Data-guided direct reprogramming of human fibroblasts into the hematopoietic lineage

Cooper Stansbury[1, 2], Jillian Cwycyshyn[1, 3], Joshua Pickard[1], Walter Meixner[1], Indika Rajapakse[1, 2, 4], and Lindsey A. Muir[1]

[1]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109 USA
[2]The Michigan Institute for Computational Discovery and Engineering, University of Michigan, Ann Arbor, MI 48109 USA
[3]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109 USA
[4]Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 USA

August 26, 2024

Abstract

Direct reprogramming of human fibroblasts into hematopoietic stem cells (HSCs) presents a promising strategy for overcoming the limitations of traditional bone-marrow transplantation. Despite the potential of this approach, our understanding of the mechanisms driving efficient autologous cell type conversion remains incomplete. Here, we evaluate a novel algorithmically predicted transcription factor (TF) recipe - GATA2, GFIB1, FOS, REL, and STAT5A - for inducing HSC-like states from human dermal fibroblasts. Using flow cytometry and long-read single-cell RNA-sequencing, we demonstrate increased $CD34^{+}$cell populations and high transcriptomic similarity to native HSCs. Additionally, we uncover transcriptional heterogeneity at both gene and isoform levels among induced HSCs, underscoring the complexity of direct reprogramming.

1 Introduction

A small number of hematopoietic stem cells (HSCs) can be used to repopulate the immune system, making them an important resource for treating blood disorders and immune deficiencies [1, 2]. Traditional HSC transplantation approaches risk immune incompatibility and graft-versus-host disease (GVHD) [1, 2], and underrepresented groups match poorly and have worse outcomes from existing donor registries (from CIBMTR data 2016-2020) [3]). Recent advances in direct cellular reprogramming via induced overexpression of transcription factors (TFs) may present an opportunity for autologous cell production [4].

Despite in-depth single cell characterization of hematopoiesis in mice [5, 6] and in humans [7–9], our understanding of the mechanics governing direct cellular reprogramming to HSCs remains limited. Many strategies for cell type conversion have been discovered [10], but it is unclear if existing strategies are optimal in terms of conversion efficiency, conversion time, and reduced risk of oncogenesis. The generation of HSC-like cells from fibroblasts has been previously shown to be technically feasible, but conversion efficiency remains low, with small fractions of initial cell populations transitioning into HSC-like states [11–15]. Further, the functional status of reprogrammed HSCs remains an open question and little is known about how new TF recipes may impact reprogramming trajectories. These results highlight our limited understanding of the underlying molecular mechanisms behind autologous HSC production.

In this study, we tested direct reprogramming of human dermal fibroblasts into the hematopoietic lineage using a novel algorithm-predicted [16] TF recipe: GATA2, GFIB1, FOS, REL, and STAT5A. We evaluated reprogrammed cells using flow cytometry and long-read single cell RNA-sequencing, finding an increased proportion of $CD34^+$ cells using the algorithm-predicted recipe and high transcriptomic overlap between reprogrammed cells and human HSCs. Finally, we identified heterogeneity in transcriptional programs among iHSCs at both gene and isoform levels.

2 Results

2.1 Novel TF-Recipe Produces CD34$^+$ Cells

There are few known TF combinations that can induce direct conversion of fibroblasts into HSCs, and the reprogramming efficiency for existing recipes is not well understood (Supplementary Figure 1). We used our previously published Data-Guided Control (DGC) algorithm to explore whether its predictions could be used to improve fibroblast-to-HSC reprogramming [16]. Among the top 25 DGC-predicted TFs were GATA2, GFI1B, and FOS, which Gomes and colleagues had previously shown to be sufficient as a 3-TF recipe for reprogramming human cells into HSC-like cells [13]. We then asked whether additional DGC-predicted TFs could be used to improve reprogramming efficiency. STAT5A and REL were selected from the top 10 DGC-predicted TFs and used with GATA2, GFI1B, and FOS in a novel 5-TF recipe (Figure 1A). Human dermal BJ fibroblasts were transduced with lentiviral vectors carrying individual TFs within TET-ON cassettes (Figure 1B). Two days after transduction, cells were treated with 1 μg/mL doxycycline (Dox) for 48-72 hours to induce TF expression.

Seven days after the first Dox treatment, reprogramming cultures were evaluated by flow cytometry for surface expression of CD90 and the HSC marker CD34 (Figure 1C-F). Primary CD34$^+$ bone marrow cells (Figure 1C) and cells transduced with our 5-TF recipe without Dox (Figure 1D) were used as controls. Approximately 1% of the cultures transduced with the 3-TF recipe became CD34$^+$ (Figure 1E) [13], while the 5-TF recipe yielded about 3-fold higher frequency of CD34$^+$ cells (Figure 1F).

Taken together, these results suggest that addition of the predicted TFs increase the efficiency of reprogramming toward an HSC phenotype.

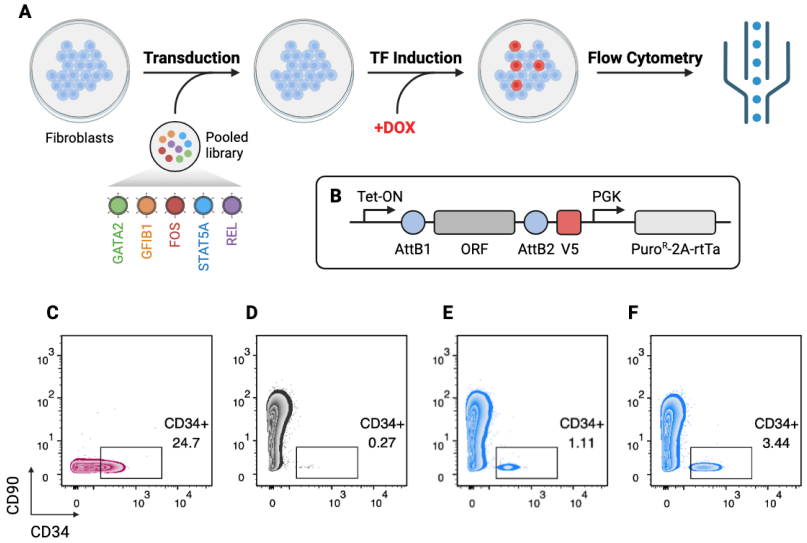


Figure 1: **Direct Reprogramming of Human Fibroblasts Produces CD34 Positive Cells. (A)** Schematic representing the process for reprogramming human fibroblasts to induced hematopoietic stem cells (iHSCs), including lentiviral transduction a transcription factor recipe, induction with doxycycline (+Dox), and quantification of $CD34^+$ cells using flow cytometry. **(B)** Cassette design. **(C-F)** Flow cytometry data showing surface expression of CD34 and CD90 in primary human $CD34^+$ bone marrow cells (C), transduced fibroblasts without Dox (D), fibroblast-to-HSC reprogramming using the 3-TF recipe (GATA2, GFI1B, FOS) [13] (E), and fibroblast-to-HSC reprogramming using the 5-TF recipe (GATA2, GFI1B, FOS, STAT5A, and REL) (F).

2.2 Single-Cell Sequencing Reveals Transcriptomic Heterogeneity in Reprogrammed Cells

We evaluated the persistence of $CD34^+$ cells in reprogramming cultures for up to 50 days. Long-term cultures showed a frequency of $CD34^+$ cells and morphology changes similar to those shown in Figure 1 (Supplementary Figure 2), supporting that reprogrammed cells remain viable and persist with time. To gain a deeper understanding of the transcriptional programs active in these reprogrammed cells, we performed long-read single cell RNA-sequencing (Figure 2A) on sorted $CD34^+$ iHSCs and also included $CD34^-$ fibroblasts as a reference for the initial cellular state (Methods 4.4). We obtained 126 million reads (Supplementary Table 2, Methods 4.4) and 8,574 single cell profiles. Single cell data were compared to population-level BJ fibroblast gene expression from [17] for fibroblast and HSC genes [18] (Supplementary Figures 4A-C). Reprogrammed cells showed decreased expression of fibroblast transcripts (Figure 2B, Supplementary Figure 4B) and elevated expression of HSC transcripts (Figure 2C, Supplementary Figure 4C). Figure 2D shows the proportion of single-cell profiles expressing hematopoietic transcripts (Figure 2E, Supplementary Table 9), including endogenous expression of the reprogramming TFs (Supplementary Table 10).

Clustering of the single-cell profiles using the Leiden algorithm [19, 20] and fibroblast and hematopoietic marker genes [18] (Supplementary Table 3) (Figure 2B) resulted in six major clusters (C1-C6). Clusters C1 and C3 expressed fibroblast-associated transcripts such as *FSTL1* and *KLF2* (Supplementary Table 4). C2 expressed key hematopoietic genes such as *MGST1* and *MPL* (Supplementary Table 6). C4, C5, and C6 showed expression of HSC and/or endothelial markers (Supplementary Table 7) such as *THY1/CD90* and *PECAM1* (Supplementary Figure 3). Control fibroblasts, which were negative for CD34 and exogenous TFs, were present in C1 and C3 (Supplementary Table 5).

Elevated expression of *CD34*, *GATA2*, and *CD164* was observed in all clusters except for the $CD34^-/TF^-$cells in cluster C3 (Figures 2F-G). These results collectively show that C2, C4, and C5 had attenuated fibroblast gene expression and increased HSC gene expression.

