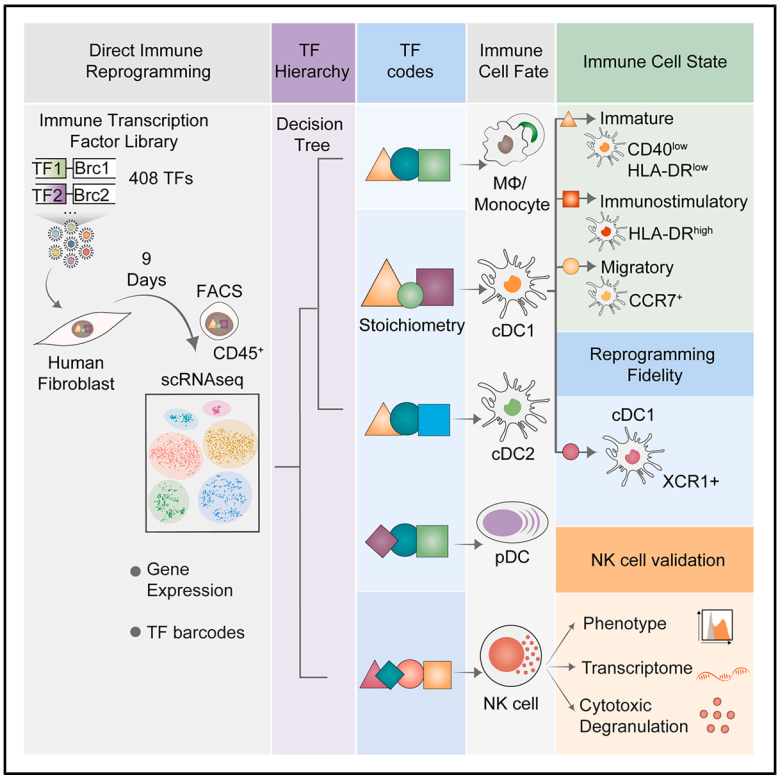
Cell Systems

Article

# A combinatorial transcription factor screening platform for immune cell reprogramming

## Graphical abstract




## Highlights

- REPROcode enables combinatorial single-cell screens of immune TFs
- Identifies TF stoichiometry, regulators of cDC1 cell state, and reprogramming fidelity
- Reveals dominant TFs that map hierarchical transcriptional networks for immune lineages
- Uncovers TF combinations for reprogramming to DC, monocyte, macrophage, and NK identities


## Authors

Ilia Kurochkin, Abigail R. Altman, Inês Caiado, ..., Ewa Sitnicka, Fabian J. Theis, Carlos-Filipe Pereira

## Correspondence

filipe.pereira@med.lu.se



## In brief

Combinatorial transcription factor (TF) networks drive cell identity, yet their logic remains elusive for inducing immune fates. Kurochkin et al. develop REPROcode, a single-cell TF screening platform that systematically maps immune reprogramming networks. REPROcode uncovers TF stoichiometry, identifies TF codes for immune lineages and states, and charts hierarchical TF frameworks.


Kurochkin et al., 2026, Cell Systems *17*, 101457
January 21, 2026 
https://doi.org/10.1016/j.cels.2025.101457

CellPress

Cell Systems

CellPress
OPEN ACCESS

Article

# A combinatorial transcription factor screening platform for immune cell reprogramming

Ilia Kurochkin,[1,2] Abigail R. Altman,[1,2,11] Inês Caiado,[1,2,3,4,5,11] Diogo Pértiga-Cabral,[1,2,3,4,5,12] Evelyn Halitzki,[1,2,12] Mariia Minaeva,[6,12] Olga Zimmermannová,[1,2,12] Luís Henriques-Oliveira,[1,2,3,4,5] Dominik Klein,[6] Malavika Nair,[1,2] Daniel Oliveira,[3,4] Laura Rabanal Cajal,[1,2] Ramin Knittel,[1,2] Cora Feick,[1,2] Markus Ringnér,[7] Marcel Martin,[8] Branko Cirovic,[9] Cristiana F. Pires,[9] Fabio F. Rosa,[9] Ewa Sitnicka,[10] Fabian J. Theis,[6] and Carlos-Filipe Pereira[1,2,3,4,9,13,*]

[1]Molecular Medicine and Gene Therapy, Science for Life Laboratory, Lund Stem Cell Center, Lund University, BMC A12, 221 84 Lund, Sweden

[2]Wallenberg Centre for Molecular Medicine, Lund University, BMC A12, 221 84 Lund, Sweden

[3]CNC - Centre for Neuroscience and Cell Biology, University of Coimbra, Largo Marquês do Pombal, 3004-517 Coimbra, Portugal

[4]CiBB - Centre for Innovative Biomedicine and Biotechnology, University of Coimbra, Coimbra, Portugal

[5]Doctoral Programme in Experimental Biology and Biomedicine, University of Coimbra, Coimbra, Portugal

[6]Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

[7]Department of Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Lund University, 223 62 Lund, Sweden

[8]Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, 171 21 Solna, Sweden

[9]Asgard Therapeutics AB, Medicon Village, 223 81 Lund, Sweden

[10]Division of Molecular Hematology, Department of Laboratory Medicine, Lund University, BMC B12, 221 84 Lund, Sweden

[11]These authors contributed equally

[12]These authors contributed equally

[13]Lead contact

*Correspondence: filipe.pereira@med.lu.se

https://doi.org/10.1016/j.cels.2025.101457

## SUMMARY

Direct reprogramming of immune cells holds promise for immunotherapy but is constrained by limited knowledge of transcription factor (TF) networks. Here, we developed REPROcode, a combinatorial single-cell screening platform to identify TF combinations for immune cell reprogramming. We first validated REPROcode by inducing type-1 conventional dendritic cells (cDC1s) with multiplexed sets of 9, 22, and 42 factors. With cDC1-enriched TFs, REPROcode enabled identification of optimal TF stoichiometry, fidelity enhancers, and regulators of cDC1 states. We then constructed an arrayed lentiviral library of 408 barcoded immune TFs to explore broader reprogramming capacity. Screening 48 TFs enriched in dendritic cell subsets yielded myeloid and lymphoid phenotypes and enabled the construction of a TF hierarchy map to guide immune reprogramming. Finally, we validated REPROcode's discovery power by inducing natural killer (NK)-like cells. This study deepens our understanding of immune transcriptional control and provides a versatile toolbox for engineering immune cells to advance immunotherapy.

## INTRODUCTION

Immunotherapy has redefined cancer treatment, with immune checkpoint blockade and chimeric antigen receptor (CAR) T cell therapies achieving unprecedented long-term responses.[1] However, clinical benefits remain limited to a subset of patients, particularly excluding immunologically "cold" solid tumors.[2] Beyond oncology, the immune system holds untapped potential to combat autoimmunity and infectious diseases, but realizing this promise requires access to a broad range of immune cell types, each specialized to perform distinct effector functions (e.g., T helper [Th]1, Th2, Th17, and regulatory).[3,4] However, sourcing immune cells remains a major bottleneck. Many are tissue-resident, difficult to isolate, or functionally compromised by disease. This restricts current immunotherapies to circulating immune cells or tumor-infiltrating lymphocytes.[1] While induced pluripotent stem cells (iPSCs) offer a renewable source of immune cells, differentiation protocols are complex, costly, and often yield immature cells resembling those derived from primitive, embryonic hematopoiesis.[5]

Direct cell fate reprogramming, whereby defined transcription factors (TFs) convert one somatic cell type into another without passing through pluripotency, offers a promising alternative to rapidly generate rare immune cells.[6] For example, the combination of PU.1, IRF8, and BATF3 (PIB) can reprogram fibroblasts and cancer cells into type-1 conventional dendritic cell (cDC1)-like cells.[7–9] This approach enables *in vivo* reprogramming through local TF delivery within tumors,[10] bypassing *ex vivo*

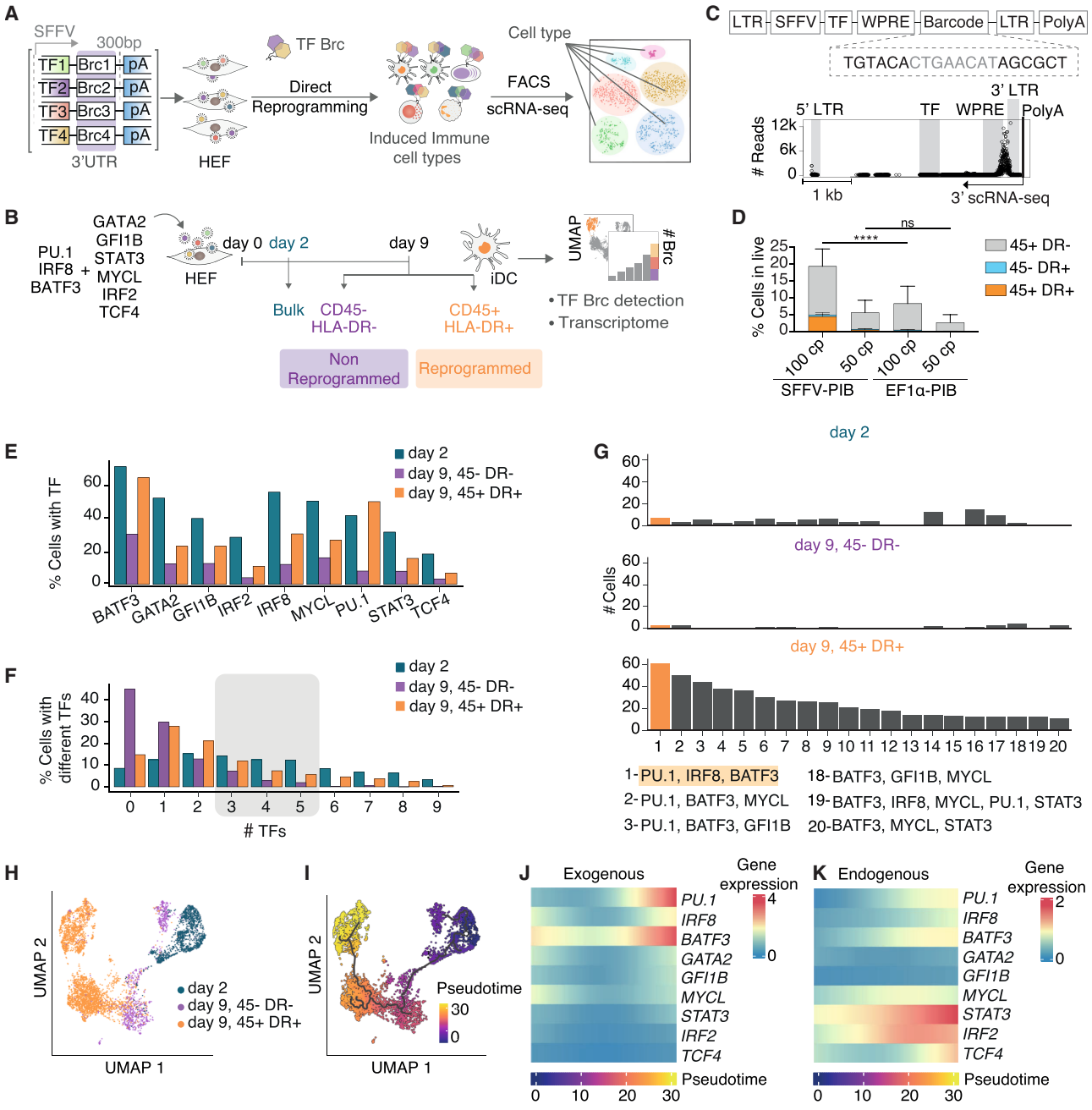



**Figure 1. REPROcode identifies instructive TF combinations for cDC1 transdifferentiation**

(A) Overview of the REPROcode platform for combinatorial TF screening. TFs were individually cloned in lentiviral vectors downstream of the constitutive splenic focus-forming virus (SFFV) promoter, coded with 8-base pair (bp) DNA barcodes (Brc) at the 3′ UTR 300 bp upstream of the poly(A) sequence (pA). A library of barcoded, lentivirally delivered immune-specific TFs was used to transduce human embryonic fibroblasts (HEFs). After 9 days, $CD45^+$ cells or subpopulations were FACS-sorted and analyzed by 3′ scRNA-seq.

(B) Platform validation using 9 TFs known to enhance or impede mouse cDC1 reprogramming,[7] including key reprogramming factors PU.1, IRF8, and BATF3 (PIB). Live (day 2), non-reprogrammed ($CD45^-HLA\text{-}DR^-$), and reprogrammed ($CD45^+HLA\text{-}DR^+$) cells (day 9) were profiled by scRNA-seq, enabling simultaneous transcriptomic profiling (UMAP) and identification of instructive TF combinations via integrated Brc.

(C) Schematic of lentiviral constructs with Brc placement optimized between woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) and long terminal repeat (LTR) for maximum read recovery in 3′ scRNA-seq. The light blue box illustrates the 6-bp flanking constant region (black) and variable 8-bp Brc (white).

(D) Flow cytometry of reprogramming efficiency at day 9 using equimolar PIB combinations at varying viral copy (cp) numbers under either the SFFV or E1Fα promoters ($n = 4$; mean ± SD; Table S1). $p$ values were calculated by one-way ANOVA and Tukey's multiple comparison test. $n$ indicates independent experiments. $^{****}p < 0.0001$; ns, non-significant.

(E) Distribution of TF-positive cells normalized to the total number of cells across conditions (live, non-reprogrammed, and reprogrammed).

*(legend continued on next page)*

manufacturing. Despite successes in preclinical models of melanoma, leukemia, and glioblastoma,[9–12] immune cell reprogramming has lagged behind advances in neural, hepatic, and cardiac reprogramming.[6] To date, validated TF cocktails exist only for macrophages, cDC1s, cDC2s, and pDC-like cells,[7,13,14] leaving most of the immune landscape, including granulocytes, monocytes, innate lymphoid cells, B cells, and T cells, largely unexplored. Collectively, these lineages comprise over 86 distinct immune cell types.[15]

A central challenge in cell fate reprogramming is identifying the minimal TF combinations to specify a target identity. Previous approaches include iterative "$n - 1$" TF dropout screens from small pools of approximately 20 TFs,[7,16] computational inference from regulatory networks,[17–20] and pooled, multiplexed screens coupled with single-cell transcriptomics.[21–24] While these methods have uncovered key reprogramming codes, including those for iPSCs,[16] hematopoietic progenitors,[25] and induced cDC1s,[7] they are often low throughput and reliant on transgenic reporter models. Predictive models based on transcriptomic or epigenomic data remain limited by assumptions of linearity and incomplete TF coverage.[17,18,26] Recent single-cell barcoding platforms that couple TF barcoding with gene expression profiling offer a scalable alternative[21,23,27,28] but cannot resolve combinatorial effects of 3–5 TFs, which often constitute the minimal instructive network for fate conversion.[29]

To address these limitations, we developed REPROcode, a high-throughput, combinatorial TF screening platform with single-cell resolution, designed to accelerate immune cell reprogramming. We first validated REPROcode by identifying and optimizing TF combinations for cDC1 induction. Using cDC1-enriched TF sets, we identified optimal TF stoichiometry, as well as co-factors that enhance PIB-driven reprogramming or drive alternative transcriptional states of cDC1s. We then applied REPROcode to a curated set of 48 TFs enriched in cDC1, cDC2, and plasmacytoid DC (pDC) lineages, enabling the identification of TF cocktails for reprogramming myeloid and lymphoid fates, and constructed a TF hierarchy map guiding immune reprogramming. Finally, we extended REPROcode to generate natural killer (NK)-like cells exhibiting defined phenotypic, transcriptional, and functional features. Together, this work establishes REPROcode as a systems-scale platform for decoding and engineering immune cell identity.

## RESULTS

### REPROcode identifies instructive TF combinations for cDC1 transdifferentiation

To evaluate whether combinations of TFs can induce immune cell fates from unrelated somatic cells in a multiplexed format, we developed REPROcode, a barcoded lentiviral library encoding immune-restricted TFs. Human embryonic fibroblasts (HEFs) were transduced with TF pools, and reprogrammed immune cells were fluorescence-activated cell sorting (FACS)-purified and analyzed by single-cell RNA sequencing (scRNA-seq). This enabled simultaneous profiling of transcriptomes and TF expression (via barcodes) at single-cell resolution to identify immune-instructive TF combinations (Figure 1A).

We first validated REPROcode by reprogramming HEFs into cDC1s using a 9-TF pool. This included the minimal cDC1-inducing combination composed by PIB,[8] along with additional enhancers (GATA2, GFI1B, and TCF4) and repressors (STAT3, MYCL, and IRF2) of *Clec9a* reporter activity previously identified in murine screens (Figure 1B).[7] Cells were FACS-sorted either at day 2 to assess initial library representation or at day 9 based on expression of CD45 and HLA-DR, markers of hematopoietic and antigen-presenting cell identity (Figures S1A and S1B).[8] We then performed scRNA-seq on non-reprogrammed ($CD45^-HLA\text{-}DR^-$), reprogrammed ($CD45^+HLA\text{-}DR^+$), and day 2 transduced cells using the 10× Chromium platform. We observed that conventional pooled virus production led to overrepresentation of TFs with smaller open reading frames (ORFs) (e.g., *MSC*, 618 bp, was overrepresented in 33.54% of the cells; Figure S1C).[21] To mitigate this bias, we used an arrayed virus production, followed by individual titration and mixing of lentiviral vectors before transduction.

Each TF construct included a unique 20-bp barcode (8-bp variable region flanked by 6-bp constants), positioned ~300 bp upstream of the poly(A) tail between the woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) and 3′ long terminal repeat (LTR) (Figure 1C). This design enabled barcode detection directly from standard polyadenylated transcripts, eliminating the need for separate barcode amplification. To ensure high TF expression, we used the splenic focus-forming virus (SFFV) promoter,[8] which increased reprogramming efficiency compared with the EF1α promoter used in iPSC systems ($CD45^+HLA\text{-}DR^+$ cells: 0.60% ± 0.11% vs. 4.74% ± 2.54% at 100 copies/cell (cp/cell); Figure 1D; Table S1).[21] All nine barcoded TFs were detected in the gene expression libraries (GEX) across samples (Figure 1E). Barcode distributions were even at day 2 and enriched in reprogrammed cells at day 9, emphasizing the contribution of combinatorial TF expression to successful reprogramming. 57.36% of reprogrammed cells contained more than one TF, and 24.87% carried 3–5 TFs (Figure 1F), demonstrating REPROcode's ability to resolve TF combinations associated with single-cell transcriptomes without additional barcode amplification steps.

To evaluate the robustness of barcode detection, we tested two multiplicities of infection (MOIs; 20 vs. 40 cp/cell). Barcode counts per cell were unaffected by MOI (Figures S1D and S1E). Approximately 10% of reprogrammed cells lacked detectable barcodes, which correlated with low unique molecular identifier (UMI) counts, indicating insufficient sequencing depth

(F) Distribution of TFs per cell normalized to the total number of cells. The 3–5 TF range (gray box) is highlighted as effective for reprogramming.
(G) Distribution of the number of cells containing combinations of 3 to 5 TF across conditions. Combinations were ordered according to the most frequent TF combination in reprogrammed cells. The 3 most frequent (left) and the 3 least frequent combinations are shown (right) (Table S1).
(H) UMAP visualization of the number of single cells captured at day 2 (1,219), non-reprogrammed (859), and reprogrammed cells (3,429), excluding cells that do not contain any TFs.
(I) UMAP with Monocle 3 pseudotime trajectory of cDC1 reprogramming (black line).
(J and K) Heatmaps showing (J) exogenous (lentiviral-derived) and (K) endogenous expression of the 9 TFs along the reprogramming trajectory ($log_2$ UMI counts).

