Single-cell Transcriptomic Landscape of Nucleated Cells in Umbilical Cord Blood

Yi Zhao[1,3,†], Xiao Li[1,†], Jingwan Wang[1], Ziyun Wan[1], Kai Gao[1], Gang Yi[4], Xie Wang[1], Jinghua Wu[1], Bingbing Fan[2], Wei Zhang[1], Fang Chen[1], Huanming Yang[1,5], Jian Wang[1,5], Xun Xu[1], Bin Li[1,4], Shiping Liu[1], Weihua Zhao[2], Yong Hou[1,*] and Xiao Liu[1,*]

[1]BGI-Shenzhen, Shenzhen 518083, China

[2]Shenzhen Second People's Hospital, First Affiliated Hospital of Shenzhen University, Shenzhen 518035, Guangdong Province, China.

[3]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China

[4]Shanghai Institute of Immunology, Shanghai JiaoTong University School of Medicine, Shanghai 200025, China; Department of Immunology and Microbiology, Shanghai JiaoTong University School of Medicine, Shanghai 200025, China

[5]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China

[†]Yi Zhao and Xiao Li contributed equally

[*]co-corresponding authors

ABSTRACT

Umbilical cord blood (UCB) transplant is a therapeutic option for both pediatric and adult patients with a variety of hematologic diseases such as several types of blood cancers, myeloproliferative disorders, genetic diseases, and metabolic disorders. However, the level of cellular heterogeneity and diversity of nucleated cells in the UCB has not yet been assessed in an unbiased and systemic fashion. In the current study, nucleated cells from UCB were subjected to single-cell RNA sequencing, a technology enabled simultaneous profiling of the gene expression signatures of thousands of cells, generating rich resources for further functional studies. Here, we report the transcriptomic maps of 19,052 UCB cells, covering 11 major cell types. Many of these cell types are comprised of distinct subpopulations, including distinct signatures in NK and NKT cell types in the UCB. Pseudotime ordering of nucleated red blood cells (NRBC) identifies wave-like activation and suppression of transcription regulators, leading to a polarized cellular state, which may reflect the NRBC maturation. Progenitor cells in the UBC also consist two subpopulations with divergent transcription programs activated, leading to specific cell-fate commitment. Collectively, we provide this comprehensive single-cell transcriptomic landscape and show that it can uncover previously unrecognized cell types, pathways and gene expression regulations that may contribute to the efficacy and outcome of UCB transplant, broadening the scope of research and clinical innovations.

KEY WORDS

Umbilical cord blood, Single-cell RNA-seq, Transcriptomics, Nucleated red blood

cell

INTRODUCTION

Human umbilical cord blood (UCB) is an excellent source of hematopoietic progenitor cells. It has been widely used for bone marrow reconstitution for decades[1, 2]. The progenitor cells contained in UCB are capable of regenerating the entire lympho-hematopoietic compartment in the host. The most notable advantage of UCB transplant is the low risk of developing graft-versus-host disease (GVHD), even when donor and recipient are partially mismatched[3]. The immune cells in cord blood are virtually free from external stimulant and infection and thus were in a relatively more naïve stage. Such immunological immaturity is the key to alleviate the severity of GVHD by decreasing the alloreactive potential of lymphocytes[2, 4]. These advantages collectively expand the clinical potential of UCB transplant in many cases including some fatal diseases. The major limitation of UCB transplant, however, is the limited and inconsistent cell dose. It has been shown that the success rate of engraftment was critically dependent on the number of nucleated cells in the donor UCB[4-6].

Although UCB is now widely used for important clinical applications, we know surprisingly little about its cellular and molecular characteristics. Especially, the composition of progenitor, lymphocyte and other nucleated cells that affect the reconstitution potency after UCB engraftment is poorly understood. Recent advances in single-cell transcriptomics technology enable the exploration of cellular heterogeneity and deduction of functional relevance[7, 8]. Single-cell RNA-seq (scRNA-seq) studies of human peripheral blood (PB) cells have revealed new insights

into immune cell composition and disease-related functional abnormalities[9-11]. Some works have been done about precursor cell, erythrocyte and certain T cell subtypes in single cell level, but only focus on the certain cell type or in mouse. sequencing the all cell type simultaneously may provide global information about regulation between cells[12-15]. However, scRNA-seq studies have not thoroughly characterized nucleated cells in UCB, despite their profound clinical significance. Thus, the purpose of the current study is to investigate the nucleated cells present in UCB and depict a landscape view of the cellular composition and their transcriptomes. Such key information will undoubtedly facilitate the clinical innovation to develop more efficient and cost-effective UCB transplant treatments.