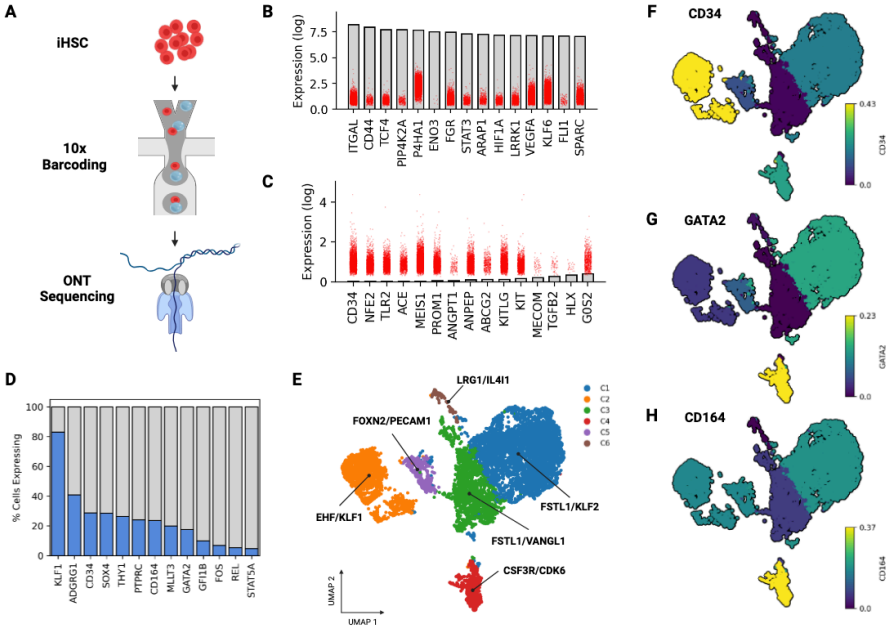


Figure 2: **Single-cell Characterization of iHSCs. (A)** Overview of iHSC single-cell long-read sequencing. **(B)** Grey bars show fibroblast marker gene expression in BJ fibroblasts at the population level. Red points show expression of fibroblast genes in individual iHSCs. **(C)** Grey bars show HSC marker gene expression in BJ fibroblasts. Red points show expression of HSC genes in individual HSCs. **(D)** The percentage of iHSCs that express key HSC transcripts. **(E)** UMAP plot showing clustering results for single-cell profiles. **(F-H)** Mean transcript expression for each cluster of *CD34* (F), *GATA2* (G), and *CD164* (H).

2.3 Reprogrammed Cells are Transcriptomically Similar to HSCs

To assess the degree to which reprogrammed cells were transcriptomically similar to HSCs, we generated a combined dataset that included our iHSC data and publicly available human hematopoietic and fibroblast single cell data, resulting in 54,347 single-cell profiles [7, 9, 21] (Methods 4.6, Supplementary Table 11). Single-cell profiles were visualized using a low-dimensional UMAP projection (Figure 3A-B, Supplementary Figure 5). iHSCs from our study were close to HSC signatures within the UMAP projections (Figure 3B) [7, 9].

To quantify the transcriptomic similarity between iHSCs and HSC and fibroblast profiles, we calculated the cosine distance between the low-dimensional representations (50 principal components) of our cells and the centroids of the HSC and fibroblast populations. iHSC signatures were closer to the HSC centroid than the fibroblast centroid (Figure 3C, Supplementary Figure 6A-B). We further evaluated select HSC markers across the integrated dataset and found that iHSCs were highly similar to HSCs and other progenitors in expression of *CD34*, *GATA2*, *GFI1B*, and other markers (Figure 3D-F, Supplementary Figure 7A-I).

Taken together, these results suggest that reprogrammed cells are transcriptomically more similar to HSCs than fibroblasts.

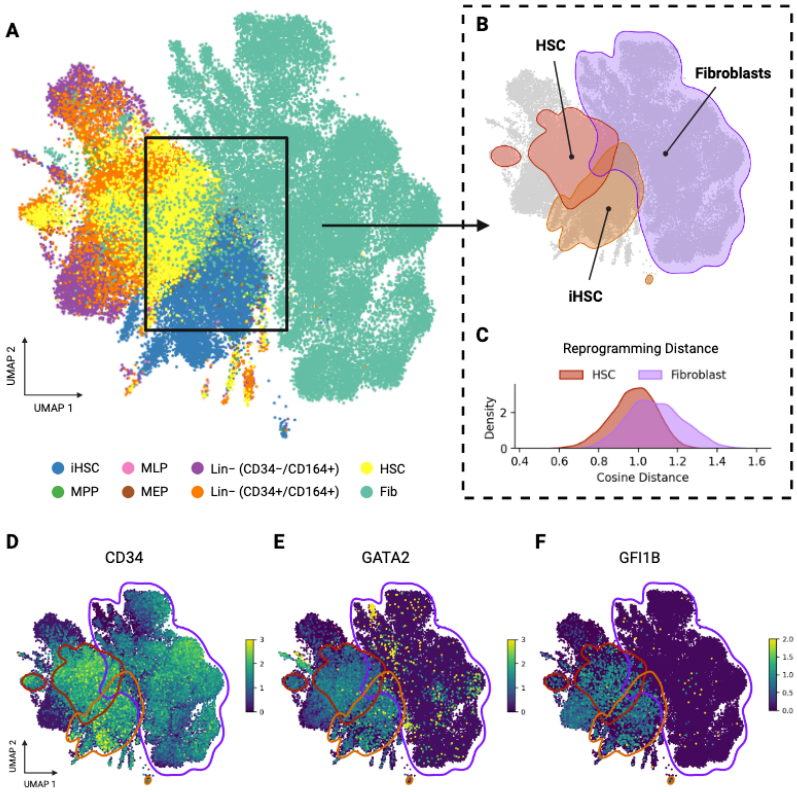


Figure 3: **Reprogrammed Cells and the Hematopoietic Landscape**. **(A)** UMAP plot showing 54,347 single-cell profiles including hematopoietic cell types, iHSCs, and fibroblasts. **(B)** Regions containing hematopoietic stem cells (HSC), fibroblasts, and iHSCs are highlighted for the UMAP projection from (A). **(C)** Distribution of cosine distances between iHSCs and the PCA centroids of HSCs (red) and fibroblasts (purple) from (A). **(D-F)** Expression levels in log CPM of *CD34* (D), *GATA2* (E), and *GFI1B* (F) with contour lines indicating the cell populations from (B).

2.4 Reprogrammed Cells Show Transcriptomic Heterogeneity

To identify iHSCs that were the most similar to HSCs, we analyzed reprogrammed cells alongside 3,243 ($CD34^+/CD45RA^-/CD90^+$) single-cell HSC profiles from Weng *et al.* [9]. Pairwise cosine distances were calculated between iHSCs and HSCs based on the expression of 25 HSC marker genes (Figure 4, Supplementary Table 8). In Figure 4A, the minimum distance between each iHSC and any HSC was measured, and cells in the distribution with a cosine distance of <1 are highlighted in red. Among iHSC clusters, C2, C4, and C5 had the lowest cosine distance, indicating higher similarity to HSCs (Figure 4B) consistent with higher expression of HSC-associated genes as in Figure 2.

We next sought to identify the transcripts that distinguished iHSCs that were closest to native HSCs. Differential expression analysis showed that hematopoiesis-related transcripts were upregulated in iHSCs that were closest to native HSCs compared to iHSCs that had the largest cosine distance (Supplementary Figures 9A-B). Notably, the closest iHSCs expressed many HSC-related transcripts that were not used in distance computations, including *TYROBP* and *MPL* (Supplementary Figures 9A). In contrast, the iHSC signatures that were the most distant from HSCs expressed fibroblast-related genes including the transcription factors *KLF2* and *FSTL1* (Supplementary Figures 9C-D).

For key HSC-related transcripts, the overall expression patterns of the most similar iHSCs resembled native HSCs (Figure 4C, Supplementary Figure 8). Furthermore, the distance between all iHSCs and HSCs was much lower than the distance between iHSCs and dermal fibroblasts, and the iHSC-to-fibroblast distance was very similar to the HSC-to fibroblast-distance (Supplementary Figure 10).

These data support that iHSC transcriptomes closely resemble those of native HSCs and highlight the transcriptomic heterogeneity among iHSCs.

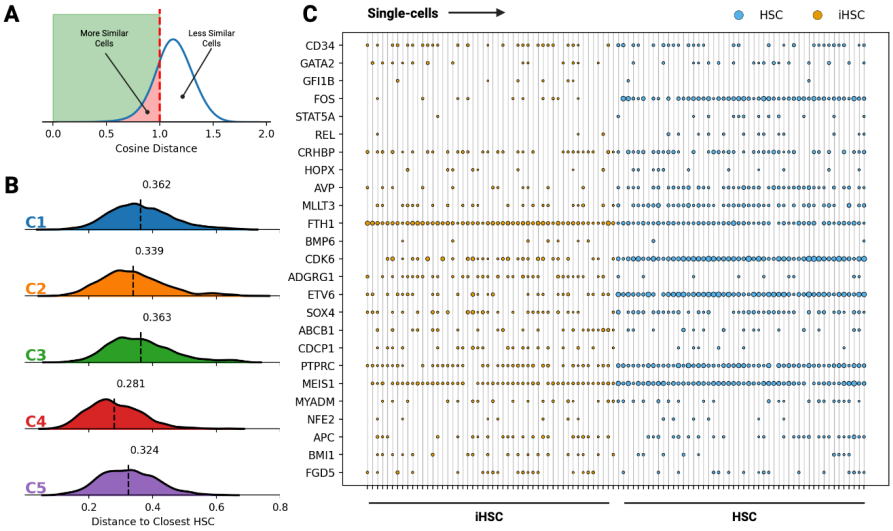


Figure 4: **Identification of reprogrammed iHSCs most similar to HSCs.** **(A)** Distribution of median cosine distances between iHSCs and gold-standard HSC profiles, with the closest iHSCs highlighted in red. **(B)** Median cosine distance distributions for iHSC clusters C1 to C5 (C6 excluded due to low-transcript counts). **(C)** 50 randomly sampled HSC signatures and 50 randomly sampled iHSC with minimum cosine distance ≤ 0.5 for key HSC transcripts.

