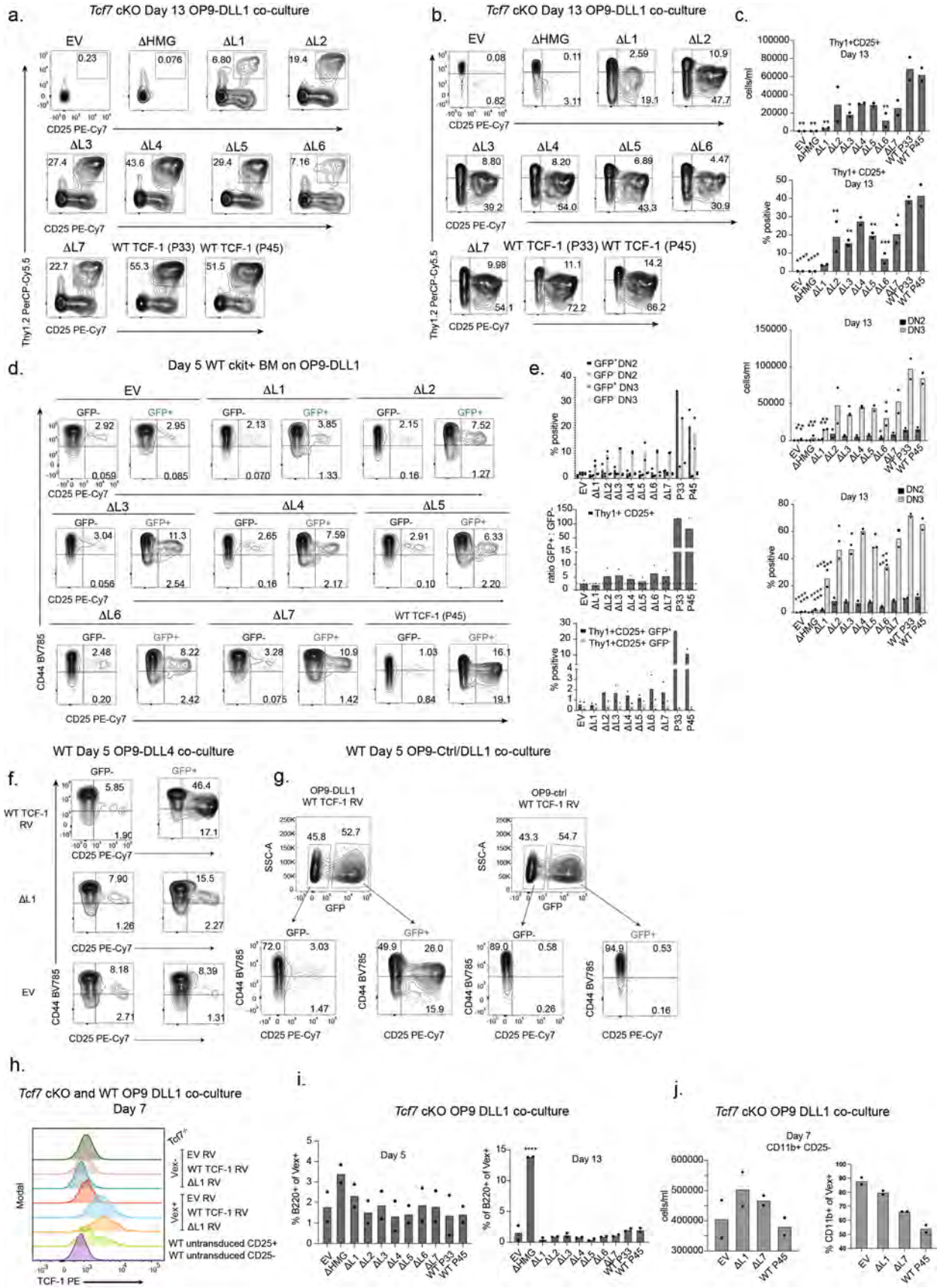


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41590-023-01599-7

Extended Data Fig. 1 | The N terminus of TCF-1 is intrinsically disordered. **a**) Size exclusion chromatography purification of affinity- and ion exchange-purified recombinant TCF-1 protein expressed in *E. coli*. Chromatogram (left) displays measured A280 vs. elution volume for a representative injection of pooled ion exchange fractions. Vertical dashed lines indicate approximate position of fractions pooled for analysis by HX-MS. Gel image displays SDS-PAGE analysis (stained with SYPRO Orange) of final protein input to HX-MS. Chromatogram (right) displays A280 vs. elution time for repeated analysis of indicated fractions from a prior Superose 6 Increase 10/300GL run during the initial purification. **b**) Plots of normalized deuterium uptake (relative to measured deuterium content after 23hrs of H-to-D exchange) at each indicated measured sample pH (pH_{meas}) for each TCF-1 peptide observation at the indicated exchange times. For observations with technical replicates (n = 3 independent samples for pH_{meas} 6.0 4sec, 10sec), center line represents mean value with error bars corresponding to standard deviation. Shaded columns indicated pH_{meas} and time conditions where approximately equivalent exchange is expected given the pH dependency of HX rates. **c**) Representative mass spectra of indicated peptide observations (generated using ExMS2[59]). Relative to the all-^{1}H sample (treated as HX time = 0sec), the change in *m/z* of each centroid distribution reflects the indicated change in mass due to deuterium incorporation. **d**) Time-scaled, back exchange-corrected deuterium content vs. time experimental data as in Fig. 1c was fit to a stretched-exponential function for each peptide across the indicated regions of TCF-1. Boxplots of estimated k_{ex} values are approximate estimates of observed *peptide-level* HX rate constants. Center line of box plots is median, limits are 1st and 3rd quartiles, and whiskers are maximum and minimum values. **e**) Comparison of approximate predicted (random coil) k_{pred} vs. observed k_{ex} *peptide-level* HX rate constants across peptides from the L1-L7 regions of TCF-1. Values for k_{ex} are as in (d). Values for k_{pred} were estimated from the stretched-exponential fitting approach using predicted deuterium content vs. time data calculated from the predicted *residue-specific* HX rate constants across each respective peptide sequence under the assumption of no protection from exchange. Pearson correlation coefficient and corresponding correlation P value are displayed.