RESULT

A single-cell transcription atlas of nucleated cells in umbilical cord blood

To acquire a cellulome map of UCB cells at single-cell resolution, we collected UCB from two donors and isolated nucleated cells for single-cell RNA-sequencing. We generated transcriptomics map of 8,947 and 10,105 single cells for the two UCB samples, sequenced to the median depth of 483,117 and 265,714 mapped reads per cell, respectively. After careful quality control and filtering by multiple criteria, single-cell data were adjusted using SVA package to minimize batch effect. As a result, we generated expression matrix robustly measuring 1,223 genes on average in each cell. Next, to identify cell populations based on their expression signatures, we computed 15 significant principle components and performed

unsupervised clustering using the Seurat pipeline. Cells were visualized in two-dimension space by t-distributed stochastic neighborhood embedding (t-SNE). To determine the unique cell subpopulations and the specific state of gene expression in UCB, we utilized the public single-cell transcriptomics dataset of peripheral blood cells for comparison. This PB dataset includes 12,000 single-cell profiles of peripheral blood mononuclear cells (PBMC) measuring 1,257 genes per cell on average, which are at comparable level with those of the UCB data.

A global view was generated to illustrate the landscape of cell composition in UCB. Eleven distinct cell populations were clustered based on their gene expression profiles in both UCB samples. Public data for PB cells were processed in parallel and clustered with UCB cells in the same dimensional reduction space (Fig. 1A). To examine the inter-individual variability and investigate whether we can capture the cellular heterogeneity of all cell subpopulations in the UCB, we compared the classification results from the two donors. The identified cell clusters were quite consistent, with all the 11 clusters found in both donors. No significant residual batch bias was observed between the two UCB samples (Supplementary Fig. 1). These results demonstrate the robustness of UCB cell classification in our data. Clusters of cells that express known markers of major immune cell types were assigned with their respective identities (Fig. 1B). All 9 major immune cell types and hematopoietic lineages found in PB were identified in UCB, while two cell types, namely nucleated red blood cell (NRBC) and granulocytes, were only present in the UCB samples. While it's well known that NRBC are universal in UCB and scarcely present in PB,

the discrepancy of granulocytes are mostly due to unidentical cell enrichment experiments (Methods). The abundance of the common cell types also varied in PB versus UCB, especially in the myeloid populations, suggesting a specific immunological capacity of UCB (Fig. 1C). Furthermore, the gene expression profiles of each common cell type in UCB showed various degrees of distinction compared to those in PB. A few differentially expressed genes in B cells, NK cells, CD14[+] monocytes and T cells illustrated the potential variations in innate immune signaling pathway activation (i.e. S100A8/9) and biological processes involves in cell proliferation, and survival (i.e. JUN/FOS pathway) at cell population level (Supplementary Fig. 1B). These differences at the population level provide a rich resource to complement transcriptomics analysis performed in bulk samples, as well as a guide to functional investigations.

Polarity of cord nucleated red blood cell

In mammal hematopoiesis, nucleated red blood cells (NRBCs) precursors, or erythroblast, undergo several developmental stages in the bone marrow and progressively decrease cellular volume and RNA content, while accumulating specific functional proteins such as hemoglobin[16, 17]. It has been known for decades that erythroblast exist in relative large numbers in cord blood[18-20]. However, little was known about whether such development processes exist in the cord blood or whether the cord blood erythroblast population was homogenous. In our dataset, we found that NRBCs constitute a significant proportion of the total UCB nucleated cells.