2.5 Alternative Splicing in Hematopoietic Transcripts

Alternative splicing (AS) is important in the regulation of hematopoiesis and during cellular responses in mature hematopoietic lineage cells [22, 23]. Major changes in splicing are also observed during cell reprogramming to pluripotent cells [24–26]. The specific transcripts undergoing AS during fibroblast reprogramming to hematopoietic stem cells (HSCs) remain largely uncharacterized. We therefore queried our iHSC single-cell long-read sequencing data for patterns of transcript expression at the isoform level, using IsoQuant to quantify isoforms [27] (Methods 4.7).

Multiple isoforms were found for a large proportion of genes, many of which had more than five isoforms (Figure 5A). Key transcription factors and HSC-related genes showed isoform diversity (Figure 5B,C). *APC* was a notable example that had four major isoforms.

Within iHSC clusters, we computed the proportion of cells expressing each isoform, using the most highly expressed isoform to represent each cluster. Clusters C2 and C4 showed distinct patterns of isoform expression across several HSC-related genes, including *CD34*, *RUNX1*, and *OCIAD2* (Figures 5D-F). For instance, three isoforms of *RUNX1* (*RUNX1-202*, *RUNX1-203*, and *RUNX1-215*), were identified in iHSC clusters. Differential expression analysis between clusters revealed clusters C2, C4, and C5 were enriched for specific isoforms of *STAT1* and *THBS1* which are implicated in blood vessel morphogenesis (Supplementary Figure 11).

These results demonstrate isoform diversity in iHSC subclusters that could contribute to heterogeneity in reprogramming outcomes.

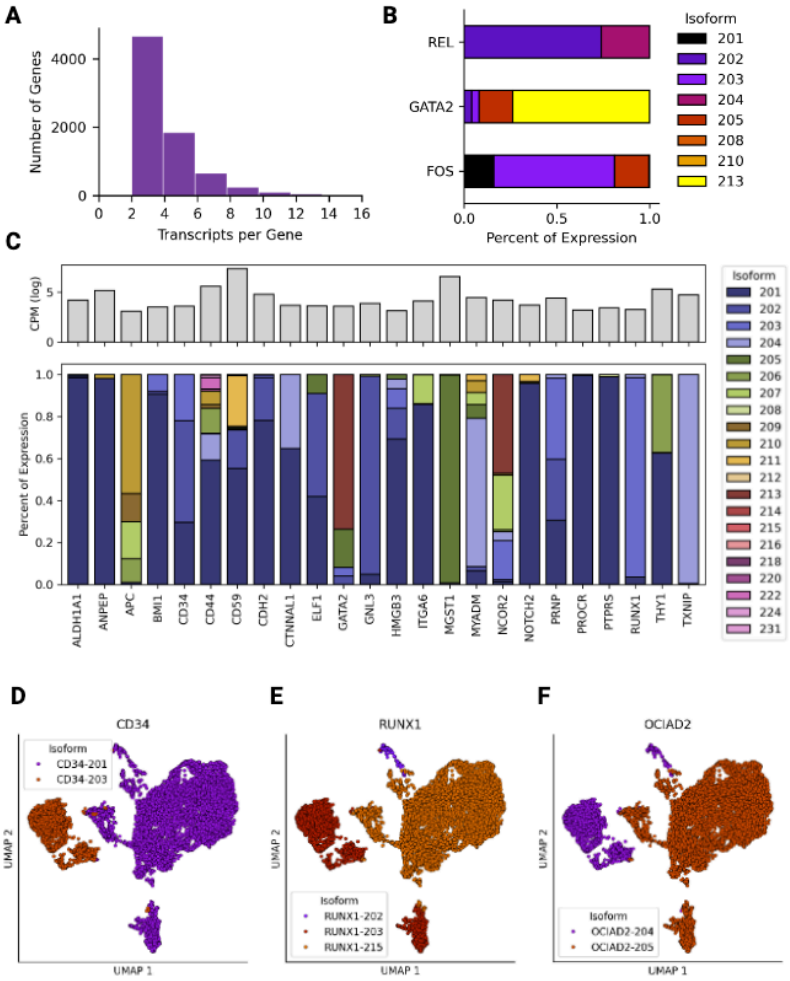


Figure 5: **Isoform Expression in iHSCs.** **(A)** The distribution of the number of detected transcript isoforms per gene in iHSC data. **(B)** Isoform as a percentage of total expression for REL, GATA2 and FOS **(C)** Isoforms as a percentage of total CPM for HSC marker genes with greater than 1,000 reads in iHSC data. **(D)** The most expressed CD34 isoforms by iHSC cluster. **(E)** The most expressed RUNX1 isoforms by iHSC cluster. **(F)** The most expressed OCAID2 isoforms by iHSC cluster.

3 Discussion

To our knowledge, this study is the first to characterize direct HSC reprogramming using single-cell long-read sequencing. Addition of algorithmically-predicted TFs improved the efficiency of reprogramming of human neonatal fibroblasts into the hematopoietic lineage and produced cells which were transcriptomically similar to native HSCs.

In our data, the iHSC clusters that were the most similar to native HSCs (C2, C4, and C5) also expressed endothelial markers such as *ADORA2A* and *CASZ1* (Supplementary Figure 12, 13). Interestingly, these clusters also expressed specific isoforms of genes implicated in blood vessel morphogenesis such as *STAT1* and *THBS1*, which also have roles in the regulation of endothelial cell proliferation (Supplementary Figure 11). Consistent with our findings, prior work has also found endothelial-like precursor cells during HSC reprogramming in both mice [28, 29] and humans [13, 14, 30]. Ours and prior studies thus support the hypothesis that direct reprogramming to HSCs may recapitulate, at least in part, the processes involved in the endothelial-to-HSC transition (EHT) in development [30–32].

Our iHSCs showed similar gene expression when compared to reprogrammed cells and umbilical cord blood cells (UCB) from previously published work (Supplementary Figures 14A-D) [13]. We hypothesize that inclusion of STAT5A and REL increased conversion efficiency through their roles in hematopoiesis and the mitigation of stress responses, respectively [33–35].

Our long-read sequencing data enabled a refined analysis of iHSCs by examining transcript isoforms. The variance in transcript isoforms by iHSC cluster suggests AS as an additional source of heterogeneity in reprogramming. However, further studies will be required to determine the overlap between iHSCs and native HSCs at the isoform level. Few datasets are available since relatively few studies have investigated the role of AS in hematopoiesis, and much of our understanding to date comes from mouse models [22, 23]. Part of the challenge in isoform analysis is that discovery and quantification is limited when using short-read sequencing data, especially at the single-cell level [36].

More is known about the role of AS in reprogramming. Regulators of AS are important in reprogramming to pluripotency and promote major changes in splicing [24–26]. Other examples in mice show that isoform switching in TFs impacts reprogramming [37, 38]. Similar to the sequential activation of distinct transcriptional programs during hematopoietic differentiation [7, 39], sequential phases of reprogramming and AS are found in reprogramming to pluripotency [25, 38, 40–43]. These studies support the notion that isoform usage can impact reprogramming and may provide higher resolution for discriminating cell states compared to gene expression alone, consistent with preliminary long-read sequencing data showing that iHSCs express multiple isoforms of many key regulators and markers. Reflecting the higher resolution provided by long-read isoform-level analysis, transcript isoforms from long-read sequencing data can also improve clustering outcomes for single cell data [44] and are used here for

refining the capture of cell state.

A limitation of our study is that our cell sorting scheme using CD34 may prematurely exclude relevant cells from consideration. Furthermore, achieving high-depth single-cell cDNA sequencing with ONT platforms is more challenging than for short-read sequencing. A lower read depth requires careful processing of the data to account for technical variation when comparing between transcriptomes, and poses challenges for single-cell isoform discovery, which is still in its infancy. Our study also only captures one time point along a continuous process in which cells likely transit through many different states. Time series single-cell analyses will be essential for resolving the temporal dynamics of reprogramming. Finally, while they are highly similar to native HSCs transcriptomically, iHSCs remain untested for differentiation and engraftment potential. Despite these limitations, our findings contribute to the broader effort to advance cellular reprogramming techniques that can generate autologous HSCs.

4 Methods

4.1 Cell Culture

All cells were cultured at 37 °C with 5% CO_2. Neonatal foreskin fibroblasts (BJ, ATCC CRL-2522, RRID:CVCL_3653) were used as the initial cell type at an initial passage number between 2 and 4, and were at 10 or fewer passages for endpoints. For reprogramming cultures, plates were coated with 0.1% gelatin. Fibroblasts were cultured in high glucose DMEM containing sodium pyruvate + 10% fetal bovine serum + 1% penicillin/streptomycin + 1% L-glutamine. Lentiviral transductions were carried out in complete medium without penicillin/streptomycin in the presence of 8 μg/mL polybrene (Millipore Sigma TR-1003-G). Integrated TF expression cassettes were induced with 1 μg/mL doxycycline hyclate (Dox, Millipore Sigma D9891) supplemented in the culture medium for 48 hours. Reprogramming cells were switched to complete MyeloCult H5100 (STEMCELL Technologies 05150) with 10e-6 M hydrocortisone (Millipore Sigma #H0888) and 1% penicillin/streptomycin 48 hours after the initial dox treatment. Normal human primary human bone marrow $CD34^+$ cells were purchased from ATCC (PCS-800-012TM Lot 80516705) and cultured in complete Myelocult H5100 for 12 - 16 hours prior to analysis. Adherent cells were dissociated using 1x TrypLE Express Enzyme (Gibco #12604-021).