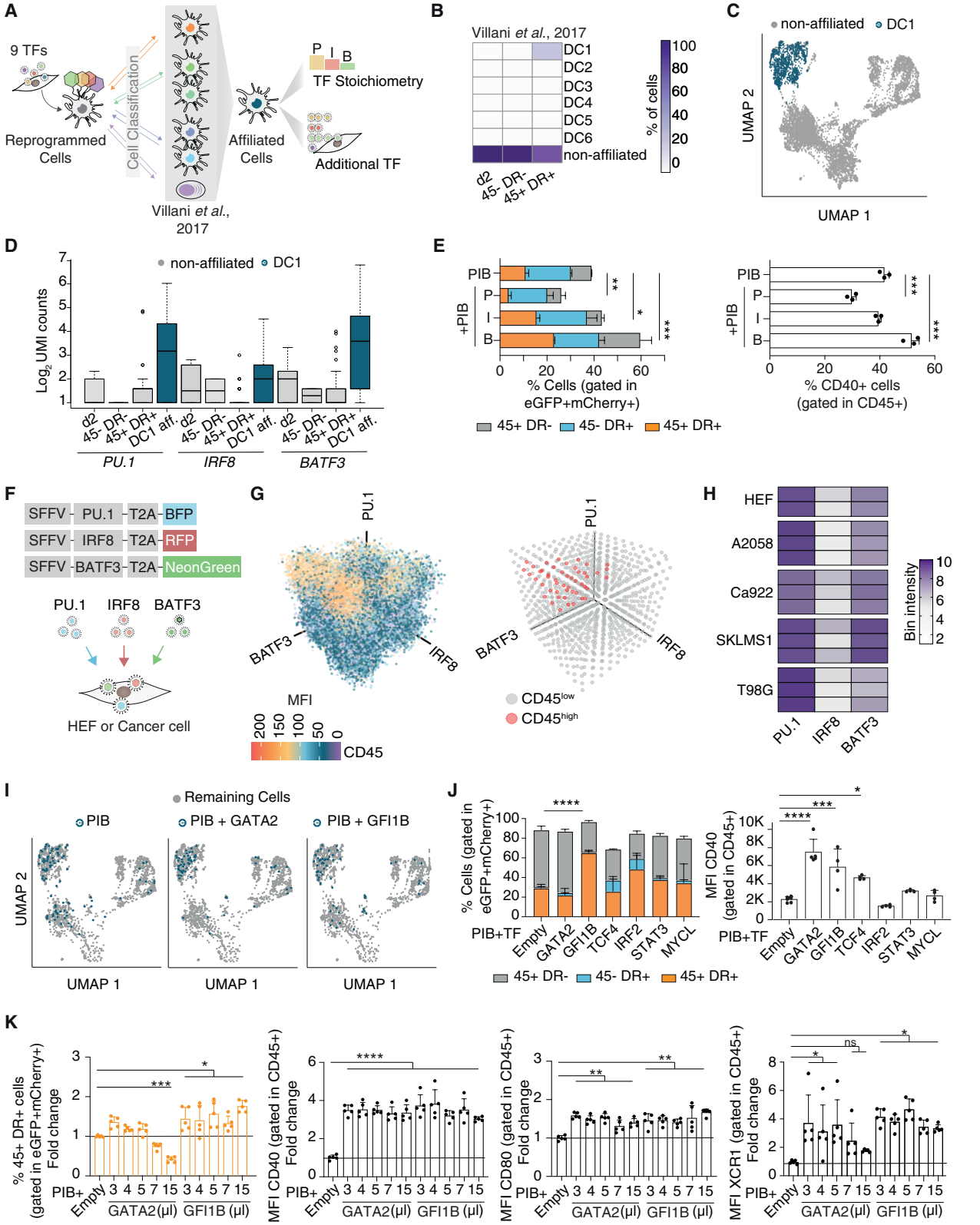



**Figure 2. Combinatorial screening reveals optimal stoichiometry and fidelity enhancers for cDC1 reprogramming**

(A) scRNA-seq data analysis strategy to identify TF stoichiometry and additional factors to enhance cDC1 reprogramming fidelity. Cells were classified using publicly available human DC transcriptional data.[31] Affiliated cells were used to infer requirements for successful reprogramming.

*(legend continued on next page)*

(Figure S1F). This highlights the importance of deep sequencing (>20,000 UMIs/cell or 40,000–80,000 reads/cell) and stringent filtering (<5,000 UMIs/cell). To further optimize barcode detection, we implemented a mismatch-tolerant algorithm allowing up to two mismatches, which increased TF recovery without reducing barcode specificity (Figure S1G). The PIB combination was the most frequent in reprogrammed cells across both MOI conditions (Figures 1G and S1H; Table S1). This combination was not enriched at day 2 or in non-reprogrammed cells, ruling out biases related to preferential co-transduction.

To reconstruct the reprogramming dynamics, we applied trajectory inference using Monocle 3.[30] UMAP visualization revealed a continuum from day 2 cells through non-reprogrammed ($CD45^-HLA\text{-}DR^-$) to two distinct $CD45^+HLA\text{-}DR^+$ reprogrammed populations, suggesting heterogeneity in reprogramming outcomes (Figures 1H, 1I, and S1I). A minor $CD45^-HLA\text{-}DR^-$ population followed a divergent trajectory, likely reflecting unsuccessful reprogramming.[8] Other cells showed intermediate states, consistent with partial reprogramming (Table S1). Pseudotime analysis revealed progressive increases in exogenous PU.1, BATF3, and more moderate IRF8 expression along the reprogramming trajectory, confirming their central role in cDC1 fate induction (Figure 1J). Activation of endogenous DC-associated TFs further supported acquisition and stabilization of a cDC1 identity (Figure 1K).

Collectively, these findings establish REPROcode as a robust and high-resolution platform for identifying minimal TF combinations that drive immune cell reprogramming in human cells.

## Combinatorial screening reveals optimal stoichiometry and fidelity enhancers for cDC1 reprogramming

We next asked whether REPROcode could resolve TF stoichiometry requirements and identify additional enhancers of reprogramming fidelity (Figure 2A). To assess reprogramming outcomes, we classified day 9 reprogrammed cells using a support vector machine (SVM) trained on a publicly available peripheral blood DC atlas.[31] All reprogrammed cells were exclusively classified as DC1 (Figure 2B) and exhibited elevated expression of *ANPEP* and *SLAMF8*, markers associated with successful human cDC1 reprogramming (Figure S2A). To investigate stoichiometric effects, we compared exogenous TF expression levels in DC1-affiliated vs. non-affiliated cells (Figure 2C). DC1-affiliated cells expressed high levels of *PU.1* and *BATF3* but lower levels of *IRF8* (Figure 2D), suggesting a refined stoichiometric requirement beyond previously established *PU.1* dosage effects.[7] To validate this, we co-transduced polycistronic PIB[8] with individual vectors encoding PU.1, BATF3, or IRF8 vectors. Supplementation with BATF3 increased both reprogramming efficiency (from 10.87% ± 1.36% to 23.03% ± 0.35%) and fidelity,[8,9] as measured by CD40 expression (from 41.93% ± 1.75% to 51.87% ± 2.75%) (Figure 2E; Table S2). We further assessed stoichiometry using fluorescently tagged TFs (PU.1-2A-BFP, IRF8-2A-RFP, and BATF3-2A-NeonGreen) co-expressed in fibroblasts and cancer cells (Figures 2F and S2B). The binned data of TF and CD45 expression revealed that high PU.1 and BATF3 levels, combined with moderate IRF8, most effectively induced reprogramming (Figures 2G and 2H; Table S2). This pattern was conserved across diverse cell types, including HEFs and cancer lines (A2058, Ca922, SKLMS1, and T98G), indicating a generalizable TF dosage relationship.

We next examined whether additional TFs enhance PIB-driven reprogramming. Among DC1-affiliated cells, co-expression of GATA2 or GFI1B (and to a lesser extent TCF4) with PIB was associated with increased frequency of DC1-affiliated cells (Figures 2I and S2C). Co-transduction of GATA2 with PIB improved reprogramming fidelity (3.5-fold; median fluorescence intensity [MFI]: 2,153 ± 315 vs. 7,441 ± 1,440), without altering efficiency (Figure 2J; Table S2). By contrast, GFI1B improved both efficiency (2.2-fold; 29.2% ± 3.5% vs. 65.0% ± 2.5%) and fidelity (2.7-fold; 2,152.8 ± 314.8 vs. 5,767.5 ± 2,029.0), consistent with

(B) Day 2 (live) and day 9 cells, identified as either non-reprogrammed ($CD45^-HLA\text{-}DR^-$; $45^-DR^-$) or reprogrammed ($CD45^+HLA\text{-}DR^+$; $45^+DR^+$) cells, were classified using human peripheral blood DC subset data[31] and the SVM algorithm. The heatmap shows the percentage of single cells affiliated to DC1-DC6 subsets. Cells not affiliating with any of the human DC subsets are indicated as unclassified.

(C) UMAP showing DC1-affiliated (blue) and non-affiliated (gray) cells.

(D) Barplot showing the distribution of lentiviral-derived (exogenous) UMI Brc counts for PU.1, IRF8, and BATF3 in DC1-affiliated (DC1 aff.) and non-affiliated (gray) cells containing the PIB combination. Day 2, $CD45^-HLA\text{-}DR^-$ ($45^-DR^-$), $CD45^+HLA\text{-}DR^+$ ($45^+DR^+$).

(E) Flow cytometry quantification of reprogrammed human embryonic fibroblasts (HEFs) reprogrammed at day 9 with a polycistronic lentiviral vector encoding PIB-IRES-dTomato co-transduced with individual lentiviral vectors encoding PU.1, IRF8, or BATF3 in combination with EGFP to increase their levels. Reprogramming efficiency was measured by the percentage of $CD45^+HLA\text{-}DR^+$ ($45^+DR^+$) cells (left) gated in live transduced cells ($eGFP^+dTomato^+$), and fidelity was measured by the percentage of $CD40^+$ cells (right) gated in $CD45^+$ cells. $n = 3$; mean ± SD; Table S2.

(F) Experimental strategy to confirm optimal levels of PU.1, IRF8, or BATF3. HEFs were co-transduced with lentiviral vectors encoding PU.1-T2A-BFP, IRF8-T2A-RFP, and BATF3-T2A-NeonGreen.

(G) Flow cytometry quantification of mean fluorescence intensity of CD45 expression in reprogrammed cancer cells (A2058) at day 3 of reprogramming. The TF axis corresponds to fluorescence level (left). Binning of CD45 signal along the PIB axes, with $CD45^{high}$ cells highlighted in red (right).

(H) Heatmap showing the scoring of individual TF bins for induction of CD45 surface expression across multiple cell lines, including HEFs and the human cancer cell lines A2058, Ca922, SKLMS1, and T98G (Table S2).

(I) UMAP visualization of cells containing only PU.1, IRF8, and BATF3 (left); PU.1, IRF8, BATF3, and GATA2 (middle); and PU.1, IRF8, BATF3, and GFI1B (right). Cells containing other combinations are colored in gray.

(J) Flow cytometry quantification of reprogramming efficiency and fidelity resultant of additional TF expression in cells co-transduced with a polycistronic lentiviral vector encoding PIB-IRES-eGFP. Additional TFs were expressed with lentiviral vectors encoding TF-IRES-dTomato. Reprogrammed ($CD45^+HLA\text{-}DR^+$; $45^+DR^+$) and partially reprogrammed populations ([$CD45^+HLA\text{-}DR^-$; $45^+DR^-$] and [$CD45^-HLA\text{-}DR^+$; $45^-DR^+$]) gated in co-transduced cells ($eGFP^+dTomato^+$) are shown (left). Quantification of median fluorescence intensity (MFI) of CD40 gated in $CD45^+$ cells (right) ($n = 4$–6; mean ± SD; Table S2).

(K) Flow cytometry quantification of reprogramming efficiency ($CD45^+HLA\text{-}DR^+$; $45^+DR^+$) (left) and MFI of CD40, CD80, and XCR1 gated in $CD45^+$ cells resultant of co-transducing PIB with increasing amounts of GATA2 and GFI1B. The amount of GATA2 and GFI1B is expressed in μL. $n = 5$; mean ± SD; Table S2.

*p* values were calculated by one-way ANOVA and Tukey's multiple comparison test. *n* indicates independent experiments. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; ns, non-significant.

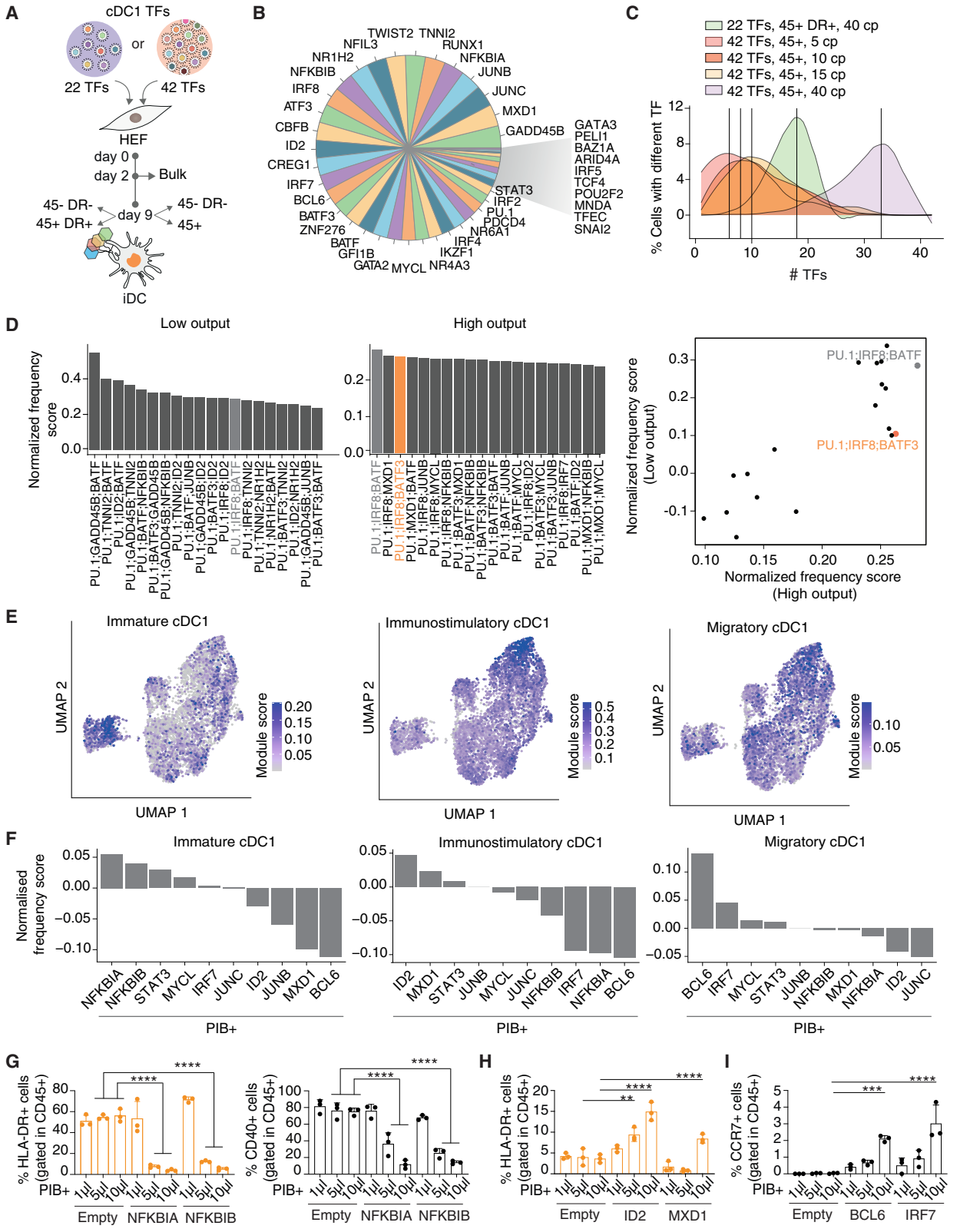



**Figure 3. Resolving instructive combinations and regulators of cell state from large TF pools**

(A) Experimental strategy to increase the pool of screened TFs during cDC1 reprogramming. Human embryonic fibroblasts (HEFs) were transduced with 22 or 42 cDC1-enriched TFs. The following populations were FACS-sorted and analyzed by scRNA-seq: (1) day 9 reprogrammed immune cells ($45^{+}DR^{+}$) and

*(legend continued on next page)*

previous murine *Clec9a* reporter studies.[7] To dissect temporal effects, we delivered GATA2 or GFI1B before or after PIB. GATA2 increased fidelity whether added before (2.5-fold; 1,593.0 ± 377.4 vs. 3,953.0 ± 837.2) or after (2.8-fold; 863.0 ± 222.0 vs. 2,447.5 ± 430.7) PIB (Figures S2D and S2E; Table S2) but reduced overall reprogramming efficiency when added before (Figure S2F, left; Table S2). GFI1B only improved efficiency when added before (4.0-fold; 9.3% ± 1.4% vs. 36.8% ± 11.7%; Figure S2F, right; Table S2).

To define optimal expression levels, we examined *GATA2* and *GFI1B* expression in DC1-affiliated vs. non-affiliated cells. Both were expressed at lower levels than the core PIB factors (Figure S2G), suggesting they act as modulators rather than primary inducers. We then varied their MOIs to assess dosage effects. Increasing GATA2 levels reduced efficiency (CD45$^+$HLA-DR$^+$: 23.2% ± 2.4% at low dose 3 μL vs. 7.1% ± 1.2% at high dose 15 μL) and decreased expression of the cDC1 marker XCR1 (MFI: 64.6 ± 33.6 vs. 30.6 ± 1.6; Figures 2K and S2H, left; Table S2). CD40 and CD80 levels plateaued beyond the lowest effective dose. By contrast, GFI1B dosage increased reprogramming efficiency (up to 30.1% ± 2.5%) without further enhancing CD40, CD80, or XCR1 expression at higher doses (Figures 2K and S2H, right; Table S2). Together, these findings reveal distinct roles for GATA2 and GFI1B: GATA2 enhances fidelity within a narrow dose window, while GFI1B boosts efficiency at higher levels. Both exhibit saturation effects, suggesting they synergize with PIB through different mechanisms.

Collectively, these results demonstrate REPROcode's ability to identify not only minimal reprogramming TF combinations but also to resolve optimal stoichiometries and fidelity-enhancing factors in a dosage- and timing-dependent manner.

## Resolving instructive combinations and regulators of cell state from large TF pools

To validate our platform under highly multiplexed conditions, we expanded the TF pool from 9 to 22 and then to 42 cDC1-enriched TFs (Figures 3A and S3A). We hypothesized that identifying instructive TF combinations from larger pools would require lower MOIs and higher cell throughput. Accordingly, HEFs were transduced with 5, 10, 15, or 40 cp/cell, FACS-sorted at day 2 (bulk) and at day 9 into non-reprogrammed (CD45$^-$HLA-DR$^-$) and reprogrammed populations (CD45$^+$HLA-DR$^+$ for 22 TFs; CD45$^+$ for 42 TFs), followed by scRNA-seq using 10× Genomics and BD Rhapsody platforms (Figure S3B). We first assessed barcode representation across the 42 TFs and observed preserved distribution even at increased pool size (Figure 3B). However, at high MOI (40 cp), cells carried a mode of 33 TFs, complicating the inference of instructive reprogramming factors. To address this, we developed a meta-combination analysis algorithm: after downsampling to "meta-combinations" followed by normalization per-cell TF load, we ranked TFs by enrichment in reprogrammed vs. non-reprogrammed cells and assembled top-ranking 3-TF meta-combinations. Using this approach, the canonical PIB combination ranked #64 at 40 cp with 42 TFs but improved to #8 in the 22-TF experiment (mode = 17 TFs/cell), underscoring the critical impact of TF load per cell (Figure S3C; Table S3). To further reduce complexity, we tested lower MOIs (5, 10, and 15 cp) with the 42-TF pool, resulting in mode TF counts of 6, 9, and 10 per cell, respectively (Figures 3C and S3D–S3F). Under these conditions, PIB meta-combination's rank improved to #13 (5 cp), #36 (10 cp), and #63 (15 cp), confirming that lower MOIs improve inference in large-pool screens. We next tested whether increasing cell recovery would enhance detection. Using BD Rhapsody in the 10 cp, 42-TF condition, we recovered 2.8× more PIB-positive cells than with 10× Genomics (629/2,273 vs. 144/1,474; Figure 3D, left; Table S3). In this optimized setting, PIB ranked third overall, and high concordance in meta-combination ranking was observed between platforms (Figure 3D, right). Together, these results demonstrate that optimizing MOI and cell throughput improves the resolution of instructive TF combinations in multiplexed screens.