Interestingly, the NRBC in the UCB samples displayed pronounced polarity defined by the divergent expression of a gene repertoire. By ordering NRBCs with differential genes identified within the cluster, we deduced a pseudotime axis that suggests a gradual change of cellular state. Evidently, the NRBC formed a linear trajectory along the pseudotime axis with no significant branching, indicating that the cell polarity resulted from a continuous changes of gene expression (Fig. 2A). Next, we modeled gene expression along this trajectory to identify genes characterized by a wave-like pattern. The most prominent of these were the genes encoding surface markers and proteins critical to the function of red blood cells, such as CD47, CD36, hemoglobin and glycophorins (Fig. 2B)[21]. The CD47 molecule has long been considered as one of the cell surface markers of primitive erythrocytes[22]. Hemoglobin genes, in contrast, are highly expressed in the relatively mature form of the erythrocyte. Thus, the polarity observed most likely reflected the maturity state of the NRBCs. An intermediate cell state that bridges the naïve state (CD47 high) and the mature state (hemoglobin high) was also observed. This intermediate stage was characterized by the elevated expression of a set of genes including those encoding glycophorins, suggesting that the cells in this stage exerted a specific function, rather than just transient intermediates. Strikingly, several key transcriptional regulators of erythrocyte homeostasis, including GATA1/2 and BCL11A[23-25], also clearly exhibited divergent patterns along the pseudotime axis (Fig. 2C). GATA1 is a well-characterized transcription factor responsible for the activation of multiple hemoglobin encoding genes in erythroid ontogeny[26], while BCL11A is a

transcription factor silencing hemoglobin encoding genes[25]. Another example was CITED2 and SOX6, transcription factors recently characterized as signature molecules specifically expressed in mouse primitive and definitive erythroblasts respectively, showed similar specificity in the naïve and intermediate cellular states as defined by the pseudotime axis[27]. In addition, a gradual decrease in the numbers of RNA molecules (represented by UMI) and expressed genes across the pseudotime axis was observed, supporting the correlation between linear polarity and the cord blood NRBCs functionality (Fig. 2D). These lines of evidence further corroborated the polarity identified in the NRBC population in UCB, and strongly indicated that the differential activation of transcriptional programs was one of the underlining mechanisms.

Molecular signatures of UCB progenitor cell

A distinct progenitor population was found in the UCB that shared a similar transcriptome profile with the hematopoietic stem cells (HSCs) in the PB dataset. However, when the t-SNE clustering was performed with the progenitor population in a finer resolution, a secondary subpopulation emerged, demonstrating the heterogeneity of progenitor population in the UCB (Fig. 3A). One subpopulation of UCB progenitor cells overlapped with HSCs in PB and specifically expressed the canonical HSC marker genes such as CD34, SOX4 and CD135 (red dots, Fig. 3A), suggesting their identity as being cord blood HSCs. Interestingly, the other subpopulation consists cells only from the UCB (blue dots, Fig. 3A) and did not

express the HSC canonical markers (Fig. 3B, 3C) despite the similarity in overall gene expression spectrum, which drove the t-SNE clustering. Surprisingly, this CD34$^-$ UCB specific progenitor population highly expressed the myeloid lineage-specific gene MS4A3 (Fig. 3C), a known signature of granulocytic-monocytic progenitors (GMPs)[28]. GMPs give rise to mast cell progenitors (MCP) and basophil progenitors (BPC), which are found in the bone marrow, spleen and gastrointestinal mucosa[29]. Furthermore, FCER1A, the gene encoding the Fc fragment of the IgE receptor, which is a surface marker frequently used in cell sorting for mast cells[30], was highly expressed in the CD34$^-$ cell population; while CCR3, a sorting marker for basophils[31, 32], was co-expressed at a comparable level. Similarly, many genes that play regulatory roles in mast cell and basophil differentiation, exemplified by ITBG7 and CSF2RB[29, 33], were co-expressed at high level as well (Fig. 3C). The collective activation of gene repertoires critical in GMP-MCP and GMP-BPC ontogeny axes strongly suggested that these cells were bi-potent progenitors or intermediate cells, similar to the basophil/mast cell progenitor (BMCP) first verified in spleens of adult C57BL/6 mice[34]. High expression levels of GATA2 and median expression of CEBPA transcription factors were also consistent with the signatures of mouse BMCP [34-36] (Fig. 3C). A critical difference between the UCB subpopulation and the mouse BMCP is that CD34 expression was turned off, suggesting limited stemness in these cells. We thus hypothesized that these cells represent the intermediates before the bifurcation during basophil and mast cell differentiation and termed them umbilical intermediate bi-potent cells (uIBC).