https://doi.org/10.1038/s41590-023-01599-7

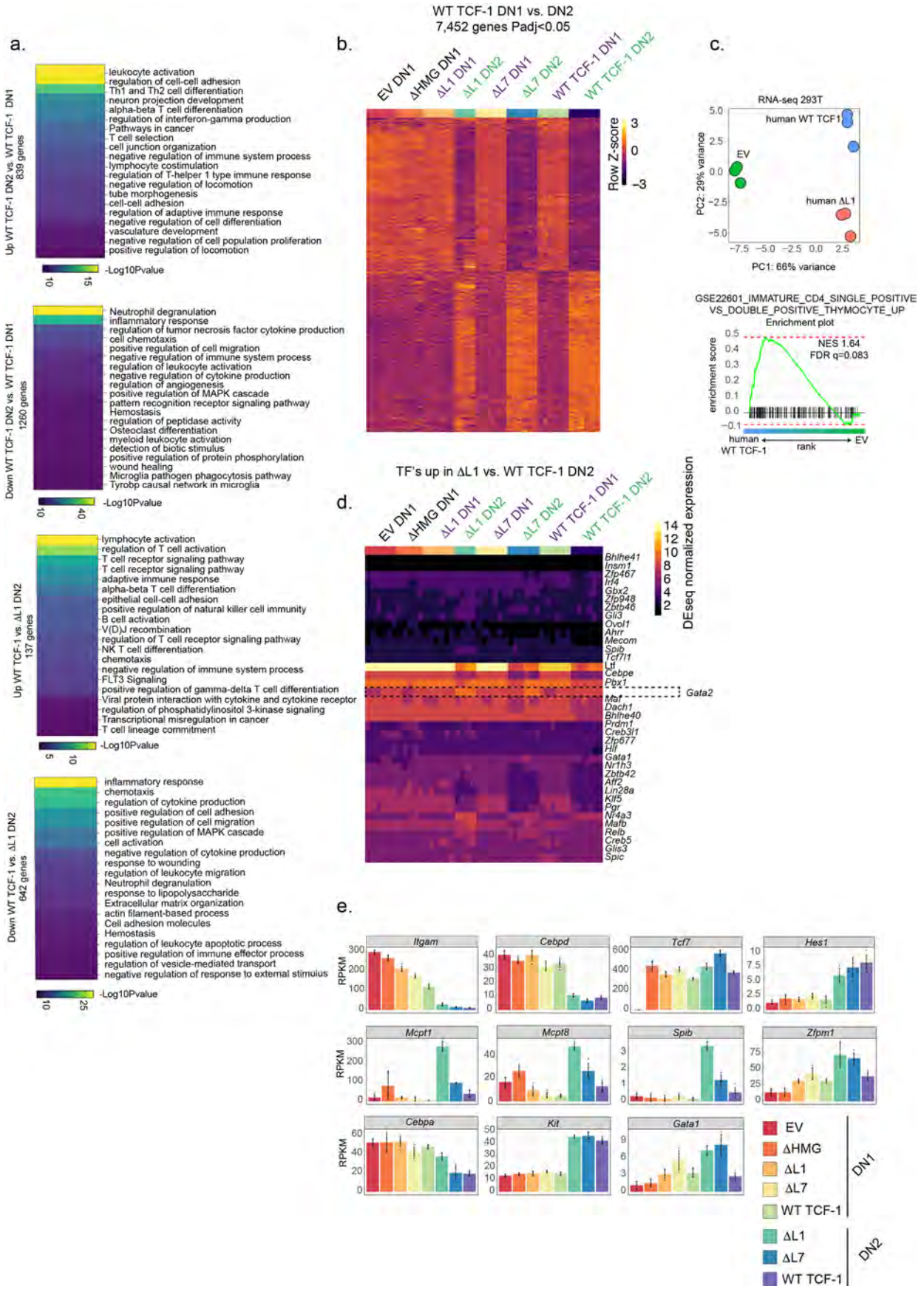


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41590-023-01599-7

Extended Data Fig. 2 | Loss of TCF-1's L1 domain limits DN1 to DN2 transition. **a**) Thy1$^+$ CD25$^+$ cells (**b**) DN1 (CD44$^+$ CD25$^-$), DN2 (CD44$^+$CD25$^+$), and DN3 (CD44-CD25+) in OP9-DLL1 co-cultures of *Tcf7* cKOs transduced with EV, WT, or mutant TCF-1 (ΔL1-7) on day 13. **c**) Quantification of Thy1$^+$ CD25$^+$ cells (top) and DN2s, and DN3s by CD44 and CD25 surface expression (bottom) in OP9-DLL1 co-cultures of *Tcf7* cKOs transduced with EV, WT, or mutant TCF-1 on day 13. **d**) DN1 (CD44$^-$ CD25$^-$), DN2 (CD44$^+$ CD25$^+$), and DN3 (CD44$^-$ CD25$^+$) cells in co-cultures of wild type (WT) ckit$^+$ bone marrow (BM) progenitors transduced with WT TCF-1, EV, or mutant TCF-1 (ΔL1-L7) (GFP$^+$) on OP9-DLL1 cells at day 5. **e**) Frequency of GFP$^+$ (transduced) and GFP$^-$ (un-transduced) DN2s and DN3s in (D) (top). Analysis of ratio of GFP+ to GFP- Thy1$^+$ CD25$^+$ cells in (d) (middle). Frequency of GFP+ and GFP$^-$ Thy1$^+$ CD25$^+$ cells in (d) (bottom). **f**) DN1 (CD44$^-$ CD25$^-$), DN2 (CD44$^+$ CD25$^+$) and DN3 (CD44$^-$ CD25$^+$) cells in WT ckit$^+$ BM progenitors transduced with WT TCF-1, ΔL1, or EV (GFP$^+$) on OP9-DLL4 after 5 days. **g**) DN1 (CD44$^+$ CD25$^-$), DN2 (CD44$^+$ CD25$^+$), and DN3 (CD44$^-$ CD25$^+$) cells in WT ckit$^+$ BM progenitors transduced with WT TCF-1 (GFP$^+$) on OP9-DLL1 cells (left) and OP9 cells (right) for 5 days. **h**) Histogram depicting TCF-1 intracellular flow cytometry in *Tcf7* cKO progenitors un-transduced (Vex$^-$) or transduced with WT TCF-1, ΔL1, or EV (Vex$^+$) as well as WT TCF-1 sufficient progenitors. **i**) Frequency of B220$^+$ Vex$^+$ (transduced) cells in *Tcf7* cKO OP9-DLL1 at day 5 (left) and 13 (right). **j**) Number (left) and frequency (right) of CD11b$^+$ CD25$^-$ cells at day 7 in *Tcf7* cKO OP9-DLL1 co-cultures. All cells (a-j) are pre-gated on SSC-A/FSC-A, Singlets, Live cell (Viability-), CD45+ and for a,b,c,i,j Vex+ and all data are representative of at least 3 independent experiments. In c, e, i,j, bars show mean from n = 2 independent animals, dots represent individual data points. In c, i, P values are determined by one-way ANOVA followed by Dunnett's multiple comparison test with WT TCF-1 (P45) as a control. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, and ****$P \leq 0.001$.