Next, we explored whether individual TFs, when added to the core PIB combination, could influence the functional state of reprogrammed cDC1s. Using established gene signatures for immature, immunostimulatory, and activated migratory cDC1 phenotypes[32] (Figures 3E and S3G), we analyzed the throughput 42-TF, 10 cp dataset to determine the frequency of cells

non-reprogrammed cells (45$^-$DR$^-$), transduced with 22 TFs; (2) day 2 (live) and day 9 reprogrammed immune cells (45$^+$) and non-reprogrammed cells (45$^-$DR$^-$), transduced with 42 TFs. 28,191 cells were sequenced in 10×, and 7,070 cells were profiled for BD Rhapsody.

(B) Pie chart showing TF distribution in day 2 (live) samples generated from HEFs transduced with 42 TFs with 40 viral copies (cp) per cell.

(C) Distribution of number of TFs per single cell in reprogrammed (CD45$^+$; 45$^+$) cells with 22 or 42 TFs, normalized to the total number of cells using 22 cp (22 TF) or 5, 10, 15, and 40 cp (42 TF) per cell. Distribution is shown as a smooth area plot. The dashed line indicates mode (17 for 22 TFs; 6, 9, 10, and 33 for 5, 10, 15, and 40 cp, respectively).

(D) To identify the most frequent TF combinations, data were down-sampled toward combinations of 4 TFs. The top 10 most frequent TFs were extracted after normalizing combination frequencies between restricted-reprogrammed cells (gene expression of both *PTPRC* and *HLA-DR*) and non-reprogrammed cells (sorted CD45$^-$HLA-DR$^-$). Combinations of 3 TFs composed from the top 10 TFs were subsequently assembled and ranked based on normalized values between restricted-reprogrammed cells and all remaining cells (shown in barplots). Low output (left, 10× Genomics) and high output (middle, BD Rhapsody). The scatterplot shows the relationship between low and high output experiments (right). The top 3-TF combination is highlighted (Table S3).

(E) UMAP visualization of cells from HEFs transduced with 42 cDC1-restricted TFs (10 cp per cell) profiled in BD Rhapsody. Module score for corresponding cDC1-state signatures[32] for immature (left), immunostimulatory (middle), and activated migratory (right) overlaid on UMAP.

(F) Barplots showing the enriched TFs in the three cell states (in addition to PU.1, IRF8, and BATF3 [PIB]). Cells expressing only one of the corresponding cDC1-state signatures (for example, expressing immature but not immunostimulatory nor activated migratory) were extracted (102, 72, and 40 cells for immature, immunostimulatory, and activated migratory, respectively). The frequency of TFs was normalized between signature-assigned and non-assigned reprogrammed (CD45$^+$) cells. See methods for thresholds applied.

(G–I) Flow cytometry quantification of CD40 (G), HLA-DR (G and H), or CCR7 (I) expression after co-transduction of NFKB1A-IRES-dTomato, NFKB1AB-IRES-dTomato, ID2-IRES-dTomato, MXD1-IRES-dTomato, BCL6-IRES-dTomato, or IRF7-IRES-dTomato with a polycistronic lentiviral vector encoding PIB-IRES-GFP. Percentage of cells within CD45$^+$GFP$^+$dTomato$^+$ cells is shown ($n$ = 3; mean ± SD; Table S3).

$p$ values were calculated by one-way ANOVA and Tukey's multiple comparison test. $n$ indicates independent experiments. $^{**}p < 0.01$; $^{****}p < 0.0001$.

co-expressing PIB plus a single additional TF within each phenotypic state (Figure 3F). This analysis revealed distinct and non-overlapping TFs preferentially associated with each state: nuclear factor κB (NF-κB) inhibitors IκB-α (*NFKBIA*) and IκB-β (*NFKBIB*)[33] were associated with the immature phenotype, *ID2*[12] and *MXD1*[34] with immunostimulatory activation, and *BCL6*[35] and *IRF7* with the migratory state. These findings suggested that specific TFs, when co-expressed with PIB, may bias the functional trajectory of reprogrammed cells. To confirm these findings, we co-transduced HEFs with PIB and one of the six TFs—NFKBIA, NFKBIB, ID2, MXD1, BCL6, or IRF7—across a range of MOIs and assessed cell surface marker expression by flow cytometry. Reprogramming efficiency remained stable across conditions (Figure S3H; Table S3); however, the addition of NFKBIA or NFKBIB led to suppression of DC activation markers. HLA-DR$^+$ cells dropped from 56.2% ± 6.2% in the PIB plus empty vector control to 3.9% ± 1.0% with NFKBIA and 5.3% ± 1.3% with NFKBIB. CD40 expression similarly declined, from 75.0% ± 4.3% to 11.7% ± 5.2% and 15.0% ± 2.0%, respectively (10 μL; Figure 3G; Table S3), indicating a shift toward an immature cDC1 state. By contrast, co-expression of ID2 or MXD1 with PIB significantly enhanced the percentage of HLA-DR$^+$ cells, suggesting synergistic promotion of immunostimulatory features. HLA-DR$^+$ cells increased from 3.6% ± 1.0% in the control to 15.0% ± 2.2% with ID2 and 8.4% ± 1.0% with MXD1 (10 μL; Figure 3H; Table S3). Finally, the addition of BCL6 or IRF7 to PIB induced expression of CCR7, a marker of migratory dendritic cells. While CCR7 was nearly undetectable in cells reprogrammed with PIB alone (0.1% ± 0.0%), its expression rose to 2.1% ± 0.2% with BCL6 and 3.0% ± 1.1% with IRF7 (10 μL; Figure 3I; Table S3), consistent with the acquisition of a migratory phenotype.

Collectively, these results demonstrate that instructive TF combinations can be resolved in highly multiplexed pooled screens and reveal additional regulators of cell state polarization.

## Construction of an immune-restricted TF library

To extend the utility of REPROcode beyond cDC1 reprogramming, we constructed a comprehensive TF library tailored to immune cell lineages. We began by compiling an initial list of 1,829 candidate TFs, generated by merging a database of DNA-binding domain-containing proteins[36] with a previously reported TF collection used in cell fate engineering.[27] To prioritize immune-relevant TFs, we analyzed gene expression data across 73 distinct immune cell types from the Tabula Sapiens dataset,[37] using differential expression analysis against a pseudo-bulk reference derived from 104 non-immune cell types (Figure 4A).[37] This ranking strategy enabled the identification of TFs highly enriched in specific immune populations, including rare subsets, a feature often associated with reprogramming activity.[29] We then aggregated the top-ranked TFs across all immune cell types and supplemented the list with 61 additional TFs previously implicated in hematopoietic reprogramming,[7,38,39] resulting in a curated set of 408 immune-restricted TFs. To confirm the immune specificity of this library, we compared expression profiles of immune genes against a control set of non-immune TFs across both the Tabula Sapiens[37] and Immune Cell Atlas datasets.[15] Immune-restricted TFs (e.g., *SPI1*, *BATF3*, *PAX5*, and *FOXP3*) displayed orthogonal expression patterns relative to non-immune TFs (e.g., *ELF3*, *EHF*, *ID4*, and *SOX9*), as visualized by histogram distributions (Figure 4B). UMAP projections further confirmed this expression dichotomy, showing clear segregation between immune and non-immune TF modules in both datasets (Figures 4C and 4D). To assess the robustness of our prioritization strategy, we applied four complementary ranking approaches: log-transformed and unlogged fold change, the fraction of cells expressing each TF within immune subsets, and an entropy-based metric using Jensen-Shannon divergence (JSD) (Figure S4A). Ranking correlations were high between log and unlogged fold change ($r = 0.96$) and between log fold change and cell fraction methods ($r = 0.87$). As expected, correlation with the JSD-based ranking was lower ($r = 0.59$), given its comparison to a uniform reference distribution. These results demonstrate the internal consistency and biological specificity of our TF selection strategy. Each TF from the final 408-member library was individually cloned into SFFV-driven lentiviral vectors (SFFV-TF-BRC; Figure 1C), enabling direct use in REPROcode combinatorial screens.

## Programming immune cell diversity

To assess the ability of our platform to generate a range of immune cell types, we selected 48 TFs from our immune-restricted library. These TFs were chosen based on enriched expression in cDC1s, cDC2s, and pDCs, as well as supporting evidence from the literature (Figures 5A and S5A). HEFs were transduced with the 48-TF pool at 10 cp/cell. On day 9, cells were sorted into non-reprogrammed (CD45$^-$HLA-DR$^-$) and reprogrammed (CD45$^+$) populations (Figure S5B) and analyzed by scRNA-seq using the BD Rhapsody across three independent experiments, yielding 49,902, 121,367, and 39,544 cells, respectively. UMAP visualization revealed clear segregation between reprogrammed and non-reprogrammed populations (Figure 5B). Reprogrammed cells showed strong downregulation of fibroblast-associated gene signatures (Figure 5C), consistent with a loss of fibroblast identity. Clustering analysis of reprogrammed cells identified 22, 29, and 25 distinct transcriptional clusters across the three experiments, highlighting the heterogeneity of reprogrammed phenotypes (Figure 5D).

To classify the identity of reprogrammed cells, we first applied CellTypist, a logistic regression-based classifier trained on 357,211 cells (329,762 of which are immune).[15] This approach categorized cells into three groups: immune cells, fibroblasts, and partially reprogrammed cells (low-confidence fibroblast assignments; Figure 5E). Among the immune cell annotations, we identified signatures corresponding to macrophages, DCs, B cells, innate lymphoid cells, and in some cases epithelial and endothelial-like profiles. To complement this, we used scTab, a deep learning-based classifier trained on 22.2 million cells,[40] which stratified populations based on frequency across reprogrammed (CD45$^+$), non-reprogrammed (CD45$^-$HLA-DR$^-$), or unchanged populations (Figures S5C and S5D). Both CellTypist and scTab consistently annotated similar reprogrammed immune identities, reinforcing the accuracy and robustness of atlas-based classification tools (Figure S5E).

To investigate how many TFs are required for immune cell reprogramming, we developed a hierarchical stepwise regression model. This framework treated cell identity (e.g., immune vs. fibroblast) as a function of TF presence and incrementally

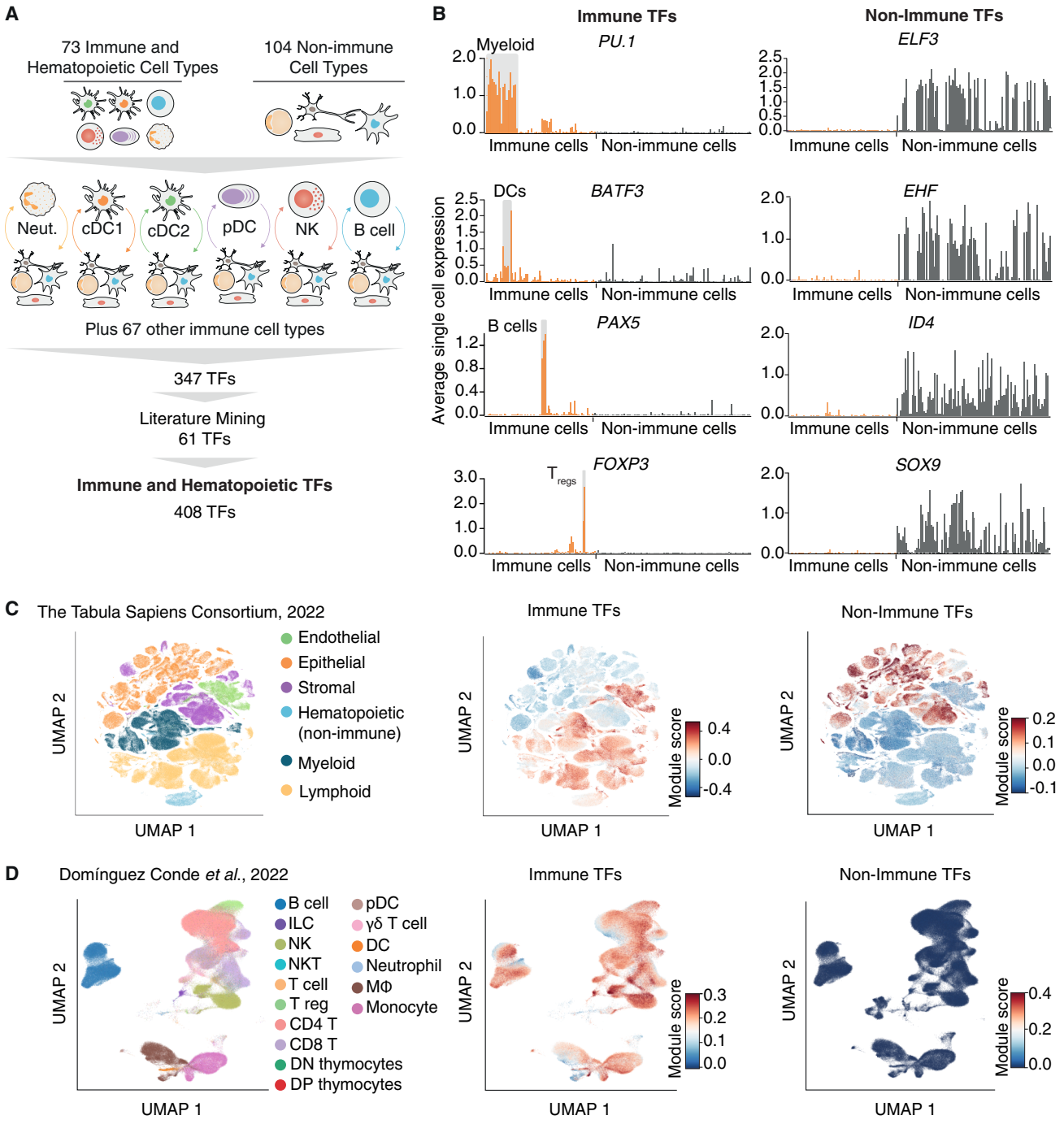


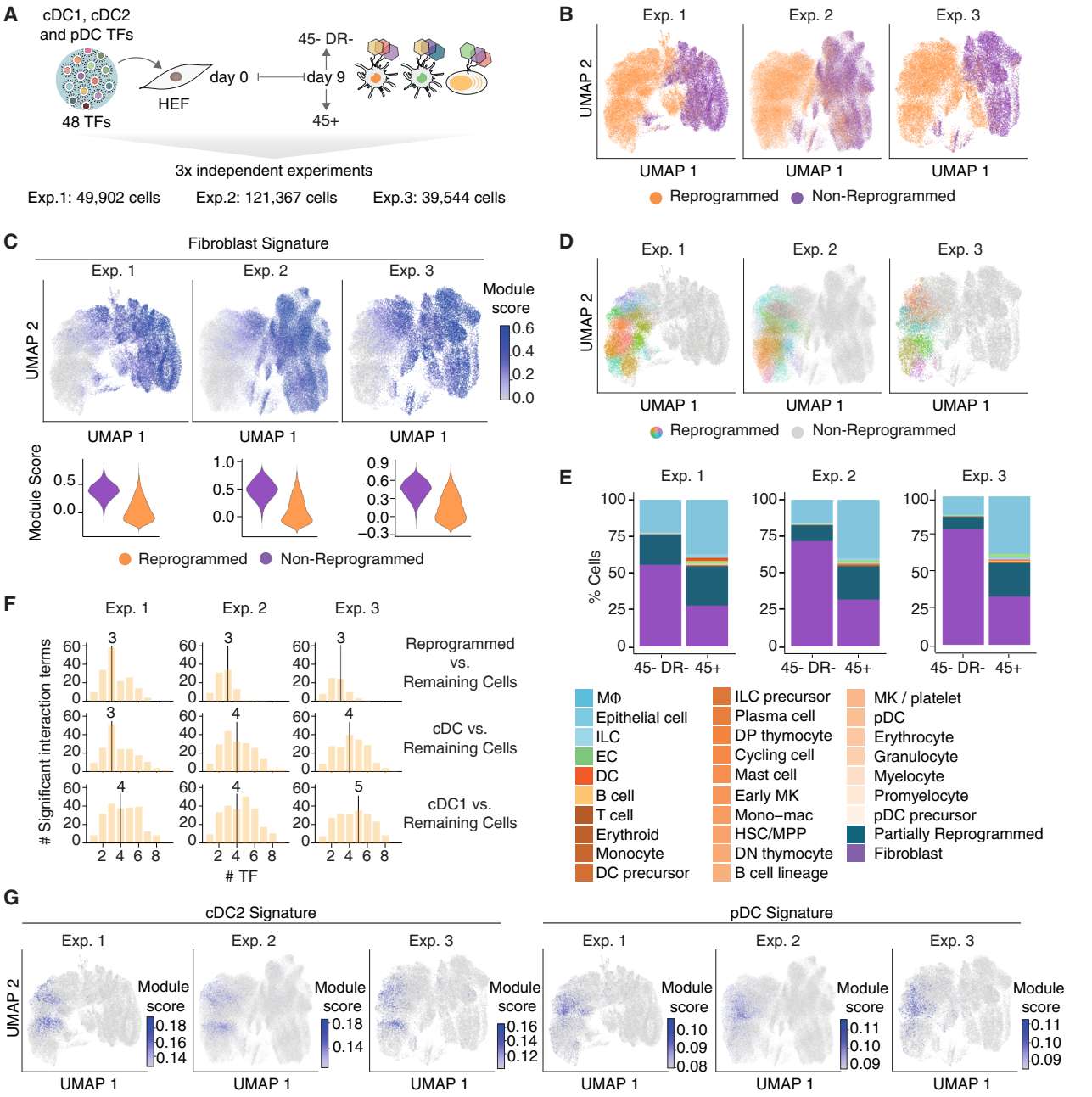

**Figure 4. Construction of an immune-restricted TF library**

(A) Schematic representation for the selection of immune TFs. A human TF library (1,829 TFs) was compiled by merging entries from the human TF database and TFome. Average expression levels of each TF across 177 cell types in the Tabula Sapiens scRNA-seq dataset[37] were calculated. Fold changes were computed between 73 hematopoietic/immune cell types and 104 non-immune cell types. TFs were ranked by immune cell type specificity and merged into a single immune-restricted TF list, which was then combined with 61 previously reported TFs involved in hematopoietic reprogramming.

(B) Bar plot showing average gene expression levels of selected TFs across immune and non-immune cell types from the Tabula Sapiens dataset.[37] Four immune TFs (*PU.1*, *BATF3*, *PAX5*, and *FOXP3*) and four non-immune TFs (*ELF3*, *EHF*, *ID4*, and *SOX9*) are displayed. The gray box highlights the immune lineage, where corresponding immune TFs are enriched. DCs. dendritic cells; $T_{regs}$, regulatory T cells.

(C) UMAP visualization of the Tabula Sapiens dataset.[37] Cells are colored by tissue origin and immune compartment (left). Module scores for non-immune (middle) and immune TFs (right) are shown.

(D) UMAP of a cross-tissue immune cell dataset,[15] with cells colored by immune lineage (left). Module scores for immune (middle) and non-immune TFs (right) are displayed. DC, dendritic cells; ILC, innate lymphoid cell; MΦ, macrophages; NK, natural killer cells; NKT, natural killer T cells; pDC, plasmacytoid dendritic cells; T reg, regulatory T cells; double-negative (DN) and double-positive (DP) thymocytes.