A spectrum of differentially expressed genes was identified in the HSC and uIBC subpopulations in UCB (partly shown in Fig. 3B). Next, we asked whether the switch of cell identities resulted from alteration of transcriptional programing that governed the differentiation process. Transcription factor enrichment analysis utilizing the comprehensive database of Encode[37] and ChEA[38] was used to detect over-represented combinations of conserved transcription factor binding sites in a given set of genes. The analysis revealed that MYC, TAF1 and MAX were the mostly enriched for activating highly expressed genes found in the HSCs as compared to uIBC. These transcription factors are well known for their roles in proliferation and cell cycle control[39-41]. Conversely, RUNX1, SPI1 and GATA2 were ranked as the top enriched transcription factors for activating the highly expressed genes in the uIBCs (Fig. 3D). These transcription factors are conventionally considered as master regulators of differentiation of the myeloid lineage[35, 42, 43]. Such functional correlation was further corroborated by the mutually exclusive expression pattern of the top enriched factors. For example, high expression levels of MYC, TAF, MAX and PML, enriched for activating HSC feature genes, were detected in the HSCs; and *vice versa*, high expression levels of GATA2, RUNX1, SP1 and RELA, were detected in the uIBC (Fig. 3E). These data collectively supported the notion that the two subtypes of cells we found in the progenitor population in UCB were indeed functionally relevant.

Heterogeneity of natural killer T cell in UCB

Effective immune response against infection, allergy and cancer generally requires coordinated activation of both the innate and adaptive immune systems. Recent studies have shown that natural killer T (NKT) cells emerge as a bridge between innate and adaptive immunity in order to mediate an immune response[44]. Unlike conventional T cells, NKT cells exhibits distinct tissue specificity under homeostatic conditions, suggesting compartmentalized functions[45-48]. Given the immune immaturity of cord blood cells, we expected an unique pattern of the NKT population. We selected all NK cells and the T cells that express *NKG7* as well as the T cells nearby on the t-SNE space in Fig. 1A from PB and UCB datasets and performed further t-SNE clustering to systemically examine their potential heterogeneity. As a result, these cells clustered into five distinct subgroups in PB and six subgroups in UCB (Fig. 4A). By analyzing the expression of the lineage marker genes *CD3* and *KLRB1*, we assigned the cell clusters with T (CD3$^+$KLRB1$^-$), NKT (CD3$^+$KLRB1$^+$) and NK (CD3$^-$KLRB1$^+$) lineages[46, 49, 50] (Fig. 4A). Interestingly, apparent heterogeneity was observed in both T and NK cells. For example, the CD3$^+$KLRB1$^-$ T cells in both PB and UCB could be further sub-grouped into cytotoxic and non-cytotoxic populations, based on the expression of cytotoxic marker genes, such as *GNLY*, *GZMK*, *GZMB* and *PRF1*. Similarly, NK cells could also be sub-grouped into GZMK$^+$ and GZMB$^+$ populations, based on the mutually exclusive expression pattern of the two granzyme genes. However, no significant difference in gene expression was detected between UCB and PB NK cells, indicating that the GZMK$^+$ and GZMB$^+$ NK populations were both constitutive of general homeostasis.

NKT cells, originally clustered as parts of the T cell population in PB and UCB, express high levels of GZMB and PRF1 but not GZMK. Interestingly, we found a unique subpopulation of NKT cells (non-cytotoxic, light blue dots in Fig. 4B) in UCB that do not express the GNLY, GZMK, GZMB or PRF1 genes, suggesting the lack of cytotoxicity mediated by the granzyme and perforin pathway (Fig. 4B). NKT cells were previously reported to have tissue-specific gene expression programs that lead to diverse functions and were termed NKT1, NKT2 and NKT17, predominantly localized in liver, lung and peripheral lymph node, respectively[46, 51-54]. In our data, the expression profile of the GZMB+ NKT cell was mostly similar to that of the NKT1 type, highlighted by signature expression of IL2RB, KLRB1, ZBTB16 and TBX21 (Supplementary Fig. 2A), but these cells did not express GATA3[55]. The non-cytotoxic NKT cell was not a part of NKT2 and NKT17, because they exhibited the expression pattern of high level of KLRB1 and GATA3[55, 56], indicating a new subtype of NKT cell in UCB. We then considered whether the development stage that shapes the different clusters, but all these two NKT cells are belong to $CD24^-CD44^+KLRB1^+$ stage 3 (Fig. 2B, Supplementary Fig. 2A). In contrast, the non-cytotoxic NKT cells highly express a spectrum of surface molecules, such as IL7R, CD40LG and CD3E that may mediate interactions with other immune cell types to coordinate immune response (Fig. 4C). Gene ontology analysis further corroborated that the highly expressed genes of the non-cytotoxic NKT cells were enriched in lymphocyte cell-cell adhesion pathways and lymphocyte proliferation,