https://doi.org/10.1038/s41590-023-01599-7

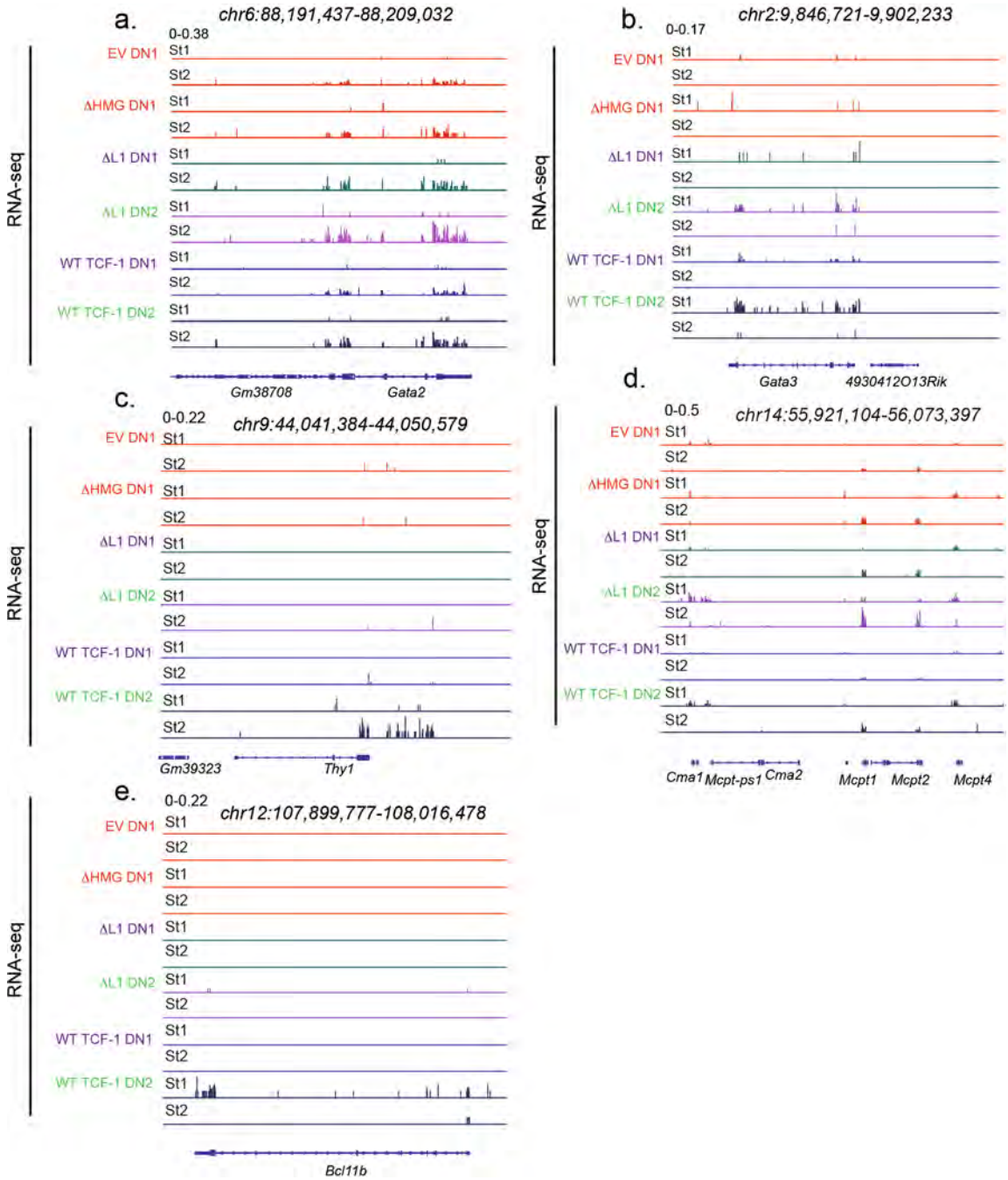


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41590-023-01599-7

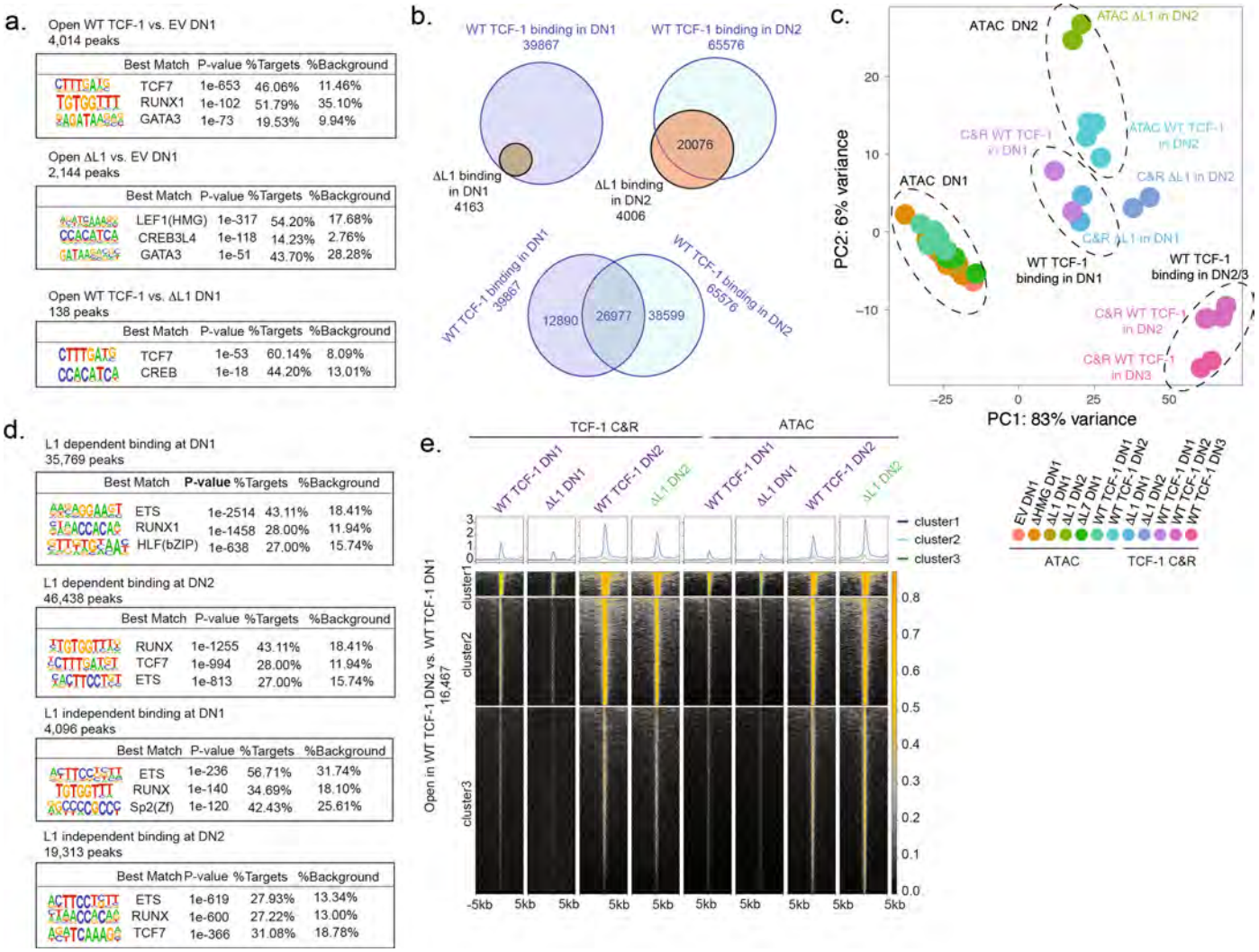
Extended Data Fig. 3 | GATA2 driven mast cell gene signature is unmasked in developing T cells lacking L1. a) Heatmaps depicting gene ontology enrichment in significantly differential gene sets (adjusted $P<0.05$, |Log2FoldChange|>1). P values are calculated using a hypergeometric test. **b**) Heatmap demonstrating differentially expressed genes (adjusted $P<0.05$) between wild type (WT) TCF-1 transduced DN1 and DN2s from *Tcf7* cKO cells on OP9-DLL1 co-cultures at day 7. P-values are calculated by the Wald test and adjusted using the Benjamini and Hochberg method. **c**) Principle component plot of RNA-sequencing on 293T human cell line transduced with empty vector (EV), wild type (WT) human TCF-1, and an internal deletion mutant lacking the analogous L1 region of human TCF-1; human ΔL1 (upper panel). GSEA depicts the enrichment of genes in GSE22601_IMMATURE_CD4_SINGLE_POSITIVE VS_DOUBLE_POSITIVE_THYMOCYTE_UP gene set within genes upregulated in 293T cells with human TCF-1 vs. EV. **d**) Heatmap depicting transcription factors differentially upregulated in ΔL1 and WT TCF-1 transduced DN2s from *Tcf7* cKO cells on OP9-DLL1 co-cultures at day 7 (adjusted $P<0.05$ and |Log2FoldChange|>1). P-values are calculated by the Wald test and adjusted using the Benjamini and Hochberg method. **e**) Bar plots depicting select gene expression (in RPKM) values in DN1 and DN2s from *Tcf7* cKO cells on OP9-DLL1 co-cultures at day 7. Bars represent mean RPKM values, error bars represent Standard deviation (SD), and individual data points are represented with dots.