H

| | Exp. 1 | Exp. 2 | Exp. 3 | Combined |
|---|---|---|---|---|
| cDC2 | PU.1,KLF4,MYCL,RBPJ<br>PU.1,KLF4,MYCL,XBP1<br>PU.1,KLF4,ETS1,MYCL<br>PU.1,KLF4,XBP1,RBPJ<br>KLF4,MYCL,XBP1,RBPJ | KLF4,JUNB,XBP1,RBPJ<br>KLF4,BATF3,JUNB,RBPJ<br>PU.1,KLF4,JUNB,RBPJ<br>KLF4,BATF3,JUNB,XBP1<br>KLF4,BATF3,XBP1,RBPJ | PU.1,KLF4,BATF3,RBPJ<br>KLF4,ARID3A,BATF3,RBPJ<br>PU.1,KLF4,ARID3A,RBPJ<br>PU.1,KLF4,ARID3A,BATF3<br>KLF4,ARID3A,JUNB,RBPJ | PU.1,KLF4,XBP1,RBPJ<br>PU.1,KLF4,JUNB,RBPJ<br>PU.1,KLF4,BATF3,RBPJ<br>PU.1,KLF4,ARID3A,RBPJ<br>PU.1,KLF4,MYCL,RBPJ |
| pDC | IRF4,RBPJ,RUNX3,PU.1<br>ETS1,IRF4,RBPJ,RUNX3<br>ETS1,IRF4,RBPJ,PU.1<br>IRF4,XBP1,RBPJ,RUNX3<br>IRF4,XBP1,RUNX3,PU.1 | JUNB,XBP1,PU.1,ZBTB46<br>ARID3A,MYCL,XBP1,ZBTB46<br>ARID3A,JUNB,MYCL,ZBTB46<br>ARID3A,MYCL,RBPJ,ZBTB46<br>ARID3A,MYCL,PU.1,ZBTB46 | IRF8,ARID3A,JUNB,MXD1<br>IRF8,JUNB,MXD1,RBPJ<br>IRF8,JUNB,MXD1,PU.1<br>IRF8,ARID3A,MXD1,RBPJ<br>ARID3A,JUNB,MYCL,XBP1 | IRF8,ARID3A,RBPJ,PU.1<br>IRF8,ARID3A,XBP1,PU.1<br>IRF8,ARID3A,XBP1,RBPJ<br>IRF8,XBP1,RBPJ,PU.1<br>IRF8,ARID3A,JUNB,RBPJ |

*(legend on next page)*

incorporated TFs into the model based on their ability to explain transcriptomic variance. Across multiple immune lineages, we found that reprogramming typically required combinations of three to five TFs. While low-order combinations were sufficient to distinguish reprogrammed from non-reprogrammed cells, higher-order interactions were necessary to capture the finer distinctions between immune subtypes, such as cDC1 and other DC subsets (Figure 5F). These results were not driven by TF co-occurrence frequency (Figure S5F), underscoring the specificity of combinatorial effects rather than simple statistical association.

As a proof-of-concept, we asked whether the canonical PIB combination could be recovered from this highly heterogeneous dataset. Overlaying a cDC1 gene signature onto UMAP projections revealed cDC1-enriched clusters in all three experiments: cluster 17 in experiment 1, cluster 15 in experiment 2, and clusters 2, 7, 15, and 17 in experiment 3 (Figures S6A and S6B). Analysis of all observed 3- and 4-TF meta-combinations showed a strong enrichment of PIB in cDC1-annotated cells (Figure S6C). When integrating data across all three experiments, PIB consistently ranked among the top five meta-combinations, even in cases where it was less prominent in individual datasets (Figure S6D; Table S4). These findings underscore the reproducibility of our approach and highlight the value of integrating multiple replicates to prioritize functionally relevant TF modules.

We next applied this approach to identify reprogramming combinations for additional DC lineages, including cDC2 (DC3) and pDC (DC6) lineages.[31] Signature overlays on UMAPs highlighted cDC2- and pDC-like populations (Figure 5G). In cDC2-enriched clusters, the most frequent TF combinations included PU.1, KLF4, and RBPJ, all known regulators of DC development[41–43] (Figure 5H, top). For pDC-like cells, combinations involving IRF8, ARID3A,[44] and XBP1[45] were enriched, consistent with their established roles in pDC biology, such as type I interferon production and pDC maintenance[46] (Figure 5H, bottom). All lineage-specific TF combinations identified in this screen are cataloged in Table S4.

To benchmark our arrayed virus production approach against pooled delivery (as used in prior studies[21]), we performed parallel reprogramming experiments using the same 48-TF set produced either individually (arrayed) or as a pooled mixture. Barcode detection via both GEX and targeted amplification showed high concordance between the two formats ($r^2$ = 0.98; Figure S7A), confirming barcode fidelity. In the arrayed format, all 48 TFs were detected at day 2 (Figure S7B), whereas in the pooled condition, ZEB2 (3,645 bp) was undetectable, and TF representation was strongly biased against longer transgenes (Pearson r = −0.57, $p = 2.7 \times 10^{-5}$). This bias was absent in the arrayed condition (r = −0.11, $p$ = 0.45; Figure S7C). At the transcriptomic level, UMAPs revealed distinct separation between reprogrammed and non-reprogrammed cells in both formats (Figure S7D). However, recovery of known instructive combinations was markedly better in the arrayed condition. While cDC1-like cells were detected in both formats (Figure S7E), PIB ranked as the top 3- and 4-TF combination only in the arrayed setting (rank 1) and ranked poorly in the pooled format (ranks 217 and 234; Figure S7F; Table S4).

Together, these results demonstrate that REPROcode enables multiplexed induction of diverse immune lineages from fibroblasts in a single experiment. The platform not only recapitulates known lineage-specifying TF combinations but also reveals additional regulators of immune identity, providing an approach for dissecting the transcriptional logic of immune cell diversity.

## TF hierarchy to induce myeloid and lymphoid immune cell identities

To further investigate TF dynamics during immune cell reprogramming, we employed a complementary approach aimed at identifying the TFs most frequently expressed in reprogrammed cells. Among these, PU.1 was the most abundant, detected in 82.77% of $CD45^+$ cells, compared with just 1.46% in $CD45^-$HLA-$DR^-$ cells. *PU.1* expression strongly overlapped with *PTPRC* (CD45) (90.34%), consistent with its established role as a pioneer TF[47] and master regulator of myeloid lineage commitment[48] (Figure 6A). Within the *PTPRC*-expressing population, we observed divergent expression patterns of *HLA-DR* (55.68%) and *CD7* (26.86%), markers of antigen-presenting myeloid cells and lymphoid cells, respectively. These patterns suggest early lineage bifurcation and were consistently observed across two additional independent experiments (Figures S8A and S8B). To dissect the regulatory hierarchy underlying immune reprogramming, we constructed a decision tree model using TF expression data from the 48-TF library. This model, which integrated TF detection with transcriptomic profiles, achieved a Pearson correlation of 0.72 ($R^2$ = 0.03 and mean squared error

**Figure 5. Programming immune cell diversity**

(A) Experimental strategy to establish a diversity screen for inducing multiple DC cell fates. Human embryonic fibroblasts (HEFs) were transduced with 48 TFs enriched in cDC1s, cDC2s, and pDCs. Day 9 reprogrammed ($CD45^+$, $45^+$) and non-reprogrammed ($CD45^-$HLA-$DR^-$ and $45^-DR^-$) cells were FACS-sorted and analyzed by scRNA-seq in three independent experiments.

(B) UMAP visualization for FACS-purified non-reprogrammed and reprogrammed cells.

(C) Module score for fibroblast gene signature overlaid on UMAP visualization (top) and violin plots showing distribution in non-reprogrammed and reprogrammed cells (bottom).

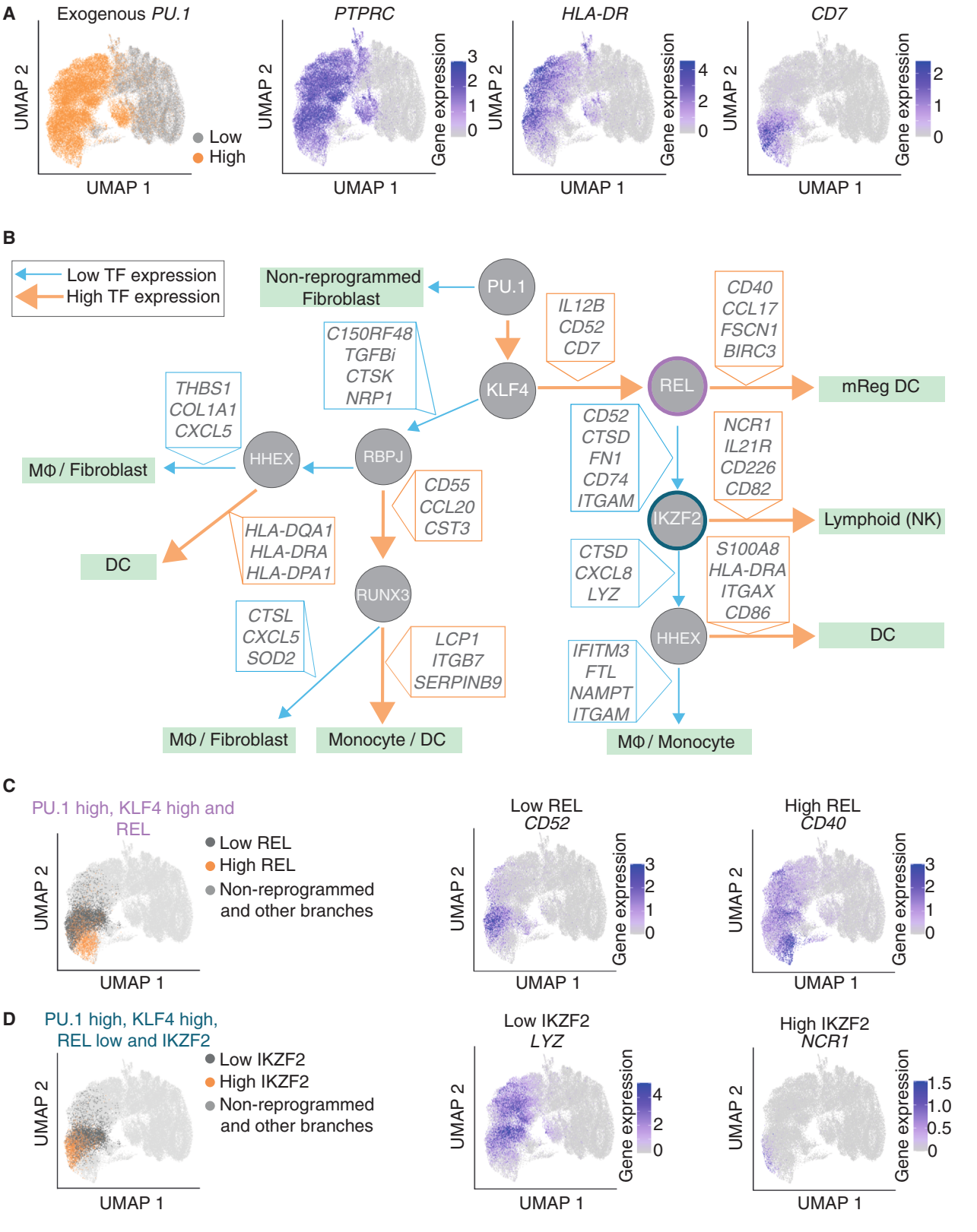
(D) UMAP visualization of reprogrammed cells ($CD45^+$) divided into Louvain clusters and non-reprogrammed cells (gray). Colors indicate distinct clusters.

(E) Cell types were annotated using CellTypist with a reference immune model. Ranked enriched immune cell types in reprogrammed cells are listed. Partially reprogrammed cells were defined as classified cells with a low confidence score. Innate lymphoid cell (ILC), hematopoietic stem cell (HSC)/multipotent progenitor (MPP), macrophage (MΦ), and double-negative (DN) thymocytes.

(F) Barplot showing TF distribution from the hierarchical stepwise regression model across three independent experiments using reprogrammed cells, cDC, and cDC1 gene signatures. The median in each condition is shown as a dashed line.

(G) UMAP visualization overlaid with module score for cDC2 (left) and pDC (right) gene signatures.

(H) Top 5 combinations of 4 TFs in cDC2-restricted cells (left) and pDC-restricted cells (right) are shown for individual experiments and for cumulative ranking across three experiments (Table S4).




(legend on next page)

[MSE] = 0.096; Figure 6B). The resulting hierarchy enabled identification of dominant TFs at key lineage-branching points and was validated across three independent experiments (Figures 6B, S8C, and S8D; Table S5). To assess robustness and reproducibility, we evaluated the model using a mixed held-out test set composed of cells from all three biological replicates (Figure S8E). Feature importance scores showed strong concordance across replicates (mean Spearman r = 0.76), and models demonstrated substantial generalizability across samples, supporting the reliability of the inferred regulatory logic, despite the limited predictive strength of the decision tree model (Figures S8E and S8F). As expected, high *PU.1* expression stratified hematopoietic ($CD45^+$) from non-reprogrammed ($CD45^-HLA\text{-}DR^-$) or fibroblast-like cells at the root of the decision tree. Within the $PU.1^{high}$ population, branching occurred based on KLF4 expression. $KLF4^{high}$ cells were enriched for *IL12B*, *CD7*, and *ITGAX*, while $KLF4^{low}$ cells expressed *NRP1*, *CTSK*, and *TGFBI* (Figures 6B and S9A; Table S5), indicating the emergence of divergent transcriptional programs. This pattern of PU.1 followed by KLF4 dominance was consistently observed in all three experiments (Figures 6B, S8C, and S8D; Table S5).

Within the $KLF4^{high}$ lineage, high *REL* expression marked a branch associated with mature regulatory dendritic cells (mRegDCs), characterized by activation markers *CD40*, *CD80*, *CD86*, and *CD83*, along with *SOX2* and *FAS* (Figure 6C; Table S5).[50] Conversely, the low *REL* branch was influenced by *IKZF2*, which directed cells toward a lymphoid fate. High *IKZF2* expression correlated with *GZMB*, *NCR1*, and *IL21R*, markers of NK cells (Figure 6D; Table S5), whereas low *IKZF2* expression redirected cells toward myeloid lineages, such as macrophages and monocytes. In this $IKZF2^{low}$ compartment, high *HHEX* expression was associated with DC markers such as *S100A8*, *HLA-DRA*, and *ITGAX*, suggesting a role for HHEX in directing DC differentiation (Figure S9B; Table S5).

In the $PU.1^{high}$-$KLF4^{low}$ branch, RBPJ guided further bifurcations into populations with non-reprogrammed/fibroblast or myeloid gene signatures encompassing macrophages and DCs (Figure S9C; Table S5). Notably, within the $RBPJ^{low}$ subbranch, cells expressing *HHEX* (alongside high PU.1 and low KLF4) co-localized with cDC1 signature-positive clusters (Figure S9D; Table S5), implicating HHEX as a potential cDC1 enhancer. In the $RBPJ^{high}$ arm, *RUNX3* expression once again delineated DC, monocyte, and macrophage transcriptional programs (Figure S9E; Table S5). While the downstream ordering of TFs varied across experiments, the resulting cell identities were consistent (Figures S8C and S8D; Table S5), suggesting that multiple TF trajectories can converge on similar immune phenotypes.

Taken together, these results demonstrate that our 48-TF screen, originally designed to target cDC1, cDC2, and pDC subsets, successfully induced a broader array of immune phenotypes, including NK-like lymphoid cells. The decision tree analysis offers a framework to explore TF hierarchies that govern lineage bifurcations, revealing how combinations of TFs interact at specific branch points to shape distinct immune cell identities.

## Induction of NK-like cells

To demonstrate REPROcode's versatility beyond dendritic and myeloid reprogramming, we applied the platform to program NK cells, a cytotoxic lymphoid lineage with major relevance for immunotherapy. To identify TF combinations capable of inducing an NK cell identity, we integrated hits from our REPROcode-derived combination analysis and decision tree framework with NK-specific TFs described in the literature. This approach led to the inclusion of EOMES, a canonical NK lineage regulator, in a set of eight TFs identified by REPROcode. Four of these TFs, TBX21, ETS1, NFIL3, and EOMES, collectively termed TENE, were enriched in both $CD56^+CD16^-$ and $CD56^{dim}CD16^+$ NK subsets (Figure S10A).

We evaluated three TF combinations for NK reprogramming: (1) the original REPROcode-derived 8-TF set, (2) the expanded 9-TF set including EOMES, and (3) the TENE 4-TF module. Flow cytometry at days 9 and 12 revealed that TENE induced the highest expression of CD56, a canonical NK surface marker (Figure S10B; Table S6), and was selected for further analysis (Figure 7A). CD56 upregulation was observed as early as day 3 post-transduction, while expression of lineage-exclusion markers CD3 (T cells) and CD19 (B cells) remained absent (Figures 7B, 7C, and S10C; Table S6), confirming specific reprogramming toward the NK lineage. TENE-transduced fibroblasts exhibited lymphoid-like morphology and robust NKG7 expression (82.8%), a protein essential for granule exocytosis (Figure 7D). Transmission electron microscopy further confirmed the presence of intracellular granules dispersed throughout the cytoplasm, consistent with cytotoxic NK cell ultrastructure (Figures 7E and S10D; Table S6). Addition of cytokines (stem cell factor, interleukin [IL]-3, FLT3L, IL-7, and IL-15) did not improve reprogramming efficiency by day 6 (Figure S10E; Table S6), suggesting that the enforced TF network is sufficient to drive NK identity without exogenous cytokine support.

Transcriptomic profiling showed that TENE-induced $CD56^+$ cells clustered closely with primary NK cells (Figure 7F). Differential gene expression across HEFs, TENE, and NK subsets

**Figure 6. TF hierarchy to induce myeloid and lymphoid immune cell identities**

(A) UMAP visualization of human embryonic fibroblasts (HEFs) transduced with a pool of 48 TFs (10 cp per cell) enriched in cDC1, cDC2, or pDC lineages. Cells were profiled using scRNA-seq via BD Rhapsody. In the left, high (orange) and low (gray) expression of *SPI1* is highlighted. Expression levels of *PTPRC*, *HLA-DR*, and *CD7* are overlaid on the UMAP (right).

(B) TF hierarchy of reprogramming-based cell fate decisions modelled using DecisionTreeRegressor.[49] Branching points were defined using differential expression gene analysis and are illustrated by high (orange arrow) and low (blue arrow) TF expression. Exemplative genes are shown at each branching point, and induced cell fates based on gene expression analysis are highlighted in green boxes. REL (purple circle) and IKZF2 (blue circle) branching points are highlighted and shown in (C) and (D). See Figure S9 for other branching points (Table S5). mReg, mature regulatory; NK, lymphoid natural killer cells, DCs, dendritic cells; MΦ, macrophages.

(C and D) (C) At the PU.1 high, KLF4 high, REL branching point, or (D) at the PU.1 high, KLF4 high, REL low, IKZF2 branching point, cells with high (orange) and low (dark gray) REL or IKZF2 expression are highlighted, compared with non-reprogrammed cells and cells from other branches (light gray; left). Expression levels of exemplificative genes for low and high REL and IKZF2 branches, including *CD52*, *CD40*, *LYZ*, and *NCR1*, are shown.

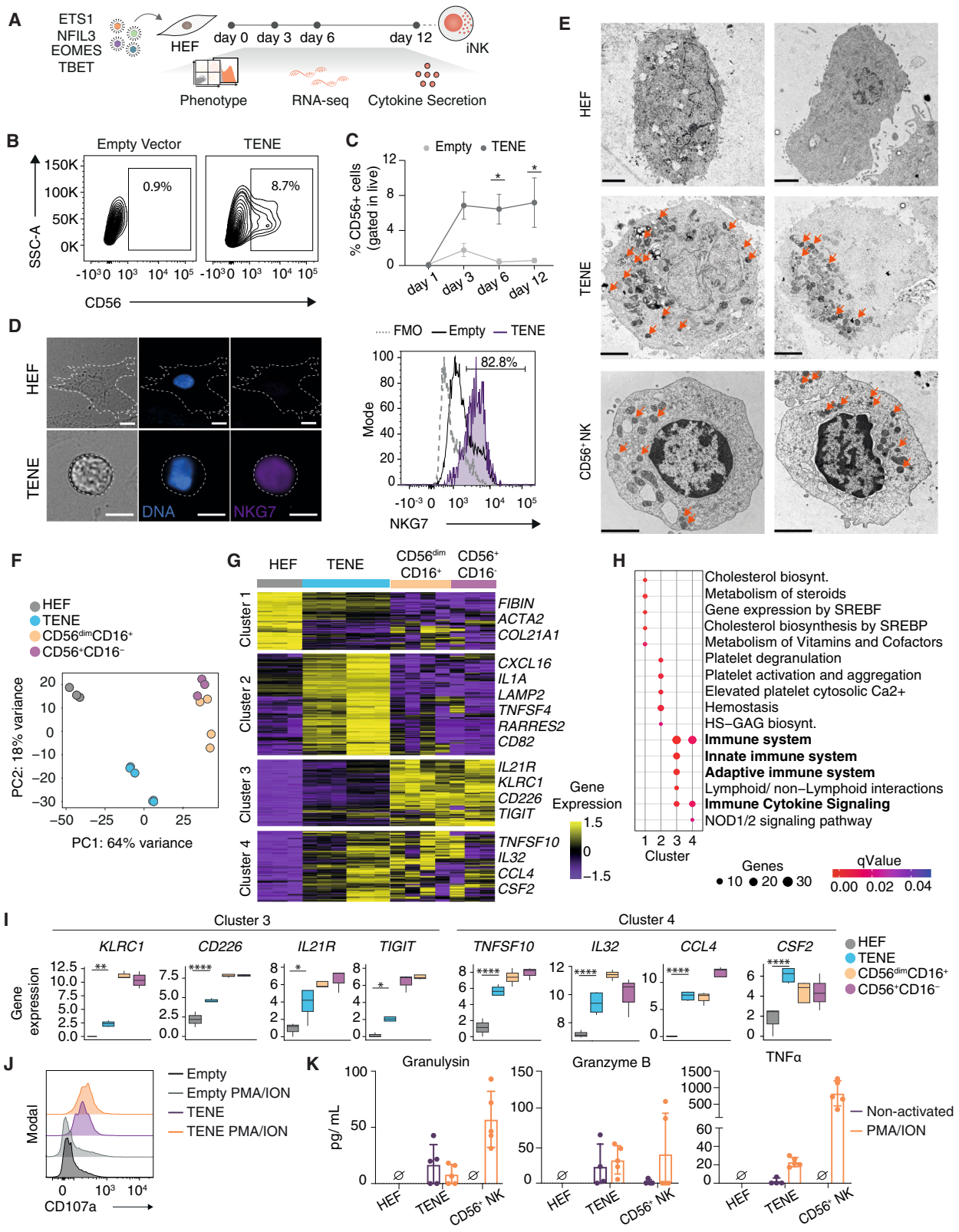



**Figure 7. Induction of NK cell identity and function in human fibroblasts**

(A) Experimental strategy to induce NK cells from human embryonic fibroblasts (HEFs) with TFs. RNA-seq was used to profile reprogrammed cells at day 12. Induced cells were characterized phenotypically and functionally.