differentiation and activation processes (Fig. 4D). Thus, we conclude that the cell composition of NKT and other cytotoxic cells vary between PB and UCB.

DISCUSSION

For the first time, we present here a single-cell level transcriptomic landscape of nucleated cells in UCB. By analyzing the expression pattern of known marker genes, we identified UCB cells belonging to almost all of the major hematopoietic lineages in the PB, covering lymphoid, myeloid and hematopoietic progenitor cells. We also observed that certain cell populations were highly enriched in UCB cells, such as the NRBCs and the uIBCs. In adults, red blood cells are generated mainly in the bone marrow from nucleated cells identified as erythroid precursors. These cells undergo morphological changes through cell divisions and gradual decrease in cell size and RNA species, increase in chromatin condensation and hemoglobin protein accumulation. Such changes have been associated with the early stages of maturation of red blood cell. Here, in our dataset we also observed such a dynamic cellular state in a linear polarity, suggesting that erythroid precursors undergo a similar maturation process in the UCB as in the bone marrow. Progenitor cell populations in UCB also appeared to be a mixture of two distinct subpopulations. The uIBC, a unique subpopulation only seen in UCB but not PB, were identified with characteristics of both basophil and mast cell signatures. A similar bipotent population (BMCP) exists in mouse spleen and is capable of divergent development. Signature gene expression, including transcription factors and surface markers were remarkably similar between

BMCP and uIBC, except that uIBC lack the expression of the conventional progenitor marker CD34. Although uIBC and HSC in UCB were globally similar in their transcriptomic profiles, the lack of CD34 made it difficult to conclude whether these uIBCs were indeed progenitors or transient intermediates captured during UCB hematopoiesis. Functional validations will be necessary to determine the potential abilities of self-renewal and lineage regeneration of these cells and substantiate the similarity with mouse BMCP at the functional level. Next, we discovered that gene expression signatures often vary greatly in the common cell types seen in both UCB and PB. We interrogated the UCB single-cell data at a finer scale and discovered a previously unknown NKT population that may be unique to UCB. NKT cells have an essential role in bridging innate and adaptive immunity against infectious diseases and tumorigenesis. Thus, NKT cells have significant therapeutic values. UCB transplants have demonstrated remarkable effectiveness in treating many types of blood cancers. Adoptive transfer of the NKT cells has been tested in animal models[57, 58], and several clinical trials are in process to test the safety and efficiency of NKT cell transfer to harness the solid tumors in human[59-62]. The enhanced understanding of the NKT cell heterogeneity in UCB would benefit our selection of appropriate source and the activation of the cytotoxicity of NKT cells to target cancer and other diseases. Therefore, we speculated that a targeted enrichment, modulation or engineering of the existing NKT populations in the UCB could lead to considerable improvement in the efficacy of enhancing protective immune responses.

Taken together, our data provides the first single-cell transcriptomic references for UCB, which could be used as a standard dataset for comparative analysis. We expect that this dataset will prove useful in uncovering the novel molecular signatures that define the cellular heterogeneity in UCB and provide markers for targeted enrichment of certain cell types of interest to researchers in multiple fields. Our dataset should also be a rich resource to formulate hypothesis of signaling pathway activation, transcription control and other mechanistic studies in the field of functional immunology in single cell level Moreover, to our knowledge, the current study was the first to performed 10X Genomics single-cell transcriptome profiling in tandem with the rolling circle amplification-based BGISEQ-500 high throughput sequencing platform, broadening the technological diversity for single-cell RNA-seq applications.