4.2 Lentiviral Vector Design

Human GATA2, GFI1B, FOS, STAT5A, and REL plasmids were procured from the DNASU repository, subcloned into pLIX-403 Tet ON lentiviral backbones (RRID:Addgene_41395), and verified by Sanger sequencing. For TF combinations labeled FH3 (GATA2, GFI1B, FOS) and FH6 (GATA2, GFI1B, FOS, STAT5A, REL), lentiviral vectors were prepared individually for each plasmid, then pooled for transductions. For recipe FH4a (GATA2, GFI1B, FOS, STAT5A, REL), plasmids were pooled in equimolar amounts followed by lentiviral packaging. FH3 and FH6 were used for flow experiments shown in Figure 1, while FH4a was used for transductions in all other experiments. AllPrep DNA/RNA/Protein Mini Kits (Qiagen 80004) were used to process cells for testing genome integration and protein and gene expression.

FH3	**FH4a**	**FH6**
GATA2	GATA2	GATA2
GFI1B	GFI1B	GFI1B
FOS	FOS	FOS
	REL	REL
	STAT5A	STAT5A

4.3 Flow Cytometry

Antibodies used for flow and/or cell sorting included anti-human CD90 PE (1:250, BioLegend Cat# 328109, RRID: AB_893442), CD34 Alexa Fluor 488 (1:250, BioLegend Cat# 343517, RRID: AB_1937204) or CD34 PE (1:250, BioLegend Cat# 343605, RRID:AB_1732033), and CD143/ACE (1:250, BioLegend Cat# 344207, RRID: AB_2783230). LIVE/DEAD® Fixable Yellow Dead Cell Stain Kit (0.8 μL per test, Invitrogen L34968) and DAPI were used as viability markers. Cells incubated with primary antibodies or LIVE/DEAD stain for 30 minutes at 4°C in Hank's balanced salt solution or phosphate buffered saline. UltraComp eBeads (Invitrogen #01-2222) and ArC Amine Reactive Beads (Invitrogen #A10628) were used for single color compensation controls. After gating for singlets and live cells, fluorescence-minus-one (FMO) controls and human primary $CD34^+$ bone marrow cells were used to aid gating for iHSCs. Flow cytometry data were collected on a Bio-Rad ZE5 Cell Analyzer and fluorescence activated cell sorting (FACS) was done using a MoFlo Astrios or a Bigfoot Spectral Cell Sorter. Veri-Cells CD34 PBMCs (Biolegend #426901 Lot B375717) were reconstituted following the manufacturer's protocol and used as an additional staining control during FACS. FlowJo v10.10.0 (Becton Dickinson) was used for flow cytometry analyses.

4.4 Single-Cell Transcriptomics

Full-length barcoded cDNA amplicons were obtained from single cells using the 10X Genomics Next GEM Single Cell 3' Kit (V3.1). Briefly, 20,000 sorted $CD34^+$ cells and 2,000 control fibroblasts were loaded onto the 10X Genomics Chromium Controller (X) to partition single cells into Gel Beads-In Emulsion (GEMs). cDNA was generated from polyadenylated mRNA transcripts, incorporating a unique molecular identifier (UMI) and a cell-specific barcode for each transcript. Barcoded cDNA was pooled, purified with SILANE magnetic beads (10X Genomics #PN-2000048), and amplified by PCR. Barcoded cDNA libraries were prepared for sequencing according to the Oxford Nanopore Technologies (ONT) 3' cDNA protocols (Table 1). Briefly, cDNA amplicons were biotin tagged by PCR with the primers shown in Table 1. Biotinylated cDNA was purified with Dynabeads™ M-280 Streptavidin beads (Invitrogen #11205D), followed by secondary PCR amplification and Agencourt AMPure XP bead clean-up (Beckman Coulter #A63881). Products were assessed for quality following ONT recommendations. Long-read cDNA libraries were prepared for sequencing as in Table 1 and sequenced on an ONT GridION or PromethION Solo 2 (P2) sequencer. Raw reads were base-called with `guppy` using the High Accuracy basecalling model and stored as fastq files.

4.5 scRNA-seq Data Analysis

We created an analysis pipeline for long-read single-cell data quality control and analysis using `Snakemake` [45]. Fastq files were demultiplexed using `blaze2`

10x Genomics		
Next GEM Single Cell 3' Kit (V3.1)		
ONT		
Sequencer	**3' Protocol Version**	**Adapter Kit**
GridION	SST_V9148_V111_REVB_12JAN2022	cDNA-PCR Sequencing Kit (SQK-PCS111)
PromethION P2 Solo	SST_V9198_V114_REVB_06DEC2023	Ligation Sequencing Kit V14 (SQK-LSK114)
Primers		
5' - [Biotin] -	CAGCACTTGCCTGTCGCTCTATCTTCCTACACGACGCTCTTCCGATCT	- 3'
5' -	CAGCTTTCTGTTGGTGCTGATATTGCAAGCAGTGGTATCAACGCAGAG	- 3'

Table 1: Protocols and primers used for single cell barcoding and ONT sequencing.

[46] with parameters `--expect-cells 10000` and `--full-bc-whitelist` set to: `3M-february-2018.txt`. Reads were aligned to GRCh38 [47] using `minimap2` [48] with the following parameters: `-ax splice -uf --secondary=no --MD`. Raw count matrices were produced using the HTSeq `htseq-count` [49]. Count matrices were stored as `anndata` [50, 51] objects and analyzed in `Python` using the library `scanpy` [20]. Single-cell expression was filtered for cells expressing more than 200 genes. Genes expressed in fewer than 3 cells were removed. We normalized reads counts to `target_sum=1e4` using `normalize_total()` and logarithmized expression values using `log1p()`. Raw counts were retained for downstream analysis.

4.6 Analysis of Publicly Available Data

Single-cell data from hematopoietic stem cells (HSCs) and related cell types were sourced from [7] and [9]. Fibroblast data were obtained from [21]. Expression data were structured as raw count matrices using the `scanpy` package [20] and the `anndata` framework [50, 51]. Original annotations from the data sources were retained. Raw counts were concatenated into a single dataset, including induced HSC (iHSC) cell profiles. Unique cell barcodes and data source labels were assigned to maintain dataset integrity. Gene annotations for protein-coding genes were taken from the GTF file for GRCh38 [47].

Cells and genes were filtered from raw count data based on minimum expression thresholds (`min_cells=3`) and cell counts (`min_genes=100`). Quality control metrics were calculated, followed by normalization (`target_sum=1e4`) and log-transformation of the data using the `scanpy` package [20]. Batch correction was conducted using the ComBat method [52] as implemented in `scanpy` [20]. The processed AnnData object, containing the raw, normalized, and batch-corrected data, along with annotations, were stored for downstream analysis.

4.7 Isoform Identification

We performed isoform identification and quantification of known isoforms using `IsoQuant`, a tool designed for genome-based analysis of long RNA reads

[27]. Briefly, the isoform quantification workflow contains the following steps: (1) construction of an annotation database from an input GTF file, (2) read-assignment to annotated isoforms based on their intron and exon structure.

The following parameters were used with `IsoQuant`: `--reference=GRCh38 --count_exons --data_type` `` `nanopore' `` `--complete_genedb --bam_tags` `` `CB,UB,RD' ``. BAM files were generated using `minimap2` [48] in splice mode, as previously described.

5 Author contributions

LAM and IR initiated and supervised the overall project goals, designed the experiments. WM, LAM, and JC processed cells and tissues, and collected sequencing data. CMS, LAM, and IR, conducted formal data analysis and interpretation and created visualizations of the data. LAM and IR supervised use of computational and bioinformatics methods that were assembled and executed by CMS. IR and LAM provided resources for the project. All authors participated in development of methodologies and writing and editing the manuscript.

6 Acknowledgements

We thank Max Wicha at the University of Michigan for valuable discussions conversations and ideas on how to present results. We thank the University of Michigan Advanced Genomics Core for assistance with 10X Genomics single cell barcoding and the Flow Cytometry Core for assistance with cell sorting. We acknowledge BioRender, which was used to assemble figures and for generating the graphical abstract. This work was supported by the National Science Foundation (NSF) award number 2035827 (IR) subcontracted from iReprogram, Inc., the Air Force Office of Scientific Research (AFOSR) award number FA9550-22-1-0215 (IR), the Defense Advanced Research Projects Agency (DARPA) award number HR00112490472 (IR and LM), iReprogram, Inc, and Oxford Nanopore Technologies (ONT). We thank Lakmal Jayasinghe and the research and support teams at ONT for helpful discussions and technical assistance.

7 Code Availability

Source code for this study is available at:
https://github.com/CooperStansbury/ONT-single-cell.

8 Data Availability

Sequencing data generated in this study will be deposited in the Gene Expression Omnibus at publication and is available upon request.

9 Conflicts of Interest Statement

IR and LM are founders of and shareholders in iReprogram, Inc., which may benefit financially from the subject matter or materials discussed in this manuscript.

References

[1] H. Joachim Deeg. "Reward from half a match". In: *Blood* 141.25 (2023).

[2] Amy E. DeZern et al. "Alternative donor BMT with posttransplant cyclophosphamide as initial therapy for acquired severe aplastic anemia". In: *Blood* 141.25 (2023).

[3] Karen Ballen et al. "Impact of Race and Ethnicity on Outcomes After Umbilical Cord Blood Transplantation". In: *Transplantation and Cellular Therapy* (2024).

[4] Haofei Wang et al. "Direct cell reprogramming: approaches, mechanisms and progress". In: *Nature Reviews Molecular Cell Biology* 22.6 (2021).

[5] Caleb Weinreb et al. "Lineage tracing on transcriptional landscapes links state to fate during differentiation". In: *Science* 367.6479 (2020).

[6] Sam Watcham, Iwo Kucinski, and Berthold Gottgens. "New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing". In: *Blood* 133.13 (2019).