*(legend continued on next page)*

identified four gene clusters (Table S6): (1) fibroblast-associated genes (*FIBIN*, *ACTA2*, and *COL21A1*), (2) genes upregulated in TENE cells (*LAMP2*, *CXCL16*, and *CSF2*), (3) NK-specific genes, and (4) genes shared between TENE and NK cells. Cluster 4 included *TNFSF10*, *IL32*, and *CCL4*, key effectors of cytotoxicity and inflammation, suggesting that TENE reprogramming induces a proinflammatory, cytotoxic transcriptional state. Reactome pathway enrichment analysis further supported this profile, showing overrepresentation of the immune system ($p = 4.3 \times 10^{-8}$ cluster 3; $p = 1.4 \times 10^{-2}$ cluster 4) and cytokine signaling pathways ($p = 3.1 \times 10^{-3}$ cluster 3; $p = 1.4 \times 10^{-2}$ cluster 4; Figures 7G–7I and S10F; Table S6). Functionally, TENE-induced NK (iNK)-like cells exhibited hallmark cytotoxic responses. CD107a expression, a marker of degranulation, was 1.8-fold higher in TENE-transduced cells following phorbol 12-myristate 13-acetate (PMA)/ionomycin stimulation (MFI: 296 ± 5 vs. 168 ± 11 for empty vector; Figure 7J). These iNK cells also spontaneously secreted granulysin and granzyme B and released tumor necrosis factor alpha (TNF-α) upon stimulation (Figure 7K; Table S6), demonstrating both granule-mediated and receptor-mediated cytotoxic potential.

Together, these findings establish TENE combination as a driver of NK-like reprogramming in human fibroblasts. The resulting iNK cells exhibit lineage specificity, effector function, and transcriptional similarity to natural NK cells. These results expand the utility of REPROcode beyond DC programming, illustrating its power to uncover testable, lineage-defining TF combinations across diverse immune cell types.

## DISCUSSION

In this study, we present REPROcode, a combinatorial, single-cell screening platform that identifies TF combinations capable of reprogramming human somatic cells into diverse immune lineages. By leveraging a curated library of 408 immune-restricted TFs and combining combinatorial barcoding with scRNA-seq, REPROcode enables high-throughput interrogation of TF pools (~50 TFs per screen). This approach allows for the simultaneous evaluation of reprogramming efficiency, TF stoichiometry, and downstream cell state specification, delivering both breadth and precision in identifying instructive TF combinations.

REPROcode advances current reprogramming strategies in several ways. First, it enables combinatorial screening directly in human cells, avoiding challenges due to species-specific differences and eliminating the need for transgenic reporter systems.[7,16,51,52] Second, it complements computational TF prediction tools[17,18,26] by functionally validating candidate combinations in an experimental framework. Third, the use of arrayed viral production avoids size-related biases common in pooled lentiviral methods,[21] ensuring equitable representation of large TFs. Fourth, single-cell transcriptomics resolves heterogeneous cell states and intermediate fates, overcoming limitations of bulk overexpression screens.[7,16] REPROcode allows for the detection of combinatorial effects at low MOI, which proved essential for resolving the canonical cDC1 reprogramming factors PIB in large screens. We also found that the BD Rhapsody system provided superior detection sensitivity and throughput compared with droplet-based platforms, further enhancing the resolution of complex TF interactions.

REPROcode builds on and extends previous single-cell platforms such as TF Atlas,[21] scTF-seq,[28] and Collide-seq,[23] which primarily focused on pairwise TF interactions or dosage effects. By contrast, REPROcode is designed to resolve higher-order combinations (3–5 TFs), capturing the cooperative and nonlinear interactions that often underlie successful reprogramming. While CRISPR-based loss-of-function tools (e.g., CROP-seq and Perturb-seq[24,53,54]) and CRISPRa[55–58] platforms are powerful for probing genetic interactions, they are less effective at inducing the high expression levels typically required for cell fate conversion. REPROcode overcomes this by directly overexpressing TF combinations at functional levels, enabling robust and efficient reprogramming across immune lineages.

Mechanistically, we showed that high levels of PU.1 and BATF3, with intermediate levels of IRF8, are critical for inducing cDC1 fate. PU.1 promotes cDC development while repressing pDC differentiation,[59] and BATF3 supports IRF8 stability via the -+32 kb enhancer, driving cDC1-specific gene programs important for cross-presentation and tumor immunity.[60,61] While IRF8 is essential for cDC1 identity,[62] its overexpression may hinder reprogramming efficiency, possibly due to inflammation-related cell stress.

We also identified TFs modulating cDC1 functional states. NFKBIA and NFKBIB promoted immature phenotypes,

(B) Representative flow cytometry plots for CD56 expression in HEFs and quantification of CD56 expression 12 days after transduction with lentiviruses encoding the 4 TFs (TENE). Lentiviruses for empty vector (SFFV-MCS) were used as a control.
(C) Kinetics of CD56 surface expression in TENE-transduced HEFs and empty vector control, 0, 3, 6, and 12 days after transduction. $n$ = 3–5; mean ± SD; Table S6. $p$ values were calculated by two-way ANOVA and the multiple Mann-Whitney comparison test. $n$ indicates independent experiments. $^{*}p < 0.05$.
(D) Immunofluorescence microscopy for HEFs and FACS-purified CD56$^{+}$ induced NK cells (iNK cells; TENE) at day 12. Staining for NKG7 (purple) and Hoechst (blue). Scale bars, 50 μM. Flow cytometry histogram for NKG7 intracellular staining at day 12. MCS is gated in live, and TENE is gated in CD56$^{+}$ cells ($n$ = 6).
(E) Transmission electron microscopy for HEFs, FACS-purified CD56$^{+}$ iNK cells (TENE), and peripheral blood CD56$^{+}$ NK cells. Arrows indicate granules defined by small-sized, electron-dense particles and the absence of mitochondrial cristae. Scale bar, 2 μM.
(F) Principal-component analysis (PCA) representation of reprogrammed and natural cells. Each dot represents a technical replicate.
(G) Heatmap showing differentially expressed genes in HEFs, NK cells (TENE), and natural peripheral blood NK cells. Genes are divided into 4 clusters (Table S6).
(H) Reactome pathway enrichment analysis of corresponding gene clusters (Table S6).
(I) Boxplot showing expression levels of relevant genes from clusters 3 and 4. $p$ values were calculated by the Wald test with Benjamini Hochberg (BH) adjustment. Relevant statistical comparisons between HEFs and iNK cells are shown. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{****}p < 0.0001$.
(J) Flow cytometry histogram for CD107a expression in HEFs transduced with empty vector (MCS) and the 4 TFs (TENE), with or without PMA/ionomycin (ION) activation. MCS conditions were gated in live cells, and TENE conditions were gated in CD56$^{+}$ cells.
(K) Secretion of granulysin, granzyme B, and TNF-α by CD56$^{+}$ iNK cells (TENE) with or without PMA/ION activation. HEFs transduced with empty vector and CD3$^{-}$CD56$^{+}$ peripheral blood NK cells were used as references. $n$ = 4–5; mean ± SD; Table S6.

consistent with their roles in dampening DC activation.[63] MXD1 and ID2 enhanced immunostimulatory programs and cDC1 maturation, aligning with their known roles: MXD1 supports cDC1 fitness and maturation,[34] while ID2 promotes cDC1 lineage commitment by antagonizing E protein activity.[64] Specifically, ID2 regulates enhancers in *Irf8*, maintaining IRF8 levels critical for cDC1 identity[65] while repressing TCF4 and therefore pDC lineages.[66] A recent study further supports this role, showing that a combination of ID2 and PIB effectively induced a mature cDC1-like phenotype in glioblastoma cells, restoring their antigen-presenting function.[12] Likewise, BCL6 and IRF7 drove migratory CCR7$^+$ cDC states, consistent with *in vivo* studies showing reduced XCR1$^+$ migratory cDCs in Bcl6-deficient mice.[35] These findings mirror known *in vivo* roles of these TFs and validate REPROcode's ability to resolve not only lineage identity but also cell state dynamics.

Beyond cDC1s, REPROcode uncovered TF combinations for programming cDC2, pDCs, macrophage, monocyte, and NK-like lineages. For pDCs, IRF8, ARID3A, and XBP1 were key regulators, aligning with their known roles in pDC specification and function.[41,44,45] cDC2 specification involved a division of labor between KLF4, linked to cDC2Bs,[42,43] and RBPJ, associated with Notch-dependent cDC2As[67] and monocyte-to-macrophage transitions.[68] These results highlight the heterogeneity of myeloid subsets and demonstrate REPROcode's resolution in dissecting lineage-specific TF modules.

Using decision tree analysis, we mapped a hierarchical TF architecture underlying immune reprogramming. At the apex, PU.1 acted as a gatekeeper for hematopoietic identity,[48] while KLF4 served as a critical branching node influencing DC subtype specification.[42,43] High KLF4 expression was associated with the development of mRegDCs, particularly when co-expressed with REL, BCL6, or IRF7. By contrast, low KLF4 expression favored cDC1 development, especially in the presence of RBPJ and HHEX. These findings align with reports linking mRegDCs to tumor immunity[50] and suggest a distinct regulatory pathway for DC specification mediated by these TF combinations.

REPROcode also enabled the generation of NK-like cells via enforced expression of TENE. iNK cells expressed canonical NK markers (CD56, NKG7, NCR1, and GZMB) and showed effector functions such as degranulation and cytokine release. ETS1, NFIL3, and TBX21 are well-established regulators of NK differentiation and function,[69–72] while EOMES promotes effector gene expression.[73] The addition of hematopoietic cytokines (e.g., IL-15 and FLT3L) did not enhance reprogramming, suggesting that the TF program alone is sufficient to initiate NK cell identity *in vitro*, echoing previous findings on FLT3L independence of induced DCs.[7,8] However, future *in vivo* studies will be essential to assess how tissue microenvironments, particularly in lymphoid organs like the spleen, might influence the maturation, survival, and function of reprogrammed immune cells.

To enhance discovery of regulators of immune identity, we integrated three computational strategies: (1) TF hierarchy inference, (2) lineage-specific TF restriction mapping, and (3) ranking critical TFs for targeted immune fates. While hierarchy analysis captures dominant reprogramming drivers, lineage-restricted enrichment improves precision for programming rare or specialized immune subtypes.

Looking ahead, REPROcode offers broad utility in regenerative medicine and immunotherapy. For example, screens targeting tolerogenic DCs could identify TF combinations that promote immune suppression in autoimmunity.[74] TF pools tailored to other lineages (~50 TFs) could also enable reprogramming into lymphoid subsets such as regulatory or γδ T cells. One limitation of the current approach is the inability to precisely control TF dosage per cell, which depends on estimated MOI. Future iterations could incorporate methods for TF copy-number titration or real-time selection of desired TF expression profiles via fluorescent reporters or RNA-based sensors.[75] Although our screen involved up to 48 TFs, the platform's full capacity remains untested. Scaling to larger libraries of 100–500 TFs would unlock deeper insights into high-dimensional TF regulatory networks.

In summary, REPROcode represents a powerful functional genomics framework for decoding the transcriptional logic of immune cell identity. By enabling systematic, high-throughput interrogation of TF combinations with single-cell resolution, REPROcode provides opportunities to uncover lineage-specifying programs and to engineer immune cells for therapeutic applications in cancer, autoimmunity, and tissue repair.

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Carlos-Filipe Pereira (filipe.pereira@med.lu.se).

### Materials availability

Plasmids generated in this study have been deposited to Addgene. Material transfer agreements (MTAs) will be necessary to obtain human polycistronic reprogramming constructs.

### Data and code availability

- ScRNA-seq and bulk RNA sequencing (RNA-seq) data have been deposited at GEO and are publicly available from the date of publication. Accession numbers are listed in the key resources table.
- This paper analyzes existing, publicly available data. The accession numbers for these datasets are listed in the key resources table.
- All original code has been deposited on Zenodo. DOI is listed in the key resources table.
- A web-based application for processed scRNA-seq data is available at https://cellreprolab.shinyapps.io/immuneReproTF/.
- Additional information required to reanalyze the data reported here is available from the lead contact upon request.

## ACKNOWLEDGMENTS

We thank the members of the Cell Reprogramming in Haematopoiesis and Immunity Laboratory for useful discussions. We thank the Center for Translational Genomics and Clinical Genomics Lund, SciLifeLab, for providing sequencing services. We would also like to thank the Lund Stem Cell Center FACS facility for cell sorting assistance. This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (866448—TrojanDC and 101189370—DART). This project was also funded by the European Innovation Council (101130218-RESYNC), Cancerfonden (23 2932 Pj), the Swedish Research Council (2020-00615), NovoNordisk Fonden (NNF22C0079466), FCT (2022.02338.PTDC), and Plano de Recuperação e Resiliência de Portugal pelo fundo NextGenerationEU (grant C644865576-00000005). The Knut and Alice Wallenberg Foundation, the Medical Faculty at Lund University, and Region Skåne are acknowledged for financial support. I.K. is supported by a Cancer Research Institute Immuno-Informatics Postdoctoral Fellowship (CRI5008)

and the Royal Physiographic Society in Lund (42846 and 45488). L.H.-O., I.C., and D.P.-C. are supported by FCT doctoral scholarships (10.54499/2020.06447.BD, PD/BD/146424/2019, and 10.54499/2022.15306.BD). B.C. was supported by the DKFZ Postdoc Internship Program and the German Helmholtz Impulse and Networking Fund. The computational resources were provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

## AUTHOR CONTRIBUTIONS

I.K. and C.-F.P. were responsible for conceptualization and experimental design. D.P.-C., M.N., D.O., and C.F. cloned the TF library. O.Z., A.R.A., and L.H.-O. FACS-sorted cells for scRNA-seq experiments. A.R.A. and I.C. generated a library for scRNA-seq. I.C., L.R.C., A.R.A., O.Z., E.H., D.P.-C., R.K., B.C., C.F.P., and F.F.R. conceptualized and performed validation experiments. I.K., M. Minaeva, D.K., M.R., and F.J.T. analyzed and interpreted scRNA-seq data. I.K., M. Martin developed barcode detection software. E.S., F.J.T., and C.-F.P. supervised. I.K., A.R.A., I.C., E.H., and C.-F.P. wrote the manuscript. All authors contributed to data interpretation and manuscript revision.

## DECLARATION OF INTERESTS

C.F.P., F.F.R., and C.-F.P. have equity interests in in and serve in management positions at Asgard Therapeutics AB, which develops cancer immunotherapies based on DC reprogramming technologies. C.F.P., F.F.R., and C.-F.P. are inventors on granted patents U.S. 11,345,891, JP 7303743, CN ZL201880005047.3, and patent application WO 2018/185709 held by Asgard Therapeutics that cover the cell reprogramming approach described here.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines and primary cell cultures
- METHOD DETAILS
  - Immune-restricted TF identification
  - Molecular cloning
  - Pooled lentiviral production
  - Lentiviral production and titration
  - Viral transduction and reprogramming
  - Flow cytometry analysis
  - Fluorescence-activated cell sorting (FACS)
  - cDC1 reprogramming with additional regulators
  - Single-cell RNA sequencing (scRNA-seq)
  - scRNA-seq data processing
  - TF barcode demultiplexing
  - Reprogramming trajectory analysis
  - scRNA-seq data analysis to evaluate the impact of additional TFs
  - Identification of candidate reprogramming TFs in multiplexed screens with >40 TFs
  - Cell type annotation with pre-trained models
  - Decision tree modeling
  - Hierarchical stepwise regression model
  - Microscopy
  - Isolation of NK populations from human peripheral blood
  - Bulk RNA sequencing
  - Degranulation and cytokine secretion
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2025.101457.

Received: November 4, 2024
Revised: July 2, 2025
Accepted: November 7, 2025
Published: January 14, 2026

## REFERENCES

1. Zimmermannova, O., Caiado, I., Ferreira, A.G., and Pereira, C.-F. (2021). Cell Fate Reprogramming in the Era of Cancer Immunotherapy. Front. Immunol. *12*, 714822. https://doi.org/10.3389/fimmu.2021.714822.
2. Ribas, A., and Wolchok, J.D. (2018). Cancer immunotherapy using checkpoint blockade. Science *359*, 1350–1355. https://doi.org/10.1126/science.aar4060.
3. Yin, X., Chen, S., and Eisenbarth, S.C. (2021). Dendritic Cell Regulation of T Helper Cells. Annu. Rev. Immunol. *39*, 759–790. https://doi.org/10.1146/annurev-immunol-101819-025146.
4. Cabeza-Cabrerizo, M., Cardoso, A., Minutti, C.M., Pereira da Costa, M., and Reis e Sousa, C. (2021). Dendritic Cells Revisited. Annu. Rev. Immunol. *39*, 131–166. https://doi.org/10.1146/annurev-immunol-061020-053707.
5. Alsinet, C., Primo, M.N., Lorenzi, V., Bello, E., Kelava, I., Jones, C.P., Vilarrasa-Blasi, R., Sancho-Serra, C., Knights, A.J., Park, J.-E., et al. (2022). Robust temporal map of human in vitro myelopoiesis using single-cell genomics. Nat. Commun. *13*, 2885. https://doi.org/10.1038/s41467-022-30557-4.
6. Wang, H., Yang, Y., Liu, J., and Qian, L. (2021). Direct cell reprogramming: approaches, mechanisms and progress. Nat. Rev. Mol. Cell Biol. *22*, 410–424. https://doi.org/10.1038/s41580-021-00335-z.
7. Rosa, F.F., Pires, C.F., Kurochkin, I., Ferreira, A.G., Gomes, A.M., Palma, L.G., Shaiv, K., Solanas, L., Azenha, C., Papatsenko, D., et al. (2018). Direct reprogramming of fibroblasts into antigen-presenting dendritic cells. Sci. Immunol. *3*, eaau4292. https://doi.org/10.1126/sciimmunol.aau4292.
8. Rosa, F.F., Pires, C.F., Kurochkin, I., Halitzki, E., Zahan, T., Arh, N., Zimmermannová, O., Ferreira, A.G., Li, H., Karlsson, S., et al. (2022). Single-cell transcriptional profiling informs efficient reprogramming of human somatic cells to cross-presenting dendritic cells. Sci. Immunol. *7*, eabg5539. https://doi.org/10.1126/sciimmunol.abg5539.
9. Zimmermannova, O., Ferreira, A.G., Ascic, E., Velasco Santiago, M., Kurochkin, I., Hansen, M., Met, Ö., Caiado, I., Shapiro, I.E., Michaux, J., et al. (2023). Restoring tumor immunogenicity with dendritic cell reprogramming. Sci. Immunol. *8*, eadd4817. https://doi.org/10.1126/sciimmunol.add4817.
10. Ascic, E., Åkerström, F., Sreekumar Nair, M., Rosa, A., Kurochkin, I., Zimmermannova, O., Catena, X., Rotankova, N., Veser, C., Rudnik, M., et al. (2024). In vivo dendritic cell reprogramming for cancer immunotherapy. Science *386*, eadn9083. https://doi.org/10.1126/science.adn9083.
11. Linde, M.H., Fan, A.C., Köhnke, T., Trotman-Grant, A.C., Gurev, S.F., Phan, P., Zhao, F., Haddock, N.L., Nuno, K.A., Gars, E.J., et al. (2023). Reprogramming Cancer into Antigen-Presenting Cells as a Novel Immunotherapy. Cancer Discov. *13*, 1164–1185. https://doi.org/10.1158/2159-8290.CD-21-0502.
12. Liu, T., Jin, D., Le, S.B., Chen, D., Sebastian, M., Riva, A., Liu, R., and Tran, D.D. (2024). Machine Learning-Directed Conversion of Glioblastoma Cells to Dendritic Cell-Like Antigen-Presenting Cells as Cancer Immunotherapy. Cancer Immunol. Res. *12*, 1340–1360. https://doi.org/10.1158/2326-6066.CIR-23-0721.
13. Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU.1 and C/EBPα/β convert fibroblasts into macrophage-like cells. Proc. Natl. Acad. Sci. USA *105*, 6057–6062. https://doi.org/10.1073/pnas.0711961105.
14. Henriques-Oliveira, L., Altman, A.R., Kurochkin, I., Ascic, E., Halitzki, E., Matei, A.-M., Pértiga-Cabral, D., Ulmert, I., Holst, S., and Nair, M.S. (2025). Anchored screening identifies transcription factor blueprints underlying dendritic cell diversity and subset-specific anti-tumor immunity.