https://doi.org/10.1038/s41590-023-01599-7



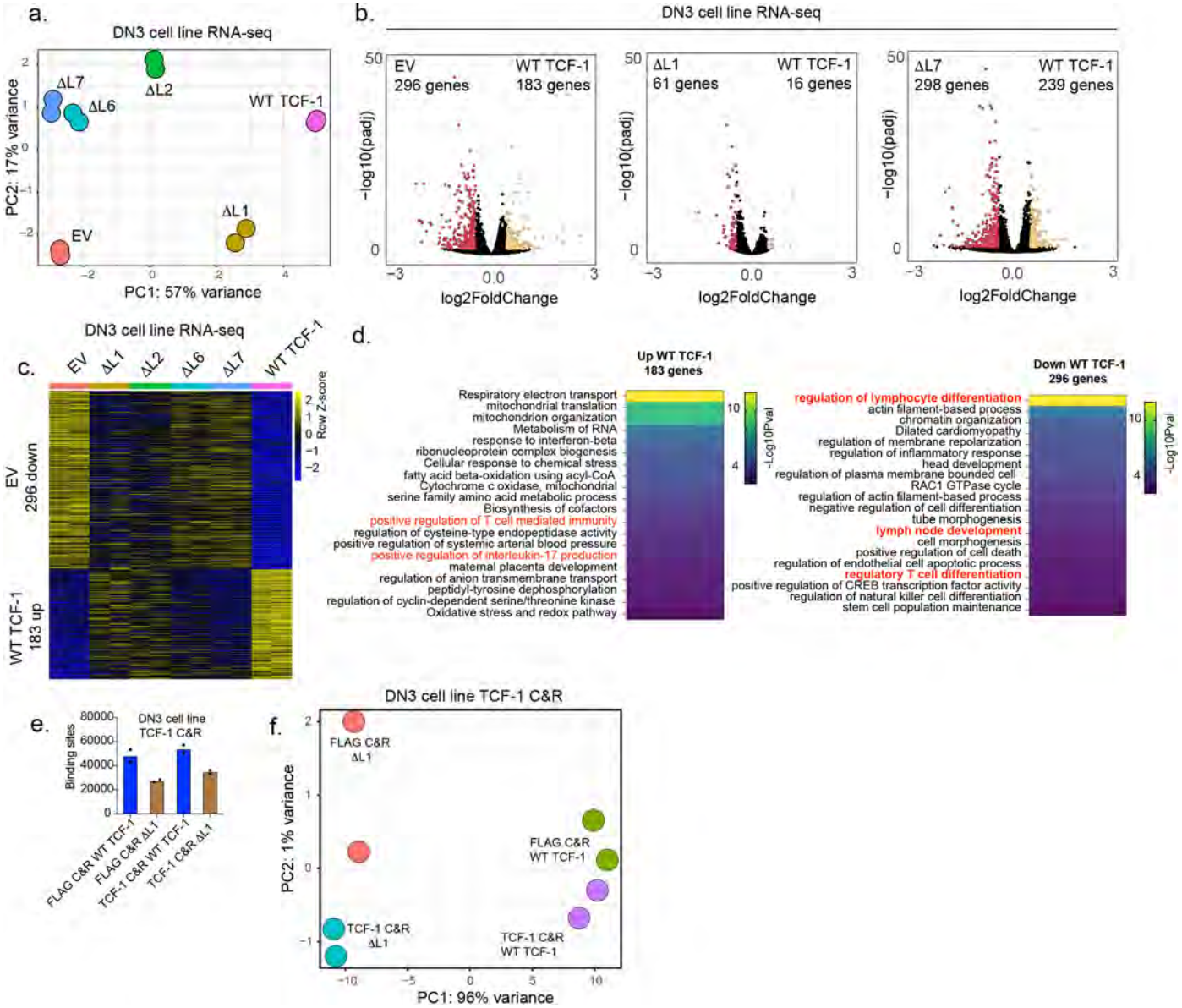
Extended Data Fig. 4 | GATA2 driven mast cell transcriptional signature is unmasked in developing T cells lacking the L1 region of TCF-1. a–**e**. Representative genome browser views of counts per million normalized strand specific RNA-seq tracks at *Gata2* (**a.**), *Gata3* (**b.**), *Thy1* (**c.**), *Mcpt1/2/4* (**d.**) and *Bcl11b* loci (**e**).