[7] Danilo Pellin et al. "A comprehensive single cell transcriptional landscape of human hematopoietic progenitors". In: *Nature Communications* 10.1 (2019).

[8] Anna Maria Ranzoni et al. "Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis". In: *Cell Stem Cell* 28.3 (2021).

[9] Chen Weng et al. "Deciphering cell states and genealogies of human haematopoiesis". In: *Nature* 627.8003 (2024).

[10] Kenichi Horisawa and Atsushi Suzuki. "Direct cell-fate conversion of somatic cells: Toward regenerative medicine and industries". In: *Proceedings of the Japan Academy, Series B* 96.4 (2020).

[11] Eva Szabo et al. "Direct conversion of human fibroblasts to multilineage blood progenitors". In: *Nature* 468.7323 (2010).

[12] Michael G. Daniel et al. "Making a Hematopoietic Stem Cell". In: *Trends in Cell Biology* 26.3 (2016).

[13] Andreia M. Gomes et al. "Cooperative Transcription Factor Induction Mediates Hemogenic Reprogramming". In: *Cell Reports* 25.10 (2018).

[14] Rita Silvério-Alves et al. "Hemogenic Reprogramming of Human Fibroblasts by Enforced Expression of Transcription Factors". In: *JoVE (Journal of Visualized Experiments)* 153 (2019).

[15] Michael G. Daniel et al. "Induction of human hemogenesis in adult fibroblasts by defined factors and hematopoietic coculture". In: *FEBS Letters* 593.23 (2019).

[16] Scott Ronquist et al. "Algorithm for cellular reprogramming". In: *Proceedings of the National Academy of Sciences* 114.45 (2017).

[17] Haiming Chen et al. "Functional organization of the human 4D Nucleome". In: *Proceedings of the National Academy of Sciences* 112.26 (2015).

[18] Oscar Franzén, Li-Ming Gan, and Johan L M Björkegren. "PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data". In: *Database* 2019 (2019).

[19] V. A. Traag, L. Waltman, and N. J. van Eck. "From Louvain to Leiden: guaranteeing well-connected communities". In: *Scientific Reports* 9.1 (2019).

[20] F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. "SCANPY: large-scale single-cell gene expression data analysis". In: *Genome Biology* 19.1 (2018).

[21] THE TABULA SAPIENS CONSORTIUM. "The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans". In: *Science* 376.6594 (2022).

[22] Fang Wang et al. "Single-cell architecture and functional requirement of alternative splicing during hematopoietic stem cell formation". In: *Science Advances* 8.1 (2022).

[23] Laura Mincarelli et al. "Single-cell gene and isoform expression analysis reveals signatures of ageing in haematopoietic stem and progenitor cells". In: *Communications Biology* 6.1 (2023).

[24] Sho Ohta et al. "Global Splicing Pattern Reversion during Somatic Cell Reprogramming". In: *Cell Reports* 5.2 (2013).

[25] Alexander Kanitz et al. "Conserved regulation of RNA processing in somatic cell reprogramming". In: *BMC Genomics* 20.1 (2019).

[26] Claudia Vivori et al. "Dynamics of alternative splicing during somatic cell reprogramming reveals functions for RNA-binding proteins CPSF3, hnRNP UL1, and TIA1". In: *Genome Biology* 22.1 (2021).

[27] Andrey D. Prjibelski et al. "Accurate isoform discovery with IsoQuant using long reads". In: *Nature Biotechnology* 41.7 (2023).

[28] Kiran Batta et al. "Direct Reprogramming of Murine Fibroblasts to Hematopoietic Progenitor Cells". In: *Cell Reports* 9.5 (2014).

[29] Raphael Lis et al. "Conversion of adult endothelium to immunocompetent haematopoietic stem cells". In: *Nature* 545.7655 (2017).

[30] Michael G. Daniel, Ihor R. Lemischka, and Kateri Moore. "Converting cell fates: generating hematopoietic stem cells de novo via transcription factor reprogramming". In: *Annals of the New York Academy of Sciences* 1370.1 (2016).

[31] Carolina Guibentif et al. "Single-Cell Analysis Identifies Distinct Stages of Human Endothelial-to-Hematopoietic Transition". In: *Cell Reports* 19.1 (2017).

[32] Yang Zeng et al. “Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing”. In: *Cell Research* 29.11 (2019).

[33] Zhengqi Wang and Kevin D Bunting. “STAT5 in hematopoietic stem cell biology and transplantation”. In: *JAK-STAT* 2.4 (2013).

[34] Madison R. Smith, Lisa R. Forbes Satter, and Alexander Vargas-Hernández. “STAT5b: A master regulator of key biological pathways”. In: *Frontiers in Immunology* 13 (2023).

[35] M. C. Turco et al. “NF-κB/Rel-mediated regulation of apoptosis in hematologic malignancies and normal hematopoietic progenitors”. In: *Leukemia* 18.1 (2004).

[36] Charlotte J. Wright, Christopher W. J. Smith, and Chris D. Jiggins. “Alternative splicing as a source of phenotypic diversity”. In: *Nature Reviews Genetics* 23.11 (2022).

[37] Hong Han et al. “MBNL proteins repress ES-cell-specific alternative splicing and reprogramming”. In: *Nature* 498.7453 (2013).

[38] Cheng-Xu Delon Toh et al. “RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming”. In: *Cell Reports* 15.12 (2016).

[39] Grigorios Georgolopoulos et al. “Discrete regulatory modules instruct hematopoietic lineage commitment and differentiation”. In: *Nature Communications* 12.1 (2021).

[40] Benjamin Cieply et al. “Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins”. In: *Cell Reports* 15.2 (2016).

[41] Mihaela Zavolan and Alexander Kanitz. “RNA splicing and its connection with other regulatory layers in somatic cell reprogramming”. In: *Current Opinion in Cell Biology*. Cell Nucleus 52 (2018).

[42] Laurent David and Jose M. Polo. “Phases of reprogramming”. In: *Stem Cell Research* 12.3 (2014).

[43] Constantinos Chronis et al. “Cooperative Binding of Transcription Factors Orchestrates Reprogramming”. In: *Cell* 168.3 (2017).

[44] Dafni A. Glinos et al. “Transcriptome variation in human tissues revealed by long-read sequencing”. In: *Nature* 608.7922 (2022).

[45] Felix Mölder et al. “Sustainable data analysis with Snakemake”. In: *F1000Research* 10 (2021).

[46] Yupei You et al. “Identification of cell barcodes from long-read single-cell RNA-seq with BLAZE”. In: *Genome Biology* 24.1 (2023).

[47] Fergal J Martin et al. “Ensembl 2023”. In: *Nucleic Acids Research* 51.D1 (2023).

[48] Heng Li. “Minimap2: pairwise alignment for nucleotide sequences”. In: *Bioinformatics* 34.18 (2018).

[49] Givanna H Putri et al. “Analysing high-throughput sequencing data in Python with HTSeq 2.0”. In: *Bioinformatics* 38.10 (2022).

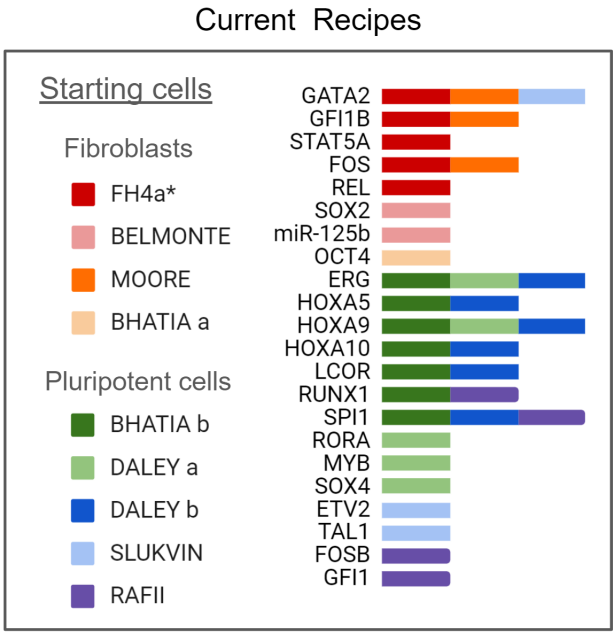
[50] Isaac Virshup et al. *anndata: Annotated data*. 2021.

[51] Isaac Virshup et al. “The scverse project provides a computational ecosystem for single-cell omics data analysis”. In: *Nature Biotechnology* 41.5 (2023).

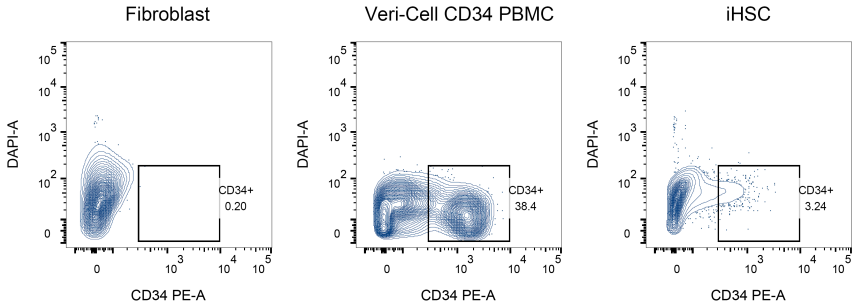
[52] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. “Adjusting batch effects in microarray expression data using empirical Bayes methods”. In: *Biostatistics* 8.1 (2007).