Immunity *58*, 2419–2438.e13. https://doi.org/10.1016/j.immuni.2025.08.001.

15. Domínguez Conde, C., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., et al. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science *376*, eabl5197. https://doi.org/10.1126/science.abl5197.

16. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676. https://doi.org/10.1016/j.cell.2006.07.024.

17. Kamaraj, U.S., Chen, J., Katwadi, K., Ouyang, J.F., Yang Sun, Y.B., Lim, Y.M., Liu, X., Handoko, L., Polo, J.M., Petretto, E., et al. (2020). EpiMogrify Models H3K4me3 Data to Identify Signaling Molecules that Improve Cell Fate Control and Maintenance. Cell Syst. *11*, 509–522.e10. https://doi.org/10.1016/j.cels.2020.09.004.

18. Rackham, O.J.L., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., , et al.; FANTOM Consortium, Suzuki, H., Nefzger, C.M., Daub, C.O. (2016). A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. *48*, 331–335. https://doi.org/10.1038/ng.3487.

19. Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell *158*, 889–902. https://doi.org/10.1016/j.cell.2014.07.021.

20. Jung, S., Appleton, E., Ali, M., Church, G.M., and del Sol, A. (2021). A computer-guided design tool to increase the efficiency of cellular conversions. Nat. Commun. *12*, 1659. https://doi.org/10.1038/s41467-021-21801-4.

21. Joung, J., Ma, S., Tay, T., Geiger-Schuller, K.R., Kirchgatterer, P.C., Verdine, V.K., Guo, B., Arias-Garcia, M.A., Allen, W.E., Singh, A., et al. (2023). A transcription factor atlas of directed differentiation. Cell *186*, 209–229.e26. https://doi.org/10.1016/j.cell.2022.11.026.

22. Parekh, U., Wu, Y., Zhao, D., Worlikar, A., Shah, N., Zhang, K., and Mali, P. (2018). Mapping Cellular Reprogramming via Pooled Overexpression Screens with Paired Fitness and Single-Cell RNA-Sequencing Readout. Cell Syst. *7*, 548–555.e8. https://doi.org/10.1016/j.cels.2018.10.008.

23. Hersbach, B.A., Fischer, D.S., Masserdotti, G., Deeksha, Mojžišová, K., Waltzhöni, T., Rodriguez-Terrones, D., Heinig, M., Theis, F.J., Götz, M., et al. (2022). Probing cell identity hierarchies by fate titration and collision during direct reprogramming. Mol. Syst. Biol. *18*, e11129. https://doi.org/10.15252/msb.202211129.

24. Duan, J., Li, B., Bhakta, M., Xie, S., Zhou, P., Munshi, N.V., and Hon, G.C. (2019). Rational Reprogramming of Cellular States by Combinatorial Perturbation. Cell Rep. *27*, 3486–3499.e6. https://doi.org/10.1016/j.celrep.2019.05.079.

25. Pereira, C.-F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C.E., Clark, N.R., Nomura-Kitabayashi, A., Kovacic, J.C., Ma'ayan, A., et al. (2013). Induction of a Hemogenic Program in Mouse Fibroblasts. Cell Stem Cell *13*, 205–218. https://doi.org/10.1016/j.stem.2013.05.024.

26. D'Alessio, A.C., Fan, Z.P., Wert, K.J., Baranov, P., Cohen, M.A., Saini, J.S., Cohick, E., Charniga, C., Dadon, D., Hannett, N.M., et al. (2015). A Systematic Approach to Identify Candidate Transcription Factors that Control Cell Identity. Stem Cell Rep. *5*, 763–775. https://doi.org/10.1016/j.stemcr.2015.09.016.

27. Ng, A.H.M., Khoshakhlagh, P., Rojo Arias, J.E., Pasquini, G., Wang, K., Swiersy, A., Shipman, S.L., Appleton, E., Kiaee, K., Kohman, R.E., et al. (2021). A comprehensive library of human transcription factors for cell fate engineering. Nat. Biotechnol. *39*, 510–519. https://doi.org/10.1038/s41587-020-0742-6.

28. Liu, W., Saelens, W., Rainer, P., Bročanin, M., Gardeux, V., Gralak, A., van Mierlo, G., Russeil, J., Liu, T., Chen, W., and Deplancke, B. (2025). Dissecting reprogramming heterogeneity at single-cell resolution using scTF-seq. Nat. Genet. *57*, 2522–2535. https://doi.org/10.1038/s41588-025-02343-7.

29. Pereira, C.-F., Lemischka, I.R., and Moore, K. (2012). Reprogramming cell fates: insights from combinatorial approaches. Ann. N. Y. Acad. Sci. *1266*, 7–17. https://doi.org/10.1111/j.1749-6632.2012.06508.x.

30. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

31. Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573. https://doi.org/10.1126/science.aah4573.

32. Meiser, P., Knolle, M.A., Hirschberger, A., de Almeida, G.P., Bayerl, F., Lacher, S., Pedde, A.-M., Flommersfeld, S., Hönninger, J., Stark, L., et al. (2023). A distinct stimulatory cDC1 subpopulation amplifies CD8+ T cell responses in tumors for protective anti-cancer immunity. Cancer Cell *41*, 1498–1515.e10. https://doi.org/10.1016/j.ccell.2023.06.008.

33. Moore, F., Buonocore, S., Aksoy, E., Ouled-Haddou, N., Goriely, S., Lazarova, E., Paulart, F., Heirman, C., Vaeremans, E., Thielemans, K., et al. (2007). An Alternative Pathway of NF-κB Activation Results in Maturation and T Cell Priming Activity of Dendritic Cells Overexpressing a Mutated IκBα1. J. Immunol. *178*, 1301–1311. https://doi.org/10.4049/jimmunol.178.3.1301.

34. Anderson, D.A., Murphy, T.L., Eisenman, R.N., and Murphy, K.M. (2020). The MYCL and MXD1 transcription factors regulate the fitness of murine dendritic cells. Proc. Natl. Acad. Sci. USA *117*, 4885–4893. https://doi.org/10.1073/pnas.1915060117.

35. Xiao, H., Ulmert, I., Bach, L., Huber, J., Narasimhan, H., Kurochkin, I., Chang, Y., Holst, S., Mörbe, U., Zhang, L., et al. (2024). Genomic deletion of Bcl6 differentially affects conventional dendritic cell subsets and compromises Tfh/Tfr/Th17 cell responses. Nat. Commun. *15*, 3554. https://doi.org/10.1038/s41467-024-46966-6.

36. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

37. Tabula Sapiens Consortium (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. https://doi.org/10.1126/science.abl4896.

38. Ebina, W., and Rossi, D.J. (2015). Transcription factor-mediated reprogramming toward hematopoietic stem cells. EMBO J. *34*, 694–709. https://doi.org/10.15252/embj.201490804.

39. Pires, C.F., Rosa, F.F., Kurochkin, I., and Pereira, C.-F. (2019). Understanding and Modulating Immunity With Cell Reprogramming. Front. Immunol. *10*, 2809. https://doi.org/10.3389/fimmu.2019.02809.

40. Fischer, F., Fischer, D.S., Mukhin, R., Isaev, A., Biederstedt, E., Villani, A.-C., and Theis, F.J. (2024). scTab: Scaling cross-tissue single-cell annotation models. Nat. Commun. *15*, 6611. https://doi.org/10.1038/s41467-024-51059-5.

41. Guerriero, A., Langmuir, P.B., Spain, L.M., and Scott, E.W. (2000). PU.1 is required for myeloid-derived but not lymphoid-derived dendritic cells. Blood *95*, 879–885. https://doi.org/10.1182/blood.V95.3.879.003k13_879_885.

42. Brown, C.C., Gudjonson, H., Pritykin, Y., Deep, D., Lavallée, V.-P., Mendoza, A., Fromme, R., Mazutis, L., Ariyan, C., Leslie, C., et al. (2019). Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. Cell *179*, 846–863.e24. https://doi.org/10.1016/j.cell.2019.09.035.

43. Zhu, Y., Cai, P., Li, Z., Zhang, S., Kong, W.T., Wang, H., Gao, F., Zeng, Y., Qian, J., Su, B., et al. (2025). Transcription factors TCF4 and KLF4 respectively control the development of the DC2A and DC2B lineages. Nat. Immunol. *26*, 1275–1286. https://doi.org/10.1038/s41590-025-02208-5.

44. Ratliff, M.L., Garton, J., Garman, L., Barron, M.D., Georgescu, C., White, K.A., Chakravarty, E., Wren, J.D., Montgomery, C.G., James, J.A., et al. (2019). ARID3a gene profiles are strongly associated with human

interferon alpha production. J. Autoimmun. *96*, 158–167. https://doi.org/10.1016/j.jaut.2018.09.013.

45. Hirai, M., Kadowaki, N., Kitawaki, T., Fujita, H., Takaori-Kondo, A., Fukui, R., Miyake, K., Maeda, T., Kamihira, S., Miyachi, Y., et al. (2011). Bortezomib suppresses function and survival of plasmacytoid dendritic cells by targeting intracellular trafficking of Toll-like receptors and endoplasmic reticulum homeostasis. Blood *117*, 500–509. https://doi.org/10.1182/blood-2010-05-284737.
46. Sichien, D., Scott, C.L., Martens, L., Vanderkerken, M., Van Gassen, S., Plantinga, M., Joeris, T., De Prijck, S., Vanhoutte, L., Vanheerswynghels, M., et al. (2016). IRF8 Transcription Factor Controls Survival and Function of Terminally Differentiated Conventional and Plasmacytoid Dendritic Cells, Respectively. Immunity *45*, 626–640. https://doi.org/10.1016/j.immuni.2016.08.013.
47. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.
48. Carotta, S., Dakic, A., D'Amico, A., Pang, S.H.M., Greig, K.T., Nutt, S.L., and Wu, L. (2010). The transcription factor PU.1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-dependent manner. Immunity *32*, 628–641. https://doi.org/10.1016/j.immuni.2010.05.005.
49. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
50. Maier, B., Leader, A.M., Chen, S.T., Tung, N., Chang, C., LeBerichel, J., Chudnovskiy, A., Maskey, S., Walker, L., Finnigan, J.P., et al. (2020). A conserved dendritic-cell regulatory program limits antitumour immunity. Nature *580*, 257–262. https://doi.org/10.1038/s41586-020-2134-y.
51. Du, Y., Wang, J., Jia, J., Song, N., Xiang, C., Xu, J., Hou, Z., Su, X., Liu, B., Jiang, T., et al. (2014). Human hepatocytes with drug metabolic function induced from fibroblasts by lineage reprogramming. Cell Stem Cell *14*, 394–403. https://doi.org/10.1016/j.stem.2014.01.008.
52. Nam, Y.-J., Song, K., Luo, X., Daniel, E., Lambeth, K., West, K., Hill, J.A., DiMaio, J.M., Baker, L.A., Bassel-Duby, R., et al. (2013). Reprogramming of human fibroblasts toward a cardiac fate. Proc. Natl. Acad. Sci. USA *110*, 5588–5593. https://doi.org/10.1073/pnas.1301019110.
53. Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301. https://doi.org/10.1038/nmeth.4177.
54. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17. https://doi.org/10.1016/j.cell.2016.11.038.
55. Domingo, J., Minaeva, M., Morris, J.A., Ghatan, S., Ziosi, M., Sanjana, N.E., and Lappalainen, T. (2024). Non-linear transcriptional responses to gradual modulation of transcription factor dosage. eLife *13*, RP100555. https://doi.org/10.7554/eLife.100555.1.
56. Liu, Y., Yu, C., Daley, T.P., Wang, F., Cao, W.S., Bhate, S., Lin, X., Still, C., Liu, H., Zhao, D., et al. (2018). CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. Cell Stem Cell *23*, 758–771.e8. https://doi.org/10.1016/j.stem.2018.09.003.
57. Black, J.B., McCutcheon, S.R., Dube, S., Barrera, A., Klann, T.S., Rice, G.A., Adkar, S.S., Soderling, S.H., Reddy, T.E., and Gersbach, C.A. (2020). Master Regulators and Cofactors of Human Neuronal Cell Fate Specification Identified by CRISPR Gene Activation Screens. Cell Rep. *33*, 108460. https://doi.org/10.1016/j.celrep.2020.108460.
58. Liu, S., Striebel, J., Pasquini, G., Ng, A.H.M., Khoshakhlagh, P., Church, G.M., and Busskamp, V. (2021). Neuronal Cell-type Engineering by Transcriptional Activation. Front. Genome Ed. *3*, 715697. https://doi.org/10.3389/fgeed.2021.715697.
59. Chopin, M., Lun, A.T., Zhan, Y., Schreuder, J., Coughlan, H., D'Amico, A., Mielke, L.A., Almeida, F.F., Kueh, A.J., Dickins, R.A., et al. (2019). Transcription Factor PU.1 Promotes Conventional Dendritic Cell Identity and Function via Induction of Transcriptional Regulator DC-SCRIPT. Immunity *50*, 77–90.e5. https://doi.org/10.1016/j.immuni.2018.11.010.
60. Theisen, D.J., Ferris, S.T., Briseño, C.G., Kretzer, N., Iwata, A., Murphy, K.M., and Murphy, T.L. (2019). Batf3-Dependent Genes Control Tumor Rejection Induced by Dendritic Cells Independently of Cross-Presentation. Cancer Immunol. Res. *7*, 29–39. https://doi.org/10.1158/2326-6066.CIR-18-0138.
61. Xu, H., Li, Z., Kuo, C.-C., Götz, K., Look, T., de Toledo, M.A.S., Seré, K., Costa, I.G., and Zenke, M. (2023). A lncRNA identifies Irf8 enhancer element in negative feedback control of dendritic cell differentiation. eLife *12*, e83342. https://doi.org/10.7554/eLife.83342.
62. Kim, S., Bagadia, P., Anderson, D.A., Liu, T.-T., Huang, X., Theisen, D.J., O'Connor, K.W., Ohara, R.A., Iwata, A., Murphy, T.L., et al. (2020). High Amount of Transcription Factor IRF8 Engages AP1-IRF Composite Elements in Enhancers to Direct Type 1 Conventional Dendritic Cell Identity. Immunity *53*, 759–774.e9. https://doi.org/10.1016/j.immuni.2020.07.018.
63. Ade, N., Antonios, D., Kerdine-Romer, S., Boisleve, F., Rousset, F., and Pallardy, M. (2007). NF-κB Plays a Major Role in the Maturation of Human Dendritic Cells Induced by NiSO4 but not by DNCB. Toxicol. Sci. *99*, 488–501. https://doi.org/10.1093/toxsci/kfm178.
64. Hacker, C., Kirsch, R.D., Ju, X.-S., Hieronymus, T., Gust, T.C., Kuhl, C., Jorgas, T., Kurz, S.M., Rose-John, S., Yokota, Y., et al. (2003). Transcriptional profiling identifies Id2 function in dendritic cell development. Nat. Immunol. *4*, 380–386. https://doi.org/10.1038/ni903.
65. Bagadia, P., Huang, X., Liu, T.T., Durai, V., Grajales-Reyes, G.E., Nitschké, M., Modrusan, Z., Granja, J.M., Satpathy, A.T., Briseño, C.G., et al. (2019). An Nfil3–Zeb2–Id2 pathway imposes Irf8 enhancer switching during cDC1 development. Nat. Immunol. *20*, 1174–1185. https://doi.org/10.1038/s41590-019-0449-3.
66. Ghosh, H.S., Cisse, B., Bunin, A., Lewis, K.L., and Reizis, B. (2010). Continuous expression of the transcription factor e2-2 maintains the cell fate of mature plasmacytoid dendritic cells. Immunity *33*, 905–916. https://doi.org/10.1016/j.immuni.2010.11.023.
67. Caton, M.L., Smith-Raska, M.R., and Reizis, B. (2007). Notch–RBP-J signaling controls the homeostasis of CD8- dendritic cells in the spleen. J. Exp. Med. *204*, 1653–1664. https://doi.org/10.1084/jem.20062648.
68. Guo, W., Li, Z., Anagnostopoulos, G., Kong, W.T., Zhang, S., Chakarov, S., Shin, A., Qian, J., Zhu, Y., Bai, W., et al. (2024). Notch signaling regulates macrophage-mediated inflammation in metabolic dysfunction-associated steatotic liver disease. Immunity *57*, 2310–2327.e6. https://doi.org/10.1016/j.immuni.2024.08.016.
69. Taveirne, S., Wahlen, S., Van Loocke, W., Kiekens, L., Persyn, E., Van Ammel, E., De Mulder, K., Roels, J., Tilleman, L., Aumercier, M., et al. (2020). The transcription factor ETS1 is an important regulator of human NK cell development and terminal differentiation. Blood *136*, 288–298. https://doi.org/10.1182/blood.2020005204.
70. Kamizono, S., Duncan, G.S., Seidel, M.G., Morimoto, A., Hamada, K., Grosveld, G., Akashi, K., Lind, E.F., Haight, J.P., Ohashi, P.S., et al. (2009). Nfil3/E4bp4 is required for the development and maturation of NK cells in vivo. J. Exp. Med. *206*, 2977–2986. https://doi.org/10.1084/jem.20092176.
71. Wong, P., Foltz, J.A., Chang, L., Neal, C.C., Yao, T., Cubitt, C.C., Tran, J., Kersting-Schadek, S., Palakurty, S., Jaeger, N., et al. (2023). T-BET and EOMES sustain mature human NK cell identity and antitumor function. J. Clin. Investig. *133*, e162530. https://doi.org/10.1172/JCI162530.
72. Fang, D., Cui, K., Cao, Y., Zheng, M., Kawabe, T., Hu, G., Khillan, J.S., Li, D., Zhong, C., Jankovic, D., et al. (2022). Differential regulation of transcription factor T-bet induction during NK cell development and T

helper-1 cell differentiation. Immunity *55*, 639–655.e7. https://doi.org/10.1016/j.immuni.2022.03.005.