https://doi.org/10.1038/s41590-023-01599-7



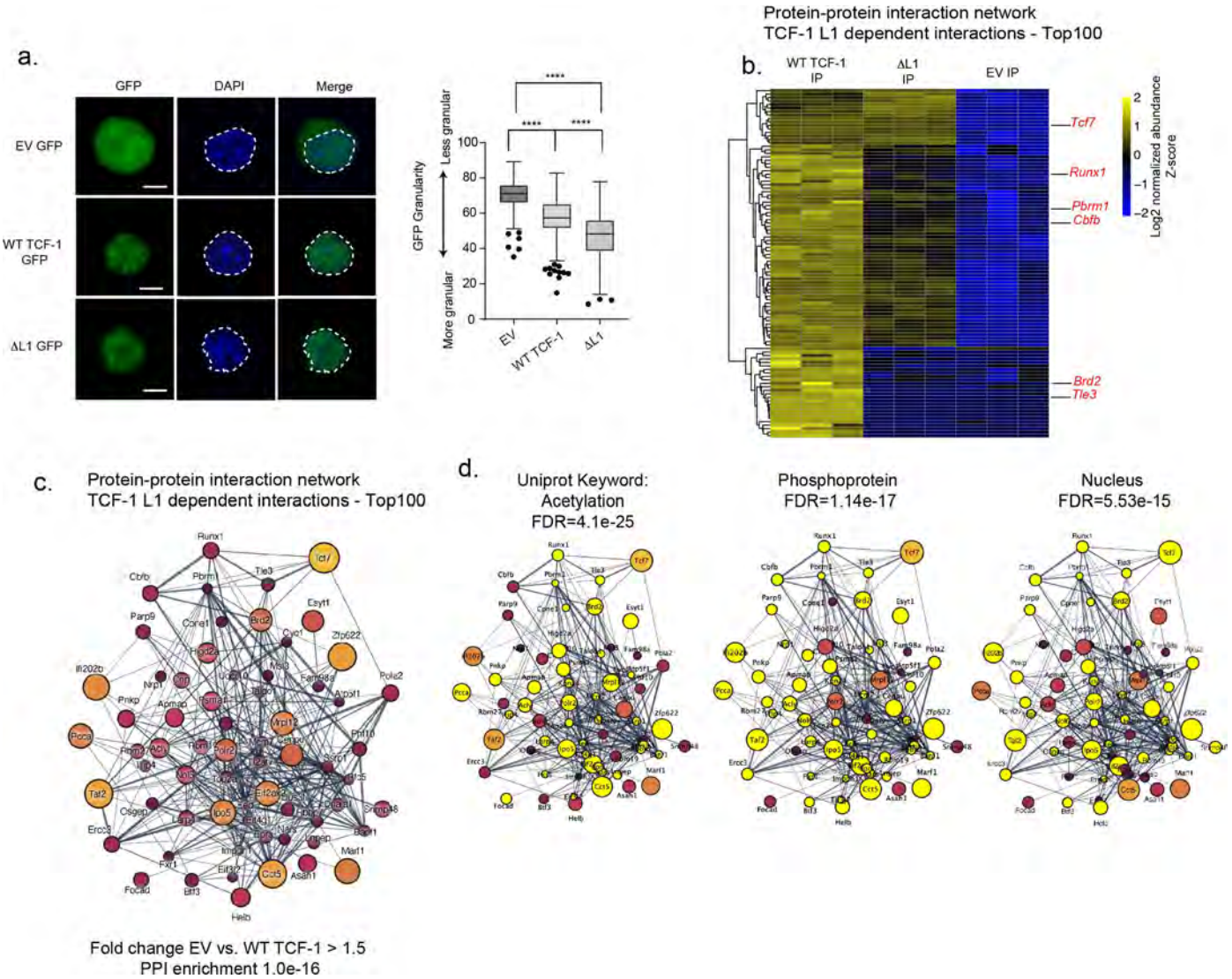
Extended Data Fig. 5 | The L1 domain of TCF-1 modulates binding and transcriptional outcomes in early T cell development independent of chromatin accessibility. a) SeqLogo depicting top enriched motifs from *de novo* HOMER motif analysis of differentially accessible ATAC-seq peaks in WT vs. EV, ΔL1 vs. EV, and WT vs ΔL1 transduced DN1 and DN2s with non-differential peaks as background. P values are calculated using a hypergeometric test. **b**) Venn-diagram representing TCF-1 CUT&RUN experiments and associated unique and overlapping WT TCF-1 and ΔL1 binding events in DN1 and DN2s. **c**) Principal component analysis of TCF-1 CUT&RUN and chromatin accessibility measurements in DN1 and DN2s. Counts of ATAC-seq and TCF-1 binding in CUT&RUN measurements were generated across the union of all peaks across all ATAC-seq and CUT&RUN conditions. **d**) SeqLogo depicting top enriched motifs from *de novo* HOMER motif analysis of L1 dependent and independent binding events in DN1 and DN2s compared to randomly generated background. P values are calculated using a hypergeometric test. **e**) Heatmap depicting TCF-1 binding events measured by TCF-1 CUT&RUN and chromatin accessibility in DN1 and DN2s at differentially accessible peaks open in WT DN2 vs. DN1s.

https://doi.org/10.1038/s41590-023-01599-7



Extended Data Fig. 6 | Loss of the L1 domain of TCF-1 has limited effect on chromatin accessibility in committed T cells. a) Principal component plot of RNA-seq on *Tcf7*$^{-/-}$ DN3 like Scid.adh cells transduced with empty vector (EV), wild type (WT) TCF-1, and internal deletion mutants: ΔL1, ΔL2, ΔL6, and ΔL7. **b**) Volcano plot demonstrating significantly differential genes comparing WT TCF-1 and EV (left), ΔL1 and WT TCF-1 (middle), and ΔL7 and WT TCF-1 (right) transduced *Tcf7*$^{-/-}$ DN3 cells. (adjusted $P<0.05$ and |Log2FoldChange|>1) P-values are calculated by the Wald test and adjusted using the Benjamini and Hochberg method. **c**) Heatmap depicting significantly up and down-regulated genes comparing WT TCF-1 and EV transduced *Tcf7*$^{-/-}$ DN3 cells. (adjusted $P<0.05$ and |Log2FC|>1). P-values are calculated by the Wald test and adjusted using the Benjamini and Hochberg method. **d**) Pathway enrichment analysis of differential gene sets depicted in B. P values are calculated using a hypergeometric test. **e**) Quantification of number of WT TCF-1 and ΔL1 binding events profiled by TCF-1 and FLAG CUT&RUN in *Tcf7*$^{-/-}$ KO DN3 cells. Bars represent mean number of binding sites from n = 2 biologically independent samples. **f**) Principal component plot of WT TCF-1 and ΔL1 binding events as measured in E.