10 Supplemental Material

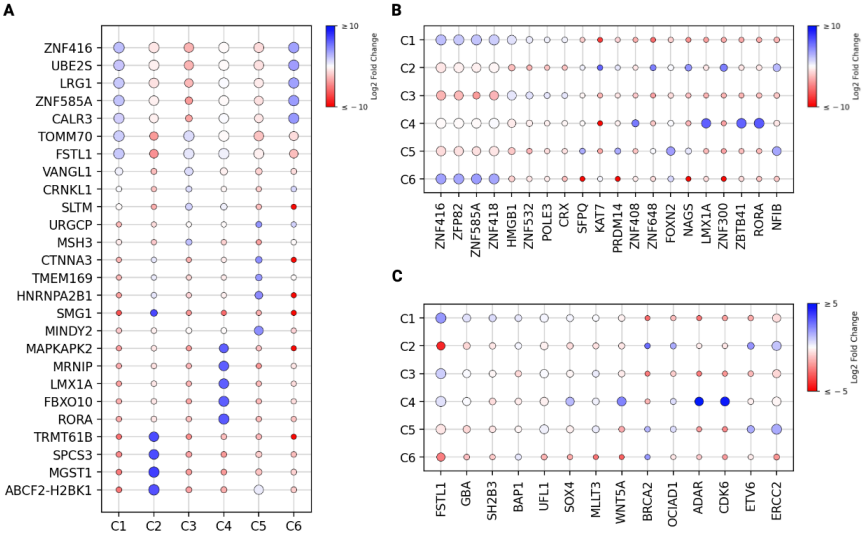
10.1 Supplementary Figures



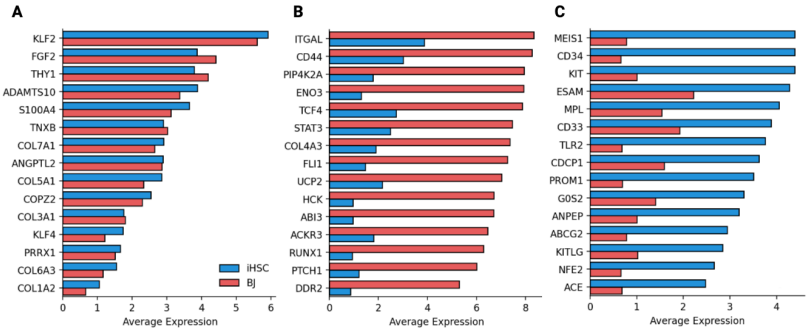
Supplementary Figure 1: **Transcription factor recipes used in reprogramming human cells into HSCs**. Left: Recipes named by group and divided by the starting cell type. Right: Transcription factors used in each recipe. *FH4a (and FH6) is the 5-TF recipe used throughout this manuscript.



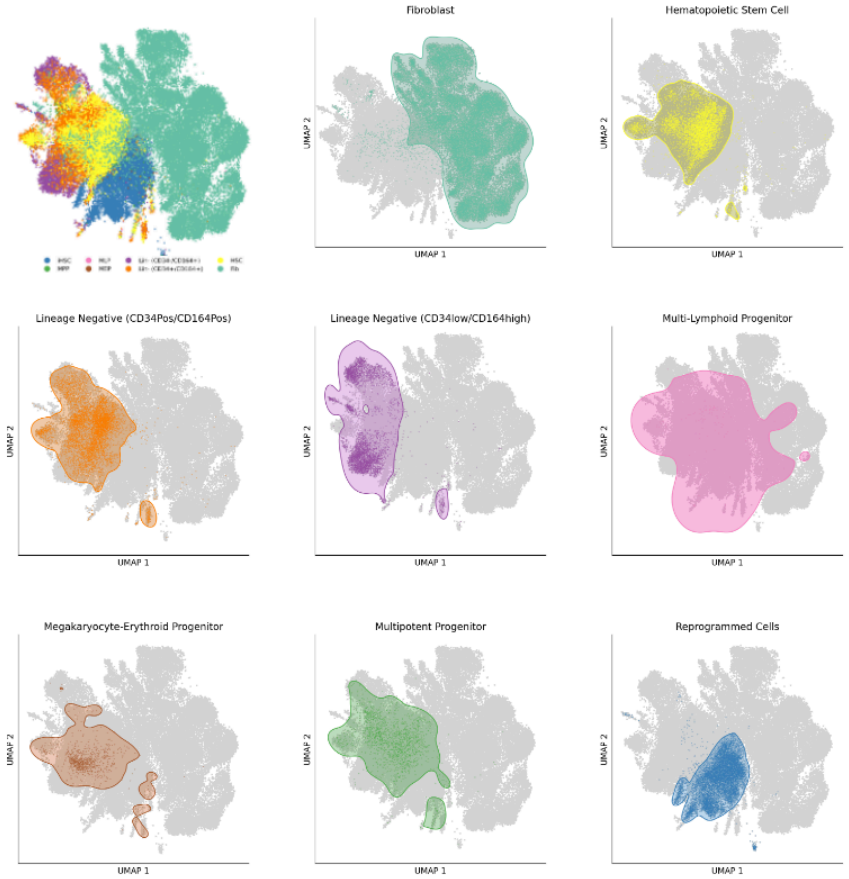
Supplementary Figure 2: **Flow cytometry of reprogramming cultures used for single cell RNA-seq**. Left: BJ fibroblast negative control. Middle: Veri-Cell CD34$^+$ PBMCs (Biolegend) positive/gating control. Right: reprogramming cultures 50 days after initial Dox treatment, showing CD34$^+$ iHSCs.



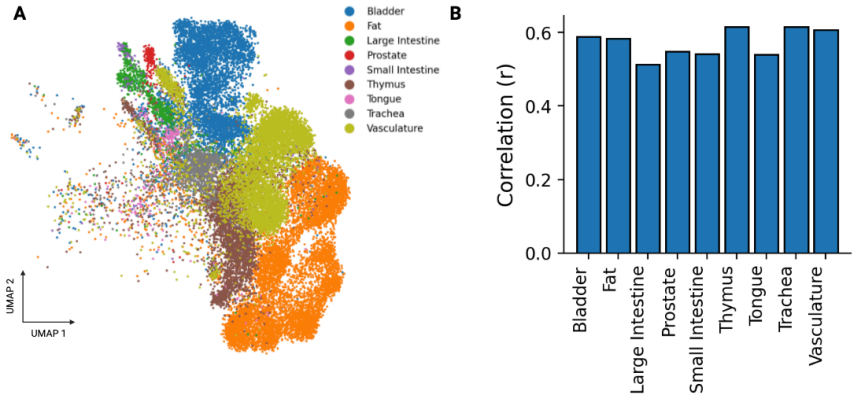
Supplementary Figure 3: **Differential Expression in iHSCs. (A)** The top 5 differentially expressed (adj. p-value ≤ 0.01 and log2 fold-change ≥ 1.0) genes per cluster. Marker size indicates the proportion (0, 1) of the cluster expressing the gene. **(B)** The top 4 differentially expressed (adj. p-value ≤ 0.01 and log2 fold-change ≥ 0.5) transcription factors per cluster. **(C)** The top 4 differentially expressed (adj. p-value ≤ 0.05 and log2 fold-change ≥ 0.25) HSC marker genes per cluster.



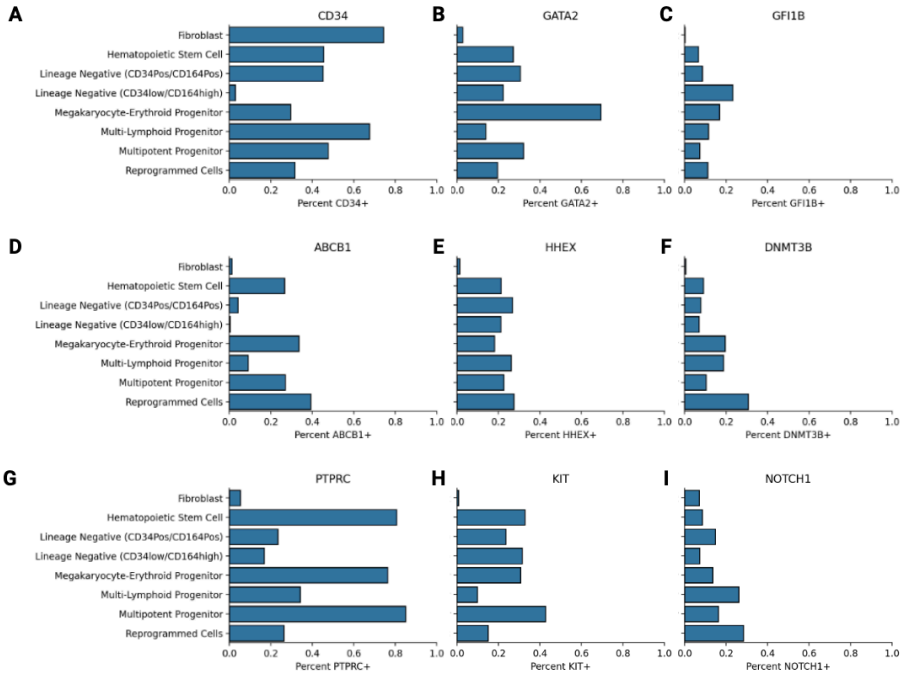
Supplementary Figure 4: **iHSC Expression Comparison with Population BJ Fibroblast. (A)** Average expression of bulk BJ fibroblast data from [17] and average expression of iHSC data for fibroblast genes. Top 15 genes are chosen based on minimum absolute difference between iHSC and BJ mean expression. **(B)** Top 15 fibroblast genes with high mean expression in BJ bulk data vs. iHSC mean expression. **(C)** Top 15 HSC genes with high expression in iHSC data vs. BJ bulk data.



Supplementary Figure 5: **Cell Types in the Hematopoietic Landscape.** UMAP coordinates of single-cells colored by broad cell type.