METHODS

Sample collection

Two umbilical cord blood samples were collected from healthy donors immediately after caesarean section with informed consents. Samples were stored in EDTA anticoagulant tubes and transported to laboratory within 1 hour. CD45$^+$ and CD45$^-$ cells were isolated from 1 mL cord blood by positive and negative selection, respectively, using Whole Blood CD45 MicroBeads (Miltenyi,130-090-872) and Whole Blood Column Kit (Miltenyi, 130-093-545). Next, the CD45+ and CD45- cells were counted by hemocytometer and mixed at the ratio of 4 to 1. The cells were further gently pipetted into a single-cell suspension and diluted to 700 cell/µL concentration. The single cell gene expression dataset of peripheral blood from healthy donors were downloaded from the https://support.10xgenomics.com/single-cell-gene-expression/datasets (processed by Cell Ranger 2.0.1). Peripheral blood mononucleated cells were isolated and processed in 10X platform (10X Genomics, USA).

Library construction and sequencing

Single-cell suspension was loaded to Single-cell 3'Chips (10X Genomics, USA) and subjected to GemCode Single-Cell Instrument (10X Genomics, USA) to generate single-cell Gel Beads in Emulsion (GEMs), per manufacture's instruction. GEMs were next subjected to library construction by Chromium™ Single-cell 3' Reagent Kits v2 (10X Genomics, USA), steps of which included RT incubation,

cDNA amplification, fragmentation, end repair, A-tailing, adaptor ligation, and sample index PCR. However, such library was originally designed to be sequenced by the Illumina sequencing platform. In order to convert the libraries to that compatible with BGI-Seq 500 sequencer, we performed a 12 cycle PCR on the libraries with BGI-Seq adaptor primers, and subsequent DNA circularization, rolling-cycle amplification (RCA) to generate DNA Nano Balls (DNBs). The purified DNBs were sequenced by BGISEQ-500 sequencer, generating reads containing 16 bp of 10X™ Barcodes, 10 bp of unique molecular indices (UMI) and 100 bp of 3' cDNA sequences. Each library was sequenced in three lanes, yielding ~1.9 billion reads in total.

Alignment and initial processing of sequencing data

CellRanger toolkit (10X Genomics, USA, version 2.0.0) was employed to align the cDNA reads to GRCh38 transcriptome. Filtered UMI expression matrices of both samples were generated with the default parameter settings except "--force-cells=4000"[63].

UMI normalization and cell clustering

The expression matrices of both samples were normalized by "cellranger aggr" function from the CellRanger toolkit, with the parameter "--normalize=mapped". In order to remove potential technical bias introduced during library construction and sequencing, SVA package was used to minimize such batch effects. The Seurat

package was then utilized to quality-control expression data for each cell, first by removing the undetected gene (i.e. expression in less than 30 cell), secondly by removing the cells with less than 400 detected genes. Cells that yielded high level (>10%) of reads that aligned to mitochondria genes were also removed. Next, the filtered expression matrices were used for unsupervised cell-clustering by the Seurat package, adapting the typical pipeline that was recommended by the authors[64]. Subsequently, the identity of each cell cluster was annotated manually, based on the previously known markers and the cluster-specific genes listed by the Seurat "FindAllMarkers" function.

Pseudotime analysis

To infer the developmental polarity of nucleated red blood cells (NRBCs), pseudotime analysis was performed by the Monocle2 package (version 2.6.4). NRBCs were ordered according to the pseudotime deduced by the top 1,000 differentially expressed genes within the NRBCs.

AUTHOR CONTRIBUTION

X.LIU. and Y.H. jointly supervised research. Y.Z. and G.Y. designed the experiments. X.W., K.G., Y.Z. and X.Z. performed the experiments. Y.Z. and J.W. processed single-cell RNA sequencing data. Y.Z., X.LI., J.W. and Z.W. analyzed the data. W.Z. and B.F. collected the cord blood. X.LI. and Y.Z. wrote the manuscript. X.LIU revised the manuscript. All authors have reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

COMPETING INTERESTS

The authors declare no competing financial interests.

REFERENCES

1.   Kurtzberg J. Update on umbilical cord blood transplantation. Current opinion in

        pediatrics. 2009;21 1:22-9.