https://doi.org/10.1038/s41590-023-01599-7



Extended Data Fig. 7 | Proteomics measurements suggest the interaction between RUNX1 and TCF-1 is dependent on the L1 domain. a) Representative immunofluorescence images depicting GFP tagged wild type (WT) TCF-1, ΔL1 mutant TCF-1 and empty vector (EV). DAPI staining of nuclei and overlay images are included (right). Boxplot of granularity of GFP signal in DN3 cells transduced with either EV, WT TCF-1 or ΔL1 fused with GFP. Granularity indicates the percentage of highest intensity elements of 8 pixels subtracted relative to the background (see Methods). Cells with a more granular pattern or punctate localization are indicated by a lower percentage. Center line of box plots represent median granularity, limits represent 1st and 3rd quartiles, whiskers represent maximum and minimum values, data points represent outliers. Cells analyzed per condition EV: n = 189, WT TCF-1: n = 237, ΔL1: n = 190. P values were determined by a two-tailed Mann-Whitney test: $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, and $^{****}P \leq 0.001$. Scale bar: 4μm. **b**) Heatmap indicating the Z score of the log2 normalized abundance of top 100 proteins detected with a higher enrichment between DN3 cells expressing WT TCF-1 and both EV and ΔL1 in mass spectrometry of a TCF-1 immunoprecipitation in DN3 cells. **c**) Depiction of L1 dependent TCF-1 protein-protein interaction network identified by mass spectrometry of a TCF-1 immunoprecipitation in DN3 cells. Node size and color indicate fold change in log normalized abundance between DN3 cells expressing WT TCF-1 and EV. **d**) Network terms corresponding to Uniprot keywords are highlighted in the network depicted in **c**.

nature portfolio

Corresponding author(s): Golnaz Vahedi

Last updated by author(s): Jul 11, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	1. In the analysis of RNA-seq data, RNA-seq samples were aligned by STAR (version 2.5.0a_alpha). HTSeq (version v0.6.1) facilitated counting RNA-seq reads on Gencode vM11 gene models. DESeq2 was subsequently applied on gene counts to identify genes differentially expressed. 2. In the analysis of ATAC-seq data, BWA (version 0.7.17-r1188) was used for alignment. Reads aligned to the mitochondrial genome or chrY as well as reads mapped to multiple genomic loci were discarded from downstream analyses. Bigwig files were generated by bedtools (version v2.27.1) genomecov and wigToBigWig normalizing tracks to tags-per-million. For peak calling, macs2 (version 2.2.6) was used. 3. In the analysis of CUT&RUN experiments The FASTQ files of CUT&RUN experiments were aligned to the bam file using BWA (version 0.7.17-r1188). In this process, minor chromosomes such as mitochondrial chromosome or chrY were removed using samtools (version 1.11). Next, duplicated reads were removed using Picard (version 2.26.7) and then the bam files were indexed using samtools. BigWig files were generated using bamCoverage (version 3.3.2) with parameters 'normalizedUsing=CPM, binsize=30, smoothLength=300, p=5, extendReads=200'. For peak calling, macs2 (version 2.1.4) was used with following commands: 'macs2 callpeak -t input_file -c control -g mm -n output_path –nomodel -f BAMPE -B –keep-dup all –broad –broad-cutoff 0.25 -q 0.25'. For the background (control), the bam file of IgG 4.CUT&RUN data was used. CUT&RUN peaks from two conditions and both replicates were merged and the number of fragments in each peak were counted with bedtools. The count data of each peak was then fed to DESeq2 for differential analysis. 5. R version 3.6.2 was used to generate boxplot, scatter plot, ecdf plot, bar plot and heatmap (ggplot2 version 3.3.3) 6. Metascape was used for gene set enrichment analysis. (https://metascape.org/gp/index.html#) 7. FlowJo (version 10.6.1) was used analysis flow cytometric data. 8. Prism 9 (Version 9.4.1 or earlier) was used to generate histograms and statistics. 9. CellProfiler (Version 4.2.5) was used to compute granularity measurements (https://cellprofiler.org/) 10. FIJI (ImageJ2 Version 2.9.0) was used for image processing.

11. ExMS2 (Kan, Z. Y., Ye, X., Skinner, J. J., Mayne, L. & Englander, S. W. ExMS2: An Integrated Solution for Hydrogen-Deuterium Exchange Mass Spectrometry Data Analysis. Anal Chem 91, 7474-7481, doi:10.1021/acs.analchem.9b01682 (2019)) was utilized for analysis of HX-MS data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