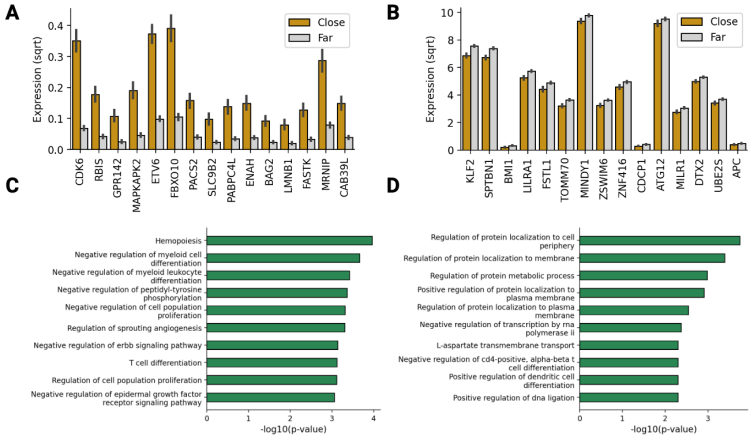


Supplementary Figure 6: **Comparison of Adult Fibroblasts with BJ Fibroblasts**. **(A)** Tissue sources of adult fibroblasts from [21] in UMAP coordinates. **(B)** Correlation of BJ fibroblast expression with adult fibroblast expression. Transcriptome-wide correlation between fibroblasts of each tissue type from [21] and BJ fibroblast from [17].

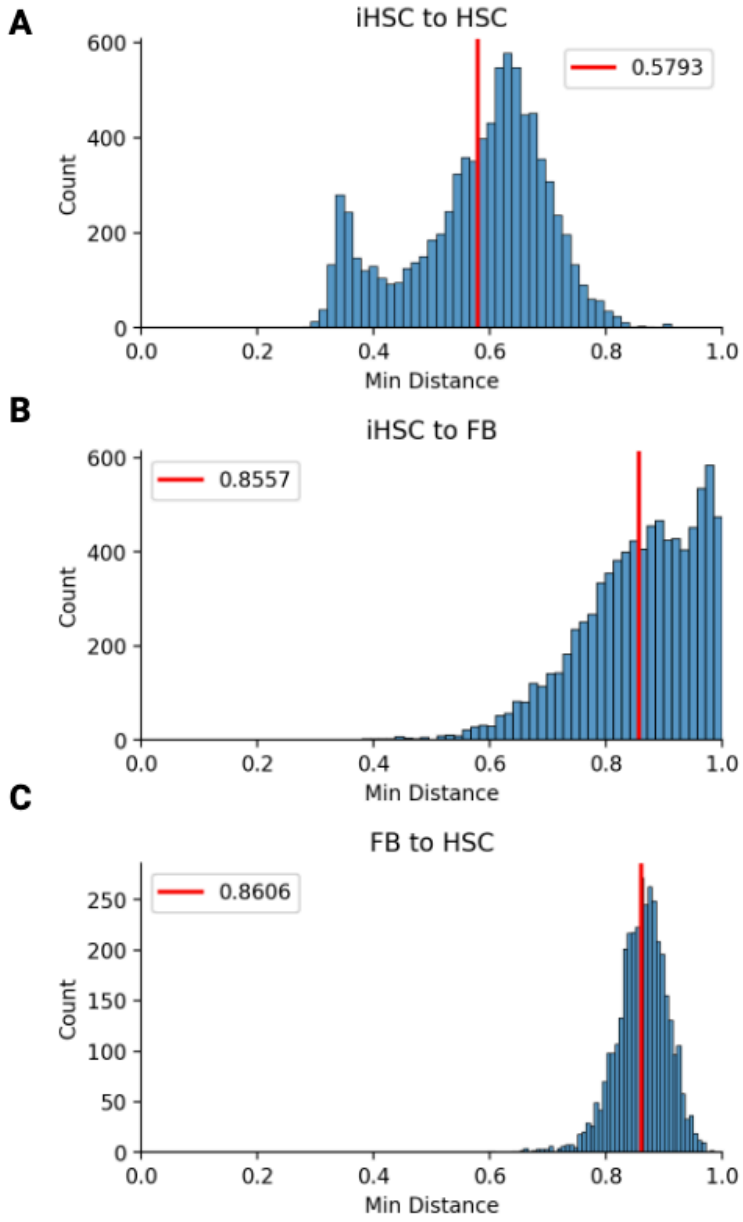


Supplementary Figure 7: **Percentage of Cell Types Expressing Key HSC Marker Genes. (A)** Percentage of cells expressing CD34 by type. **(B)** Percentage of cells expressing GATA2 by type. **(C)** Percentage of cells expressing GFI1B by type. **(D)** Percentage of cells expressing ABCB1 by type. **(E)** Percentage of cells expressing HHEX by type. **(F)** Percentage of cells expressing DNMT3B by type. **(G)** Percentage of cells expressing PTPRC by type. **(H)** Percentage of cells expressing KIT by type. **(I)** Percentage of cells expressing NOTCH1 by type.

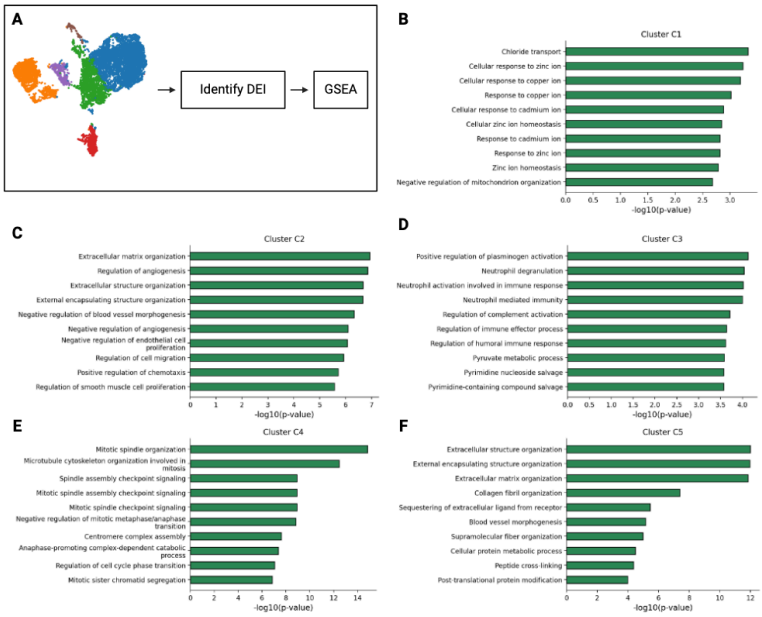
Supplementary Figure 8: **Most and Least Similar iHSC to HSC.** Comparison of the most (top) and least (bottom) similar iHSC profiles to HSCs from [9]. The most similar cells are 50 randomly sampled iHSC profiles with cosine distance less than 0.5 to native HSC profiles. The least similar cells are 50 randomly sampled iHSC profiles with cosine distance greater than 0.8 to native HSC profiles.



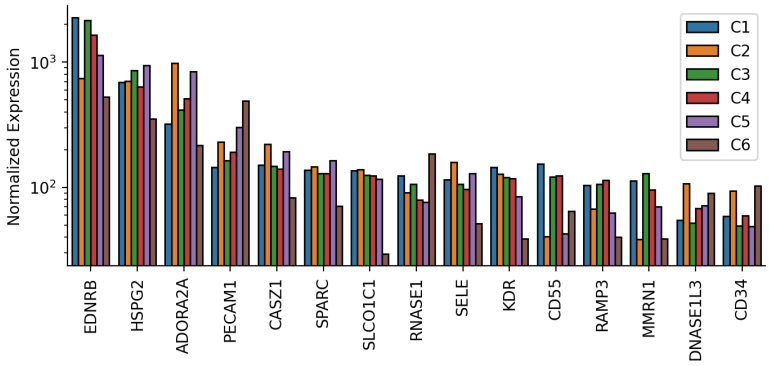
Supplementary Figure 9: **Differential expression analysis of optimally reprogrammed cells. (A)** Top 15 DEG ($\alpha = 0.05$, FC ≥ 0.25) between cells closest to HSC signatures from [9] vs. cells furthest from HSC signatures by cosine distance. Genes are unregulated in most similar iHSC profiles relative to least similar profiles. **(B)**Top 15 DEG ($\alpha = 0.05$, FC ≥ 0.25) between cells closest to HSC signatures from [9] vs. cells furthest from HSC signatures by cosine distance. Genes are unregulated in least similar iHSC profiles relative to most similar profiles. **(C)** Gene ontology enrichment analysis for the top 20 DEGs ($\alpha = 0.05$, FC ≥ 0.25) for closest iHSCs. **(D)** Gene ontology enrichment analysis for the top 20 DEGs ($\alpha = 0.05$, FC ≥ 0.25) for furthest iHSCs.



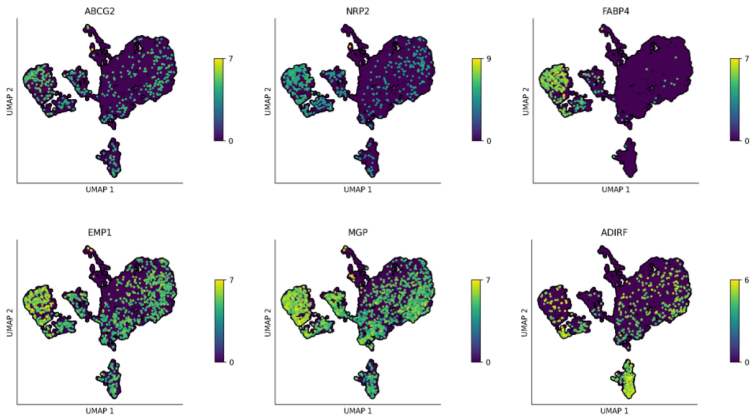
Supplementary Figure 10: **Transcriptome-wide single-cell minimum distance distributions.** **(A)** The distribution of minimum cosine distances between all iHSC cells and native HSC cells from [9]. Distances are computed over the shared genes. Expression was normalized to CPM. **(B)** The distribution of minimum cosine distances between all iHSC cells and adult fibroblast cells from [21]. **(C)** The distribution of minimum cosine distances between all native HSC cells from [9] and adult fibroblast cells from [21].