73. Wagner, J.A., Wong, P., Schappe, T., Berrien-Elliott, M.M., Cubitt, C., Jaeger, N., Lee, M., Keppel, C.R., Marin, N.D., Foltz, J.A., et al. (2020). Stage-Specific Requirement for Eomes in Mature NK Cell Homeostasis and Cytotoxicity. Cell Rep. *31*, 107720. https://doi.org/10.1016/j.celrep.2020.107720.
74. Zhuang, Q., Cai, H., Cao, Q., Li, Z., Liu, S., and Ming, Y. (2020). Tolerogenic Dendritic Cells: The Pearl of Immunotherapy in Organ Transplantation. Front. Immunol. *11*, 552988. https://doi.org/10.3389/fimmu.2020.552988.
75. Kaseniit, K.E., Katz, N., Kolber, N.S., Call, C.C., Wengier, D.L., Cody, W.B., Sattely, E.S., and Gao, X.J. (2023). Modular, programmable RNA sensing using ADAR editing in living cells. Nat. Biotechnol. *41*, 482–487. https://doi.org/10.1038/s41587-022-01493-x.
76. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
77. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
78. Rual, J.-F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.-O., et al. (2004). Human ORFeome version 1.1: a platform for reverse proteomics. Genome Res. *14*, 2128–2135. https://doi.org/10.1101/gr.2973604.
79. Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P., et al. (2007). hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics *89*, 307–315. https://doi.org/10.1016/j.ygeno.2006.11.012.
80. Heumos, L., Schaar, A.C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M.D., Strobl, D.C., Henao, J., Curion, F., et al. (2023). Best practices for single-cell analysis across modalities. Nat. Rev. Genet. *24*, 550–572. https://doi.org/10.1038/s41576-023-00586-w.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| APC/Cyanine7 anti-human CD45 Antibody | Biolegend | Cat #304014; RRID: AB_314402 |
| PE/Cyanine7 anti-human HLA-DR Antibody | Biolegend | Cat #307616; RRID: AB_493588 |
| APC anti-human CD45 Antibody | Biolegend | Cat #304037; RRID: AB_2562049 |
| Brilliant Violet 785 anti-human CD40 Antibody | Biolegend | Cat #334340; RRID: AB_2566211 |
| APC anti-human CD80 Antibody | Biolegend | Cat #305220; RRID: AB_2076147 |
| APC anti-human CD197 (CCR7) Antibody | Biolegend | Cat #353214; RRID: AB_10917387 |
| APC/Fire 750 anti-human XCR1 Antibody | Biolegend | Cat #372608; RRID: AB_2687371 |
| PE anti-human CD56 | Biolegend | Cat #, 304606; RRID: AB_314448 |
| APC anti-human CD56 | Biolegend | Cat #, 304610; RRID: AB_314452 |
| APC/Fire 750 anti-human CD107a (LAMP-1) Antibody | Biolegend | Cat #328654; RRID: AB_2890761 |
| FITC anti-human CD19 Antibody | Biolegend | Cat #302256; RRID: AB_2564143 |
| PE/Cyanine7 anti-human CD3 Antibody | Biolegend | Cat #300420; RRID: AB_439781 |
| NKG7 (E6S2A) Rabbit mAb | Cell Signaling Technology | Cat #84835T |
| Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 | Invitrogen | Cat #A-11008; RRID: AB_143165 |
| PerCP/Cyanine5.5 anti-human CD16 | Biolegend | Cat #302028; RRID: AB_893262 |
| Bacterial and virus strains | | |
| Competent *E. coli* DH5a | NEB | Cat #C2987I |
| Stellar Competent Cells | Takara | Cat #636763 |
| Chemicals, peptides, and recombinant proteins | | |
| Polybrene | Sigma-Aldrich | Cat #TR-1003-G |
| Polyethylenimine (PEI) | Sigma-Aldrich | Cat #408727 |
| Recombinant FLT3L | Miltenyi | Cat #130-096-480 |
| Human SCF Recombinant Protein | Peprotech | Cat #300-07-10UG |
| Human IL-3 Recombinant Protein | Peprotech | Cat #200-03-10UG |
| Human IL-7 Recombinant Protein | R&D systems | Cat #207-IL-010 |
| Human IL-15 Recombinant Protein | Peprotech | Cat #200-15-10UG |
| Critical commercial assays | | |
| In-Fusion HD Cloning Kit | Takara | Cat #639650 |
| Lenti-X Concentrator | Takara | Cat #631232 |
| Lenti-X qRT-PCR titration kit | Takara | Cat #631235 |
| LEGENDplex Human CD8/NK Panel (13-plex) w/ FP V02 | Biolegend | Cat #741186 |
| Chromium Single Cell 3' v2 Reagent Kit | 10X Genomics | Cat #120237 |
| Rhapsody 8 Lane Cartridge | BD Biosciences | Cat #666262 |
| BD Rhapsody Enhanced Cartridge Reagent Kit | BD Biosciences | Cat #664887 |
| Rhapsody cDNA Kit | BD Biosciences | Cat #633773 |
| BD Rhapsody WTA Amplification Kit | BD Biosciences | Cat #633801 |
| Rhapsody Targeted mRNA & AbSeq Amp Kit | BD Biosciences | Cat #633774 |
| Rhapsody Custom Panel 2-99 Genes/TARGET | BD Biosciences | Cat #633777 |
| Deposited data | | |
| Single cell RNA-seq barcoded dataset | This paper | GEO: GSE298471 |
| Bulk RNA-seq reprogrammed NK cells | This paper | GEO: GSE298487 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| The Human Transcription Factors Database | Lambert et al.[36] | N/A |
| Human TFome Database | Ng et al.[27] | GEO: GSE159786 |
| Tabula Sapiens Dataset | Tabula Sapiens Consortium et al.[37] | GEO: GSE201333 |
| Human cDC1 reprogramming Dataset | Rosa et al.[8] | GEO: GSE162650 |
| TF Atlas Dataset | Joung et al.[21] | GEO: GSE216481 |
| DC1 immuno-regulatory module Dataset | Maier et al.[50] | GEO: GSE131957 |
| Distinct cDC1 subpopulation Dataset | Meiser et al.[32] | GEO: GSE200174 |
| Human blood myeloid cell dataset | Villani et al.[31] | GEO: GSE94820 |
| Cross-tissue Immune Cell Atlas | Domínguez Conde et al.[15] | ArrayExpress: E-MTAB-11536 |
| **Experimental models: Cell lines** | | |
| Human Embryonic Fibroblasts (HEFs) | Zaal Kokaia, Henrik Ahlenius, and Ella Quist | N/A |
| HEK293T | ATCC | Cat #CRL-3216 |
| A2058 cell line | Ascic et al.[10] | N/A |
| Ca922 cell line | Ascic et al.[10] | N/A |
| SKLMS1 cell line | Ascic et al.[10] | N/A |
| T98G cell line | Zimmermannova et al.[9] | N/A |
| **Oligonucleotides** | | |
| TF cloning primers | This paper | Available in Table S9 |
| BD Rhapsody targeted primers | This paper | Available in Table S9 |
| **Recombinant DNA** | | |
| psPAX2 | Didier Trono, via Addgene | RRID: Addgene_12260 |
| pMD2.G | Didier Trono, via Addgene | RRID: Addgene_12259 |
| pRRL.PPT-SFFV-TF plasmids | This paper | Available in Table S8 |
| **Software and algorithms** | | |
| FlowJo v10 | Becton, Dickinson | https://www.flowjo.com/ |
| CytoML v.16.0 | Bioconductor | https://www.bioconductor.org/packages/release/bioc/html/CytoML.html |
| flowCore v.2.16.0 | Bioconductor | https://bioconductor.org/packages/release/bioc/html/flowCore.html |
| flowWorkspace v.4.16.0 | Bioconductor | https://www.bioconductor.org/packages/release/bioc/html/flowWorkspace.html |
| PRISM v9 | GraphPad | https://www.graphpad.com/scientificsoftware/prism; |
| R v4.2.3 | R Core Team | https://www.r-project.org |
| Cell Ranger v6 | 10X Genomics | https://www.10xgenomics.com |
| STAR v2.7.5 | Dobin et al.[76] | https://github.com/alexdobin/STAR |
| BD Rhapsody Sequence Analysis Pipeline v2.2.1 | BD Biosciences | https://www.bdbiosciences.com/en-sg/products/software/rhapsody-sequence-analysis-pipeline |
| Seurat v4.3.0 | Hao et al.[77] | https://satijalab.org/seurat/ |
| Monocle3 | Cao et al.[30] | https://cole-trapnell-lab.github.io/monocle3/ |
| scTab | Fischer et al.[40] | https://pklab.med.harvard.edu/felix/data/scTab-checkpoints.tar.gz |
| DecisionTreeRegressor | Pedregosa et al.[49] | https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html |
| Celltypist v1.6.2 | Domínguez Conde et al.[15] | https://www.celltypist.org/ |
| Immune-restricted TF identification | This paper | Zenodo: https://doi.org/10.5281/zenodo.17348409 |
| TF barcode demultiplexing | This paper | Zenodo: https://doi.org/10.5281/zenodo.17348409 |
| Decision tree modeling | This paper | Zenodo: https://doi.org/10.5281/zenodo.17348409 |

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Cell lines and primary cell cultures

Human embryonic fibroblasts (HEFs, passages 6 to 10) derived from 6-weeks old human embryos and human embryonic kidney (HEK) 293T cells were maintained in growth media (Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) heat-inactivated fetal bovine serum, antibiotics (10 μg/mL penicillin and streptomycin), and 2 mM L-glutamine (complete DMEM). HEFs were grown on 0.1% porcine gelatine-coated dishes. Cancer cell lines were maintained in DMEM (A2058 and SKLMS1), RPMI 1640 (Ca922) or EMEM (T98G) growth medium supplemented with 10% (v/v) fetal bovine serum (FBS), antibiotics (10 μg/mL penicillin and streptomycin), and 2 mM L-glutamine. Media were changed every 2-3 days and cells were maintained at 37°C and 5% (v/v) $CO_2$.

Cell dissociation from tissue culture plates was performed with TrypLE Express for 5-8 min at 37°C. All cells were maintained at 37°C and 5% (v/v) CO2. All tissue culture reagents were from Thermo Fisher Scientific unless stated otherwise.

## METHOD DETAILS

### Immune-restricted TF identification

We compiled a curated list of human TFs by combining the Human Transcription Factors[36] and Human TFome[27] databases, resulting in 1,829 TFs. For the selection of TFs for immune reprogramming, we utilized publicly available scRNA-seq datasets, including Tabula Sapiens.[37] We used pre-defined annotations such as cell_ontology_class, based on the Cell Ontology, and free_annotation based on free text, to categorize cell types as either non-immune or immune and hematopoietic. To normalize the data, we divided the counts of each gene by the total counts in that cell and multiplied by 10,000 to obtain unlogged normalized counts. Subsequently, we applied natural-log transformation using log1p to obtain normalized counts for each gene in each cell. We then used this list to filter immune-restricted TFs. We devised four scoring methods for each TF (See Table S7), which aimed to evaluate their potential impact on reprogramming within each immune cell type: (1) Fold-change expression score method, which compares the average TF expression value in individual immune cells when compared to non-immune cells. The fold-change expression score is calculated by subtracting the TF expression level of the non-immune cell type from the TF expression level of the immune cell type. (2) Fold-change unlogged expression score method, which uses unlogged normalized counts for calculation of expression values. The fold-change unlogged expression score is calculated by subtracting the log1p of the TF non-immune expression level from the log1p of the TF immune cell type expression level. (3) Fraction of cells with expression score method, which determines the fraction of cells expressing each TF within the immune and non-immune cell types. The fraction of cells with expression score is obtained by subtracting the fraction of non-immune TFs from the fraction of TF-immune cell types. (4) Jensen-Shannon divergence (JSD) score method, which measures the divergence of TF expression patterns between immune and non-immune cell types using a vector-based approach. The vector is composed of the first element being a normalized value in immune cell type and the remaining elements the corresponding normalized value for each non-immune cell type, separately. Next, the vector is normalized to sum up to unity. Then, an ideal vector of the same length with the first value set to 1 and the remaining values set to 0 is created. A score for each TF and cell type is calculated as the Jensen-Shannon divergence between these two vectors. The TFs were ranked separately for each immune cell type based on their scores in the respective cell type. We then calculated for each scoring method and each TF: (1) the best rank across all immune cell types, (2) a rank product across all immune cell types (the geometric mean of the ranks across all immune cell types), (3) a p-value and a percentage of false positive predictions (pfp) for the rank product, as estimated using the R package RankProd, and (4) the rank of the rank products across all immune cell types. For the fold-change expression score method and the fraction of cells with expression score method, the number of TFs selected as immune-restricted TFs was determined as the median number of TFs across immune cell types with scores larger than zero. For the Jensen-Shannon divergence score, the number of selected TFs was based on the TFs with rank product p-values smaller than 0.01. Due to the high correlation observed among the methods — with the exception of JSD, which relies on an idealized comparison approach — we chose to use fold-change enrichment, the most widely adopted method, to define immune TFs. The list of immune TFs was generated by compiling the enriched TFs per individual immune cell type. The ranked lists corresponding to each immune cell type are accessible at https://cellreprolab.shinyapps.io/immuneReproTF as a resource for identifying candidate TF to program individual immune cell identities of interest.

### Molecular cloning

The REPROcode TF library, comprising 408 TF immune ORFs, was constructed in two stages (See Tables S8 and S9). First, a library of barcoded plasmids was generated, with each cellular barcode containing a 20-bp sequence composed of a 6-bp constant sequence flanking an 8-bp variable sequence. This design allows for mismatches, accommodating for potential sequencing errors. The barcodes were positioned between the WPRE and the LTR, a non-disruptive region that optimally situates them upstream of the poly-A tail for efficient detection. Barcodes were cloned into the lentiviral vector pRRL.PPT-SFFV-MCS-Brc (SFFV), using In-Fusion cloning (Takara Bio, 639650) and the restriction enzyme XbaI, with PCR-based addition of overlap regions, following the manufacturer's protocols. Second, individual ORFs were cloned into individual barcoded vectors. Due to lentiviral packaging limitations, TF ORFs exceeding 5,800 bp were excluded, narrowing the total to 408 TFs. ORFs for the selected Human TFs were sourced from the CCSB Human ORFeome collection,[78,79] MORF collection,[21] Addgene.org, Group's internal collection, or synthesized (twist bioscience). Amplified TF ORFs with the cloning overhangs for In-Fusion cloning were then inserted into the multiple cloning site (MCS) of

the barcoded vector using the restriction enzymes BstBI and AgeI, resulting in the complete construction of the library with a total of 408 plasmids. All constructs were verified through Sanger sequencing (Eurofins Genomics) and deposited in Addgene.org. The list of selected TFs, along with their respective barcodes and sequences, is provided in Table S8.

To induce cDC1 reprogramming, we used the lentiviral vector pRRL.PPT.sf.hPIB.i2eGFP or pRRL.PPT.sf.hPIB.i2dTomato[8] expressing human PU.1, IRF8, and BATF3. Empty MCS pRRL.PPT.sf.MCS.i2eGFP or pRRL.PPT.sf.hPIB.i2dTomato served as controls throughout reprogramming experiments. For validation experiments, the coding sequences of candidate TFs were cloned into the MCS of lentiviral vectors containing the constitutive SFFV promoter, an internal ribosomal entry site and eGFP or dTomato (pRRL.PPT.sf.MCS.i2eGFP or pRRL.PPT.sf.MCS.i2dTomato).

### Pooled lentiviral production

To prepare the library of 48 TF plasmids for pooled lentiviral production, plasmids were joined at equal concentrations (5 ng/μL) and electrotransformed into Endura ElectroCompetent Cells following the manufacturer's protocol. Following recovery, we plated transformed bacteria onto LB agar plates with ampicillin and incubated for 12 hours at 37°C. We harvested colonies by scraping and immediately isolated plasmid DNA by NucleoBond Xtra Midi kit.

### Lentiviral production and titration

To produce individual lentivirus for each barcoded TF, 5 million HEK293T cells were seeded per 15 cm plate 48 hours prior to transfection. For transfection, a mixture of 7.5 μg of the lentiviral packaging vector psPAX2.G, 2.5 μg of envelope vector pMD2, and 10 μg of lentiviral vector with 60 μL of PEI (1 mg/ml, linear 25 kDa, Polysciences) in Opti-MEM was prepared and incubated for 15 minutes at room temperature. Following incubation, the mixture was added dropwise to HEK293Ts cells in DMEM without FBS or Pen-Strep. Transfection media were replaced with DMEM complete after 6 hours. Supernatants containing viral particles were collected 48, 60, and 72 hours after transfection. The supernatants from the three collections were combined, passed through a 0.45 μm low protein binding filter, and concentrated with the Lenti-X Concentrator (Takara) according to manufacturer's instructions. Lentiviral titers were determined using the Lenti-X qRT-PCR titration kit (Takara) according to manufacturer's instructions.

### Viral transduction and reprogramming

To induce reprogramming, HEFs were counted and seeded (500,000 per 10-cm plate) on 0.1% gelatin-coated plates 6 hours prior to transduction. For screening experiments, viral copies (5, 10, 15, 20, or 40 copies) per TF per cell were calculated individually and volumes were pooled (See Table S10) in DMEM complete supplemented with 8 μg/mL polybrene, added to the cells, and incubated overnight. For validation experiments, concentrated virus was functionally titrated, and an appropriate volume of each lentivirus was added accordingly. The next day, after 12-16 hours, transduction media were replaced with DMEM complete supplemented, marking day 0 of reprogramming. Media were changed every 2-3 days for the remainder of reprogramming.

To directly compare reprogramming efficiency between lentiviral vectors with either the SFFV or EF1a promoter, individual TFs were added at either 50 or 100 copies per TF per cell. Plasmids from (Joung et al.)[21] encoding PU.1 (TFORF1652), IRF8 (TFORF3130), and BATF3 (TFORF3500) were ordered from Addgene.

For experiments aimed at validating TF stoichiometry or assessing the impact of additional TFs on PIB-mediated cDC1 reprogramming, HEFs were seeded at a density of 250,000 cells per gelatin-coated 6-well plate. To explore the effects of additional PU.1, IRF8, and BATF3 on the reprogramming process, HEFs were co-transduced with lentiviral particles containing the PIB-IRES-dTomato cassette and lentiviral particles encoding individual PU.1, IRF8, and BATF3 tagged with eGFP. For the validation of additional factors to support the reprogramming process, HEFs were co-transduced with lentiviral particles containing the PIB-IRES-eGFP and increasing amounts of an additional TF-IRES-dTomato (ranging from 3 to 15μL).

To examine TF expression levels by flow cytometry, lentiviral vectors encoding PU.1-T2A-mTagBFP2, IRF8-T2A-miRFP670nano, and BATF3-T2A-mNeonGreen were obtained from VectorBuilder. Lentiviruses encoding for individual TFs were co-delivered either to HEFs or to the human cancer cell lines A2058, Ca922, SKLMS1, and T98G. Cells were seeded at a density of 12,500 cells per 12-well plate and transduced with lentiviral particles encoding for individual TFs at 12 cp for PU.1 and IRF8, and at 1.5 for BATF3, to achieve effective co-delivery. At day 9 post-transduction, cells were analyzed by flow cytometry.

### Flow cytometry analysis

Analysis of surface marker expression was performed in dissociated cell suspensions. To prepare cells for flow cytometry analysis, cells were detached using TrypLE Express and resuspended at a concentration of $10^6$ cells per 100 μL of DPBS containing 5% FBS. Cells were then incubated with appropriate concentrations of antibody at 4°C for 30 min, with the addition of mouse serum (1/100; GeneTex). Following the staining, cells were washed, resuspended in DPBS containing 5% FBS, and kept at 4°C until analysis. To exclude dead cells, 7-Aminoactinomycin D (7AAD, Thermo Fisher Scientific) or 4′,6-diamidino-2-phenylindole (DAPI) were added shortly before analysis. Flow cytometry analysis was performed with LSR Fortessa and LSR Fortessa X20 (BD Biosciences). Flow cytometry data were analyzed using FlowJo v.10.0.7 (FlowJo LLC). Gates were determined according to fluorescence minus one (FMO) controls. All flow cytometry analysis were performed in live-single gates, unless otherwise specified. For analyzing the combinatorial effects of TF levels in cells lines transduced with TFs coupled to fluorescent proteins, events within live-single-cell gates were extracted using FlowJo and subsequently processed in R using the following packages: CytoML (v.16.0), flowCore (v.2.16.0), and flowWorkspace (v.4.16.0). As part of data cleanup, the top and bottom 0.5% of events per TF channel were excluded to remove

extreme outliers. Events for each fluorescent reporter channel were then divided into ten intervals, representing TF levels from lowest (1) to highest (10). Three-dimensional bins were created to capture specific combinations of these three TF levels. Median CD45 intensity was calculated within each bin to assess reprogramming efficiency and to identify the optimal TF level combination.