1. Publicly available datasets used in the study: GSE82044,
2. The data generated in this study is publicly available on GEO: GSE213238
3. Mass Spectrometry Data are available via ProteomeXchange with identifier PXD043586.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample sizes of two biological replicates demonstrated the differences between experimental groups and were reproducible and significant, No statistical method was used to determine sample size.
Data exclusions	No data were excluded from the analyses.
Replication	All experiments were independently replicated at least twice.
Randomization	No experiment in this study required randomization.
Blinding	No experiment in this study required blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Anti-mouse/human CD44 BV785 (IM7); dilution- 1:400; Biolegend Cat# 103041; RRID:AB_11218802
Anti-mouse CD45 BV650 (30-F11); dilution- 1:400; Biolegend Cat# 103151; RRID:AB_2565884
Anti-mouse CD90.2 (Thy1.2) PerCPCyanine5.5 (30-H12); dilution- 1:300; Biolegend Cat#105338; RRID:AB_2571945
Anti-mouse CD117 (C-kit) PE (2B8); dilution- 1:300; Biolegend Cat# 105807; RRID:AB_313216
Anti-mouse CD25 PE/Cyanine7 (PC61); dilution- 1:350; Biolegend Cat# 102015; RRID:AB_312864
Anti-mouse/human B220/CD45R (RA3-6B2) APC; dilution- 1:300; Biolegend Cat# 103211; RRID:AB_312996
TruStain FcX™ (anti-mouse CD16/32) Antibody; dilution- 1:200; BioLegend Cat# 101320; RRID:AB_1574975
Anti-mouse/human CD11b APC (M1/70); dilution- 1:200; Biolegend Cat# 101211; RRID:AB_312794
Anti-mouse Ly-6A/E (Sca1) PE (D7); dilution- 1:200; Biolegend Cat# 122513; RRID:AB_756198
Anti-mouse Ter119 APC (TER-119); dilution- 1:200; Biolegend Cat# 116211; RRID:AB_313712
Anti-mouse CD3e APC (145-2C11); dilution- 1:200; Biolegend Cat# 100311; RRID:AB_312676
Anti-mouse NK1.1 APC (PK136); dilution- 1:200; Biolegend Cat# 108709; RRID:AB_313396
Anti-mouse Ly-6G/Ly-6C GR1 APC (RB6-8C5); dilution- 1:200; Biolegend Cat# 108411; RRID:AB_313376
Anti-mouse TCRgd APC (GL3); dilution- 1:200; Biolegend Cat# 108411; RRID:AB_313376
Anti-mouse TCRb APC (H57-597); dilution- 1:200; Biolegend Cat# 109211; RRID:AB_313434
Anti-mouse CD11c APC (N418); dilution- 1:200; Biolegend Cat# 117309; RRID:AB_313778
Anti-mouse CD19 APC (1D3/CD19); dilution- 1:200; Biolegend Cat# 152410; RRID:AB_2629839
Monoclonal Anti-Flag M2 antibody Millipore Sigma; dilution- 1:1000; Cat# F1804; RRID:AB_262044
Anti-mouse TCF-1 (S33-966) PE; dilution- 1:200; BD Biosciences Cat# 564217; RRID:AB_2687845
Anti-mouse TCF1/TCF7 (C63D9) Rabbit mAb; dilution- 1:200; CST Cat# 2203; RRID:AB_2199302
Anti-RUNX1 / AML1 antibody (ab23980) Abcam; dilution- 1:200; Cat# ab23980; RRID:AB_2184205
RUNX1 Antibody (A-2); dilution- 1:200; Santa Cruz Cat# sc-365644; RRID:AB_10843207
Vinculin Antibody (H-10); dilution- 1:200; Santa Cruz Cat# sc-25336; RRID:AB_628438
Anti-rabbit IgG, HRP-linked Antibody; dilution- 1:2000; CST Cat# 7074
Anti-mouse IgG, HRP-linked Antibody; dilution- 1:2000; CST Cat# 7076
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 568; dilution- 1:200; Invitrogen Cat# A-11004

Validation

All antibodies have been validated by their manufacturer and in citations by other studies. For flow cytometric antibodies from Biolegend: "Specificity testing of 1-3 target cell types with either single- or multi-color analysis (including positive and negative cell types)."

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

Scid.adh cell line, a pro-T cell line derived from spontaneous thymic lymphomas 53, was a kind gift from Warren Pear's lab at the University of Pennsylvania, HEK293T and NIH3T3 cells from ATCC (Cat# CRL-3216; RRID:CVCL_0063 and Cat# CRL-1658 RRID:CVCL_0594). OP9-DLL1, OP9-DLL4, and OP9-Ctrl cells were a kind gift from Ivan Maillard's lab at the University of Pennsylvania.

Authentication

Cell lines were not authenticated.

Mycoplasma contamination

Mycoplasma contamination were tested periodically in all cell lines, no mycoplasma contamination was detected.

Commonly misidentified lines
(See ICLAC register)

Commonly misidentified cell lines were not used.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

Tcf7eGFP Mice (strain # 030909), Vav-iCre transgenic mice (Strain #008610), and C57BL6/J (strain # #000664) from Jackson laboratories.

Wild animals

This study did not involve wild animals.

Reporting on sex	No sex-based analysis was conducted.
Field-collected samples	The study did not involve any field-collected samples.
Ethics oversight	All mice were bred and housed in an American Association for the Accreditation of Laboratory Animal Care (AAALAC) accredited vivarium at the University of Pennsylvania. All husbandry and experimental procedures were performed according to the protocol reviewed and approved by the Institutional Animal Care and Use Committee (IACUC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Single-cell suspensions were prepared from the bone marrow (BM) removed from the femur and tibiae of 6-8 week old C57BL6/J or Tcf7 cKO mice. Ckit+ BM cells were enriched for with EasySep™ Mouse CD117 (cKIT) Positive Selection kit according to manufacturer instructions. Enriched cells where co-cultured on OP9 monolayers or stained for LSK sorting. For LSK sorting cells were stained with LD Aqua, a combination of lineage antibodies (Ter119, CD3, NK1.1, GR1, TCRgd, TCRb, Cd11c, Cd19, B220, CD11b), Sca1, and Ckit and were sorted for Viabillity-, Lin-, Ckit+, Sca1+. Ckit+ or sorted LSK cells were activated in IMDM media supplemented with 20% FBS, 1% penicillin streptomycin, SCF(100ng/ml), IL-6 (5ng/ml) and IL-3(10ng/ml)) overnight. Transduced cells where plated the following day on OP9 monolayers in OP9 media supplemented with 5 ng/mL Flt-3L and 1 ng/mL IL-7 for 5, 7, or 13 days. Co-cultures were passaged by gently disrupting cells, passage through a 40um cell strainer (Falcon) and transferred onto new OP9 monolayers every 4-5 days. Cells from co-cultures were stained with L/D APCef780, and fluorescent antibodies to B220, CD44, CD45, Thy1.2, Ckit, CD25, and CD11b.
Instrument	Data were collected on an LSRII running DIVA software (BD Biosciences)
Software	FlowJo software v10.6.1.
Cell population abundance	Post sort fraction purity was determined by re-analyzing sorted sample with original sort gates, purity was routinely above 90% for all experiments.
Gating strategy	All cell populations were pre-gated on SSC/FSC, Singlets and Viability Dye- (live cells), and CD45+. Cells were additionally gated on transduced cells (GFP+ or Vex+) as well as DN1 (CD44+ CD25-) and DN2 cells (CD44+ CD25+).

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.