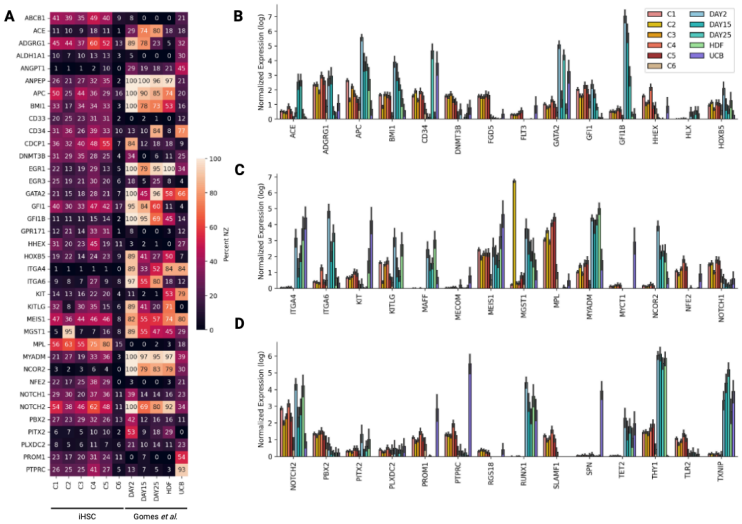
Supplementary Figure 11: **Enrichment analysis of differentially expressed isoforms. (A)** Differentially expressed isoforms were computed for iHSC clusters, excluding cluster C6. Enrichment analysis against the the Gene Ontology 'biological process' hierarchy was performed using the genes corresponding to the top 50 differentially expressed isoforms, where genes were sorted by their log2 fold change ($p-$value ≤ 0.05 Wilcoxon-rank sum test). **(B)** Pathway results for cluster C1. **(C)** Pathway results for cluster C2. **(D)** Pathway results for cluster C3. **(E)** Pathway results for cluster C4. **(F)** Pathway results for cluster C5.



Supplementary Figure 12: **Mean expression of endothelial marker genes in iHSC clusters.** Expression of endothelial marker genes by cluster in iHSC. Expression is was sorted based on the mean value and the top 15 genes are shown.



Supplementary Figure 13: **UMAP plot of endothelial marker gene expression.** Expression of endothelial marker genes shown on iHSCs in UMAP coordinates. Colorbars represent log-normalized expression values.



Supplementary Figure 14: **Comparison to Gomes *et al.*** **(A)** Percentage of cells with non-zero expression for HSC marker genes in iHSCs by cluster and reprogramming stages from [13]. **(B - D)** Mean expression by cluster of HSC marker genes for iHSC cells and single-cells from [13]

10.2 Supplementary Tables

Sample	Total Reads	Total Bases	Read N50	Base Quality
Gridr1	11,030,820	8,119,964,837	751	12.48
Gridr2	13,534,992	9,876,965,416	747	12.14
Gridr3	9,599,263	6,880,813,324	730	11.88
Gridr4	9,213,117	6,313,606,949	703	11.81
P2r5	29,881,286	19,901,617,912	688	14.40
P2r6	23,412,245	15,719,109,142	694	14.80
P2r7	23,576,025	15,494,282,065	674	14.49
Run6x0426	730,345	424,665,479	605	12.61
Run5x0314	66,824	55,488,659	872	10.77
Run1x0312	87,233	67,651,211	805	10.81
Run6x0314	69,623	57,776,834	869	10.80
Run7x0426	44,175	25,704,825	603	12.55
Run2x0313	77,992	64,362,375	866	10.73
Run7x0315	1,703,510	1,429,772,292	882	13.96
Run2x0416	797,533	573,911,854	750	12.64
Run8x0426	254,923	145,402,360	592	10.58
Run4x0314	72,658	60,805,557	877	10.89
Run3x0314	68,552	57,371,647	879	10.86
Run1x0415	286,350	201,657,420	718	13.05
Run4x0422	951,594	573,321,872	631	12.76
Run9x0503	486,740	280,702,344	591	10.24
Run3x0418	644,957	389,159,440	629	12.50

Supplementary Table 2: **iHSC Sequencing Summary**

Cluster	Cell Count	Proportion
C1	3927	0.458601
C2	1752	0.204601
C3	1708	0.199463
C4	602	0.070302
C5	447	0.052201
C6	127	0.014831

Supplementary Table 3: **Number of Cells by Cluster.** The count and and proportion of cells by cluster.

Cluster	Gene	LFC	Adj. P-value	Prop. Expressing
C1	KLF2	0.829	0.0	1.0
C1	FSTL1	2.021	0.0	0.998
C1	P4HA1	0.006	0.0	0.987
C2	P4HA1	0.251	0.0	0.977
C2	KLF6	0.706	0.0	0.815
C2	PALLD	4.532	0.0	0.672
C3	KLF2	0.702	0.0	0.989
C3	FSTL1	0.92	0.0	0.951
C3	TCF21	0.451	0.0	0.917
C4	TREML4	0.471	0.0	0.796
C4	NR4A2	0.227	0.028	0.789
C4	PAMR1	0.334	0.0	0.679
C5	P4HA1	0.521	0.0	0.993
C5	KLF6	1.271	0.0	0.969
C5	NR4A2	0.659	0.0	0.877

Supplementary Table 4: **Differential Expression of Fibroblast Genes.**

Cluster	CD34$^-$/ TF$^-$	Cell Count
C1	False	3411
C1	True	516
C2	False	1571
C2	True	181
C3	False	1415
C3	True	293
C4	False	509
C4	True	93
C5	False	374
C5	True	73
C6	False	101
C6	True	26

Supplementary Table 5: **Number of CD34$^-$/ TF$^-$Cells by Cluster.** CD34$^-$cells have no CD34 transcripts. Similarly, TF$^-$have no transcripts for GATA2, GFI1B, FOS, REL or STAT5A. In addition, TF$^-$cells have no evidence of V5 tags present in the lentiviral vector construct.

Cluster	Gene	LFC	Adj. P-value	Prop. Expressing
C1	NOTCH2	0.513	0.0	0.541
C1	APC	0.918	0.0	0.5
C1	MEIS1	0.398	0.0	0.465
C2	MGST1	7.464	0.0	0.947
C2	MPL	0.542	0.0	0.633
C2	ADGRG1	0.314	0.001	0.44
C3	BMI1	0.238	0.009	0.344
C3	KITLG	0.265	0.007	0.303
C4	MPL	0.487	0.0	0.752
C4	NOTCH2	0.292	0.001	0.62
C4	ADGRG1	0.352	0.0	0.598
C5	MPL	0.808	0.0	0.803
C5	CDCP1	0.507	0.0	0.546
C5	THY1	0.443	0.007	0.394

Supplementary Table 6: **Differential Expression of HSC Genes.**

Cluster	Gene	Fold Change (Log2)	Adj. P-value	Prop. Expressing
C1	EDNRB	1.136	0.0	0.988
C1	CD55	1.069	0.0	0.602
C1	KDR	0.372	0.0	0.586
C2	ADORA2A	1.386	0.0	0.902
C2	CASZ1	0.496	0.0	0.648
C2	PECAM1	0.416	0.0	0.647
C3	EDNRB	0.453	0.0	0.946
C3	HSPG2	0.226	0.0	0.932
C3	MMRN1	0.516	0.0	0.553
C4	ADORA2A	0.351	0.0	0.939
C4	PECAM1	0.251	0.006	0.741
C4	CD55	0.341	0.0	0.64
C5	HSPG2	0.48	0.0	0.975
C5	ADORA2A	1.049	0.0	0.953
C5	PECAM1	1.059	0.0	0.875
C6	PECAM1	0.026	0.026	0.315

Supplementary Table 7: **Differential Expression of HSC Genes.**

Gene 1	Gene 2	Gene 3
CD34	GATA2	GFI1B
FOS	STAT5A	REL
CRHBP	HOPX	AVP
MLLT3	FTH1	BMP6
CDK6	ADGRG1	ETV6
SOX4	ABCB1	CDCP1
PTPRC	MEIS1	MYADM
NFE2	APC	BMI1
FGD5		

Supplementary Table 8: **HSC Marker Genes.**

Gene	n Cells	Percent
CD34	2,457	28.7
CD164	2,029	23.7
PTPRC	2,058	24.04
THY1	2,260	26.4
MLLT3	1,705	19.91
CRHBP	2,287	26.71
DUSP1	2,357	27.53

Supplementary Table 9: **Key Hematopoietic Transcripts Expression in iHSCs**. The number and proportion of cells expressing key hematopoietic genes.

Gene	n Cells	Percent	n Cells (V5)	Percent (V5)
GATA2	1,502	17.54	148	1.73
GFI1B	851	9.94	0	0.0
FOS	583	6.81	104	1.21
STAT5A	406	4.74	5	0.06
REL	451	5.27	16	0.19

Supplementary Table 10: **Endogenous Expression of Reprogramming TFs.**

Cell Type	Count
Fibroblasts (FB)	26,586
Induced Hematopoietic Stem Cells (iHSC)	8,574
Lineage Negative CD34 Positive CD164 Positive (LinNegCD34PosCD164Pos)	6,343
Hematopoietic Stem Cells (HSC)	5,798
Lineage Negative CD34 low CD164 high (LinNegCD34lowCD164high)	4,265
Multipotent Progenitors (MPP)	1,795
Megakaryocyte Progenitors (MKP)	863
Myeloid Lymphoid Progenitors (MLP)	123

Supplementary Table 11: **Counts of Cell Types in the Hematapoetic Landscape Dataset**