### Fluorescence-activated cell sorting (FACS)

After 2 or 9 days of reprogramming, HEFs were dissociated using TrypLE Express, centrifuged, and resuspended in PBS with 0.04% bovine serum albumin (BSA). Cells were incubated with mouse-serum (1:100, GeneTex), mouse anti-human fluorescently labelled anti-CD45 and anti-HLA-DR antibodies for 30 minutes at 4°C. DAPI was added shortly before analysis to allow for the exclusion of dead cells. Single-live double negative (DN; $CD45^{-}HLA\text{-}DR^{-}$), single positive (SP; $CD45^{+}$), or double positive (DP; $CD45^{+}HLA\text{-}DR^{+}$) cells were then purified on a BD FACSAria III (BD Biosciences) using a 100 μm nozzle. Transduced cells at day 2 were also FACS-sorted for single-live. Cells were sorted directly into 1.5 mL Eppendorf tubes with PBS with 0.04% BSA. All cells were sorted on a BD FACSAriaIII (BD Bioscience).

### cDC1 reprogramming with additional regulators

To validate the impact of additional TFs on cDC1 reprogramming, we performed co-transductions of hPIB-eGFP with lentiviral particles encoding individual TFs (GATA2, GFI1B, STAT3, IRF2, TCF4, MYC, NFKBIA, NFKBIB, ID2, MXD1, BCL6 and IRF7) upstream IRES-dTomato. We seeded 600K HEFs per 12-well cell culture plates pre-treated with 0.1% porcine gelatine, and co-transduced them with hPIB-eGFP lentiviral particles to induce reprogramming and with TF-dTomato lentiviral particles to elicit the expression of candidate genes. Transduction with TF-dTomato was performed at 3 different timepoints: 48 hours before the reprogramming initiation (day 2), simultaneously with reprogramming induction (day 0), or 72 hours after (day +3). Empty eGFP and dTomato lentiviral vectors were used as experimental controls. To evaluate reprogramming efficiency, we analyzed the percentage of fully reprogrammed ($CD45^{+}HLA\text{-}DR^{+}$) and partially reprogrammed ($CD45^{+}HLA\text{-}DR^{-}$/$CD45^{-}HLA\text{-}DR^{+}$) cells within single-live $eGFP^{+}dTomato^{+}$ populations. To evaluate reprogramming fidelity, we further analyzed the expression of CD40, XCR1, CD80, and CCR7, gated in $CD45^{+}$.

### Single-cell RNA sequencing (scRNA-seq)

For low-output sequencing, single-cell suspensions containing between 5,000 and 15,000 cells for each condition [(day 2 (live cells), day 9 DN (double negative, $CD45^{-}HLA\text{-}DR^{-}$), day 9 SP (single positive, $CD45^{+}$), day 9 DP (double positive, $CD45^{+}HLA\text{-}DR^{+}$)] were prepared in PBS with 0.04% BSA and loaded on a 10x Chromium (10x Genomics), following the manufacturer's protocol. Chromium Single Cell 3' v2 Reagent Kits were used to prepare the scRNA-seq libraries according to the manufacturer's instructions. Library quantification and quality assessment were performed using an Agilent Bioanalyzer with a High Sensitivity DNA analysis kit (Agilent). Equimolar indexed libraries were pooled and subjected to paired-end 26 x 98 bp sequencing on an Illumina NovaSeq 6000. The resulting coverage was approximately 18,000 reads per cell.

For high-output sequencing approximately 30,000 to 50,000 cells were sorted for each condition [day 9 DN ($CD45^{-}HLA\text{-}DR^{-}$) and day 9 SP ($CD45^{+}$)] for 42 TFs (10 cp/TF/cell). Sorted cells were then labelled with BD AbSeq Ab-Oligos using the BD Human Single-Cell Multiplexing Kit according to the manufacturer's instructions. Following washing, oligo-tagged cells were pooled in even ratios, and approximately 20,000 cells were loaded onto the BD Rhapsody Cartridge. For higher multiplexed barcoded pools (48 TFs; 10 cp/TF/cell), approximately 55,000 day 9 DN ($CD45^{-}HLA\text{-}DR^{-}$) and day 9 SP ($CD45^{+}$) sorted cells were loaded per lane onto the BD Rhapsody 8-Lane Cartridge. Subsequent steps were performed using the BD Rhapsody Cartridge Reagent Kit or the BD Rhapsody Enhanced Cartridge Reagent Kit, as specified in the manufacturer's protocol. Following cell capture bead retrieval, cDNA synthesis was immediately initiated using the BD Rhapsody cDNA Kit. Subsequently, scRNA-seq libraries were prepared using either the BD Rhapsody WTA Amplification Kit for gene expression libraries, or the Rhapsody Targeted mRNA & AbSeq Amp Kit combined with the Rhapsody Custom Primer Panel (Table S9) for barcode amplification libraries. Library quantification, quality, and pooling were performed as described above. Libraries were sequenced on an Illumina NovaSeq 6000 or NovaSeqX using paired-end 100bp sequencing mode. Resulting coverage was approximately 70,000 reads per cell.

### scRNA-seq data processing

For the 10x Genomics platform, paired-end sequencing reads of scRNA-seq were processed using the publicly available software Cell Ranger v6.1.2. First, we used cellranger mkfastq to convert binary base call files to FASTQ files and to decode multiplexed samples simultaneously. Next, we applied cellranger count to FASTQ files and performed alignment to human (hg38) genome assemblies using STAR v2.7.6a.[76] Then, we combined output files from each run to produce one single matrix using cellranger aggr. For the BD Rhapsody platform, paired-end sequencing reads of scRNA-seq were processed using the publicly available software BD Rhapsody Sequence Analysis Pipeline v2.2.1. First, we used bcl2fastq v2.20 to convert binary base call files to FASTQ files and to decode multiplexed samples simultaneously. Next, we applied BD Rhapsody Sequence Analysis Pipeline to FASTQ files and performed alignment to human (hg38) genome assemblies using STAR. The sparse expression matrix generated by cellranger and BD Rhapsody Seven Bridges analysis pipeline was used as input to the Seurat library v4.3.0.[77] Cells and genes were included based on quality control thresholds, following these criteria: (1) total number of unique molecular identifiers (UMIs) detected per cell greater than 3 lower median absolute deviations (MADs) or greater than 5,000 UMIs/cell (for 22, 42, 48 TFs datasets); (2) number of genes detected in each single cell greater than 3 lower MADs; (3) percentage of counts in mitochondrial genes less than 10% (10x Genomics) or 25% (BD

Rhapsody). For the 10x Genomics, data normalization was performed using SCTransform, and 3000 variable features were identified. We used the first 30 principal components for subsequent UMAP and clustering analyses. For the BD Rhapsody, data normalization was performed using "LogNormalize" with a scale factor of 10,000, and 5000 variable features were identified. We used the first 30 principal components for subsequent UMAP and clustering analyses.

### TF barcode demultiplexing

Reads containing the TF barcode sequence were extracted from the processed BAM files produced by the 10X Genomics and BD Rhapsody pipelines. Reads that contained the TF barcode motif were identified by searching for 20-bp barcodes (6-bp constant flanking sequences and 8-bp variable sequence in the middle). We developed a custom python script: extracting reads from the unmapped read fraction of the BAM file, searching for constant 6-bp flanking regions, and reporting the sequence between them. Those sequences were further annotated using a user supply list of correspondence between TF and 8-bp sequence. The resulting output was the tabular file containing cell barcode, UMI, TF barcode sequence and associated TF. The resulting matrix was further imported into R and used to calculate the number of UMI counts for each exogenous TF. Next, we matched each cell with the corresponding TF based on observation of cells with non-zero UMI counts for a particular TF. This approach was also used for barcode demultiplexing from barcode amplified library.

### Reprogramming trajectory analysis

To order cells along a reprogramming pseudotime, we used Monocle 3 v1.3.1.[30] The analysis was conducted on UMAP embeddings with default settings, with the exception of the parameter "use_partition=FALSE," which assumes that all cells in the dataset descend from a common transcriptional ancestor. The root of the trajectory was automatically selected as cells at day 2 of reprogramming. Subsequently, we ordered the cells according to pseudotime and fitted smooth splines with 3 degrees of freedom for both endogenous and exogenous TF levels. Exogenous TF levels were defined as those associated with a specific barcode originating from the lentiviral construct. This approach allowed us to visualize the trajectory of reprogramming and assess the dynamics of TF expression during this process.

### scRNA-seq data analysis to evaluate the impact of additional TFs

To estimate the effect of adding an individual TF to the successful reprogramming combination, we first extracted from the dataset only cells that either contain the PIB TF combination alone (3TFs = PU.1 + IRF8 + BATF3) or PIB + 1 additional TF. Next, we calculated the number of cells in each of these seven distinct TF combinations, separately within the DC1 classified cells (positive population) and within the remaining cells (negative population). For each TF combination and population separately, we calculated a normalized fraction by dividing the number of cells by the number of cells that contained PIB alone. Finally, we subtracted the fractions for the negative population from the fractions for the positive population to generate a normalized frequency score for each additional TF.

### Identification of candidate reprogramming TFs in multiplexed screens with >40 TFs

Typically, combinations of 3-5 TFs are used in reprogramming. In multiplexed screens with 9 TFs in total, we observed sufficient cell numbers with combinations of 3-5 different TFs. However, when increasing the total number of TFs in the screens, most cells contained more than 5 TFs. Therefore, we developed a computational framework to identify candidate reprogramming TF networks from cells containing a larger number of TFs. The computational framework was performed separately starting with 4-TF meta-combinations. Here, we outline the steps of the framework: 1) All possible combinations of 4 TFs were listed given the total set of TFs in the screen. For example, for 48 TFs there were 194,580 possible 4-TF combinations. 2) Multiple 4-TF meta-combinations were assigned to each cell, since cells typically contained more than 4 TFs. 3) The number of cells containing each of those combinations separately within the negative population was calculated (sorted negatively for a particular marker; for example, $CD45^{-}HLA\text{-}DR^{-}$) and the positive population (e.g., a distinct cluster enriched with cells expressing cDC1 gene signature), since, for a more general screen with a large number of TFs, multiple cell types could be induced simultaneously. 4) For each population (positive and negative) separately, a normalized fraction for a TF combination was calculated by dividing the number of cells by the maximum number of cells found for any TF combination in that population. 5) The fractions for the negative population were subtracted from the fractions for the positive population to generate a normalized frequency score for each 4-TF meta-combinations. 6) The top-1000 4-TF meta-combinations with the largest normalized frequency scores were taken and the frequency of each TF appearing in these TF combinations was counted 7) 1000 4-TF meta-combinations that were not included in the top-1000 were randomly subsampled and counted by in how many of these each TF appears. 10000 such subsamplings were generated and an empirical p-value for each TF was estimated as the fraction of subsamplings having at least the same count as the count for the top 1000. 8) Candidate reprogramming TFs for the positive population were selected as having empirical p-values less than 0.05. 9) Steps 1-5 were repeated for a shortened reprogramming candidate list by comparing a cell type specific-cluster of cells against the remaining cells from the positive population.

To estimate the divergence of DC states obtained by adding an individual TF to the PIB combination, we first extracted cells that only contained PIB + 1 additional TF from the dataset. Next, we calculated the number of cells in each of these distinct TF combinations separately within cDC1 states (immature, mature, and migratory[32]) and within remaining cells. For each TF combination and population, we separately calculated a normalized fraction by dividing the number of cells for the maximum number of cells found for any TF combination in the population. Finally, we subtracted the fractions for the negative population from the fractions for the positive population to generate a normalized frequency score for each additional TF.

To extract cDC2 and pDC TF combinations, we used the same logic described above (steps 1-7) and included an additional step to link combinations to a critical TF for each DC lineage.

### Cell type annotation with pre-trained models

For preprocessing the gene expression matrix, we applied a standard single-cell data procedure.[80] Initially, cells with total counts exceeding ±5 MADs were excluded, as well as cells with mitochondrial counts above 22%. The remaining cells were normalized, using either CPM (Counts Per Million) normalization (specifically for CellTypist label transfer) or size-factor normalization (for all other analyses), followed by log1p transformation. Highly variable genes were identified using the Scanpy package with the 'seurat' flavor setting. Cell type annotations were performed using two methods: CellTypist[15] and scTab.[40] For CellTypist, we employed the pre-trained Immune_All_High model (https://www.celltypist.org/models),[15] which includes diverse cell types such as fibroblasts and immune cells. CellTypist was run with default settings, expect for the majority voting pruning, which was omitted due to the dynamic nature of our system. Additionally, we utilized the scTab model, pre-trained on 22 million cells (https://pklab.med.harvard.edu/felix/data/scTab-checkpoints.tar.gz), to predict cell labels and calculate uncertainty scores, following the authors' guidelines.

### Decision tree modeling

The gene expression matrix was pre-processed as described above. For the TF barcode count matrix, we applied size-factor normalization, followed by log1p transformation. We then trained a DecisionTreeRegressor[49] with 'max_depth=6' on the size-factor-normalized, log-transformed barcode matrix, using as target variables an intersected set of highly variable genes (HVGs) identified independently for each sample (n = 2,522). Using HVGs as predictive targets improves comparability across samples by focusing on consistently informative features. Model performance was evaluated via five-fold cross-validation, yielding mean squared error (MSE) values of 0.096 (Experiment 1), 0.155 (Experiment 2), and 0.13 (Experiment 3), with corresponding $R^2$ values of 0.03, 0.04, and 0.02, respectively. To assess structural similarity of the models across biological replicates, we computed Spearman correlations between the feature importance vectors. We observed high concordance: Exp. 1 vs. 2: r = 0.86; Exp. 2 vs. 3: r = 0.67; Exp. 1 vs. 3: r = 0.74. We further assessed model generalizability using an independent test set composed of equal numbers of cells from each experiment. Predictive performance remained high regardless of whether the test data originated from the same experiment as the training set or from a different replicate (Pearson r = 0.83, 0.85, and 0.94 for models trained on Exp. 1, 2, and 3, respectively), indicating strong cross-sample robustness. To assign cell lineages to each terminal branch of the decision tree, we scored lineage-specific marker gene signatures using Scanpy's score_genes function, adjusting for gene set length. Cell types were designated based on unique enrichment of specific signatures within branches. Finally, we performed differential gene expression analysis using Seurat FindMarkers function with logfc.threshold=0.1, comparing cell populations separated at each branching point as defined by the regression tree. Individual marker gene expression patterns were visually inspected to validate lineage assignments.

### Hierarchical stepwise regression model

To examine the relative contribution of TFs to cellular identity commitment, we employed a stepwise hierarchical regression approach. This method enabled us to evaluate the marginal effects of TFs and determine the optimal size of their combinatorial input. Prior to model construction, features were pre-selected using F-regression (SelectKBest method from scikit-learn) to reduce the dimensionality of the interaction term space. Selected features were then incrementally added to the regression model along with all corresponding interaction terms. At each step, we assessed changes in the Akaike Information Criterion (AIC) to evaluate the added predictive value of the combinations. To mitigate overfitting and reduce computational complexity, interaction term order was restricted to a maximum of ten TFs. We further evaluated the contribution of each interaction term using two-way ANOVA to identify statistically significant interactions influencing the model's predictions. All statistical analyses were performed using the statsmodels v.014.4 package.

### Microscopy

For immunofluorescence, FACS-purified CD56-positive ($CD56^+$) cells were seeded overnight in black 96-well plate (ibidi). Cells were fix with 4% paraformaldehyde for 20 min, permeabilized with 0.4% Triton X-100 and blocked for 30 min with 2.5% bovine serum albumin and 10% normal goat serum (Abcam). Staining with primary antibodies was perform overnight at 4°C. Cells were then washed and incubated with secondary antibodies for 1 h 30 min. Hoechst 33342 solution (1 μg/ mL) was added to stain nucleus for 5 min and washed once before visualization. Cells were imaged on CellDiscoverer 7 microscope (Zeiss), at 20x magnification. Data acquisition, image analysis and export were performed with the ZEN 2.5 blue software (Zeiss). For transmission electron microscopy (TEM), 1x 106 $CD56^+$ cells were purified by FACS, pellet by centrifugation and fixed with 2% glutaraldehyde, 2% paraformaldehyde in Sorensen's phosphate buffer overnight at 4 °C. Fixed cell pellets were washed 3x with Sorensen's buffer and post fixed in 1% osmium tetroxide. Samples were dehydrated using increasing concentrations of acetone and embedded in Polybed 812 epoxy resin. Ultrathin sections (50 nm) were then cut from the center of the sample using a Leica EM UC7 ultrotome (Leica Microsystems) equipped with a diamond knife. Sections were mounted on pioloform coated grids, contrasted with 4% uranayl acetat solution and imaged on Thermo Fischer Talos L120C.

### Isolation of NK populations from human peripheral blood

Peripheral blood mononuclear cells (PBMCs) were isolated from healthy donors leucocyte concentrates provided by the Department of Transfusion Medicine, Skåne University Hospital. After diluting the concentrate with PBS (1:1), the isolation protocol was

performed with Lymphoprep tubes (Fisher Scientific) according to the manufacturer's instructions. Erythrocytes were removed by BD Pharm Lyse (BD Biosciences). For NK populations, cells were directly isolated by FACS. When mentioned, cells were purified either for $CD3^{-}CD56^{+}CD16^{-}$ and $CD3^{-}CD56^{dim}CD16^{+}$ NK cells, or simply $CD3^{-}CD56^{+}$ NK lymphocytes. Lymphocytes were selected by FSC and SSC, excluding monocytes and granulocytes.

### Bulk RNA sequencing

For bulk RNA-seq, total RNA from 3,000–7,000 FACS-sorted HEFs or TENE-induced cells was extracted using TRIzol (ThermoFischer Scientific) following the manufacturer's protocol. Total RNA from 10,000 natural cells ($CD56^{+}CD16^{-}$ and $CD56^{dim}CD16^{+}$ PBMC-NK cells) was used for comparison. cDNA was transcribed with SMARTSeq v4 Ultra-low input RNA kit (TakaraBio), amplified with 8 PCR cycles, and purified with AMPure XP beads (Beckman). cDNA was quantified using High sensitivity DNA kit on Bioanalyzer 2100 (Agilent). Library preparation was performed using the Nextera XT DNA library preparation kit (Illumina) or SMART-Seq v4 PLUS Kit (Takara) according to manufacturer's protocol. Libraries were normalized with beads, pooled, and sequenced on Illumina NextSeq 500 platform. For analysis, paired-end reads were mapped to the human genome (Ensembl, release 93) using STAR v2.5.3a. Resulting gene counts were additionally processed with R package DESeq2 and normalized using RLE method. DESeq2 package was used for performing differential expression analysis based on Wald test. We defined upregulated genes by a fold change (FC) > 0 and Benjamini Hochberg (BH) corrected p-value < 0.05 and downregulated genes by FC < 0 and BH-corrected p-value < 0.05. PCA was performed using plotPCA function from DESeq2 package. Functional enrichment analysis of differentially expressed genes was conducted using the Reactome database via the EnrichR R package, and the top five enriched categories were selected.

### Degranulation and cytokine secretion

To test for cytotoxic cytokine secretion, day 12 FACS-sorted $CD56^{+}$ or HEFs transduced with a combination of 4 TFs were seeded at a density of 10,000 cells per well on a tissue-culture 96-well plate. HEF transduced with lentiviruses for empty vector (SFFV-MCS) was used as control. $CD56^{+}$ $CD3^{-}$ peripheral blood NK cells seeded at the same density in non-tissue culture 96-well plate were used as reference. The next day, cells were activated using 100 ng/mL Phorbol 12-myristate 13-acetate (PMA) and 500 ng/mL Ionomycin and incubated for 4h. Non-activated cells were kept for comparison. After incubation, 25 μL of supernatant was harvested and processed with BD LEGENDplex Human CD8/NK Panel (13-plex), according to the manufacturer's instructions. Data were acquired with BD LSRFortessa and analyzed in LEGENDPlex software. For degranulation, non-activated and activated cells were stained for CD107a and analyzed by FACS.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Statistical analysis

Statistical analyses were performed with GraphPad Prism 9 software. Data are shown as mean ± SD. P values are shown when relevant (*P < 0.05; **P < 0.01; ***P< 0.001; ****P < 0.0001). Statistical tests and parameters for each experiment are reported in the respective figure legend.