2.   Paloczi K. Immunophenotypic and functional characterization of human umbilical cord

        blood mononuclear cells. Leukemia. 1999;13 Suppl 1:S87-9.

3.   Park SK and Won JH. Usefulness of umbilical cord blood cells in era of hematopoiesis

        research. International journal of stem cells. 2009;2 2:90-6.

4.   Rocha V, Wagner JE, Jr., Sobocinski KA, Klein JP, Zhang MJ, Horowitz MM, et al.

        Graft-versus-host disease in children who have received a cord-blood or bone marrow

        transplant from an HLA-identical sibling. Eurocord and International Bone Marrow

        Transplant Registry Working Committee on Alternative Donor and Stem Cell Sources.

        The New England journal of medicine. 2000;342 25:1846-54.

        doi:10.1056/NEJM200006223422501.

5.   Laughlin MJ, Barker J, Bambach B, Koc ON, Rizzieri DA, Wagner JE, et al.

        Hematopoietic engraftment and survival in adult recipients of umbilical-cord blood

        from unrelated donors. The New England journal of medicine. 2001;344 24:1815-22.

        doi:10.1056/NEJM200106143442402.

6.   Migliaccio AR, Adamson JW, Stevens CE, Dobrila NL, Carrier CM and Rubinstein P.

        Cell dose and speed of engraftment in placental/umbilical cord blood transplantation:

        graft progenitor cell content is a better predictor than nucleated cell quantity. Blood.

bioRxiv preprint doi: https://doi.org/10.1101/346106; this version posted June 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2000;96 8:2717-22.

7.   Navin NE. The first five years of single-cell cancer genomics and beyond. Genome

research. 2015;25 10:1499-507. doi:10.1101/gr.191098.115.

8.   Tanay A and Regev A. Scaling single-cell genomics from phenomenology to

mechanism. Nature. 2017;541 7637:331-8. doi:10.1038/nature21350.

9.   Eltahla AA, Rizzetto S, Pirozyan MR, Betz-Stablein BD, Venturi V, Kedzierska K, et al.

Linking the T cell receptor to the single cell transcriptome in antigen-specific human T

cells. Immunology and cell biology. 2016;94 6:604-11. doi:10.1038/icb.2016.16.

10.   Proserpio V and Mahata B. Single-cell technologies to study the immune system.

Immunology. 2016;147 2:133-40. doi:10.1111/imm.12553.

11.   Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of Infiltrating T

Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell. 2017;169 7:1342-56

e16. doi:10.1016/j.cell.2017.05.035.

12.   Gaublomme JT, Yosef N, Lee Y, Gertner RS, Yang LV, Wu C, et al. Single-Cell

Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. Cell. 2015;163

6:1400-12. doi:10.1016/j.cell.2015.11.009.

13.   Patil VS, Madrigal A, Schmiedel BJ, Clarke J, O'Rourke P, de Silva AD, et al.

Precursors of human CD4(+) cytotoxic T lymphocytes identified by single-cell

transcriptome analysis. Science immunology. 2018;3 19

doi:10.1126/sciimmunol.aan8664.

14.   Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al.

Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell.

2015;163 7:1663-77. doi:10.1016/j.cell.2015.11.013.

15.   Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human

haematopoietic stem cell lineage commitment is a continuous process. Nature cell

biology. 2017;19 4:271-81. doi:10.1038/ncb3493.

16.   Migliaccio AR. Erythroblast enucleation. Haematologica. 2010;95 12:1985-8.

doi:10.3324/haematol.2010.033225.

17.   Ji P, Murata-Hori M and Lodish HF. Formation of mammalian erythrocytes: chromatin

condensation and enucleation. Trends in cell biology. 2011;21 7:409-15.

doi:10.1016/j.tcb.2011.04.003.

18.   Hebbar S, Misha M and Rai L. Significance of maternal and cord blood nucleated red

blood cell count in pregnancies complicated by preeclampsia. Journal of pregnancy.

2014;2014:496416. doi:10.1155/2014/496416.

19.   Hermansen MC. Nucleated red blood cells in the fetus and newborn. Archives of

disease in childhood Fetal and neonatal edition. 2001;84 3:F211-5.

20.   Merenstein GB, Blackmon LR and Kushner J. Nucleated red-cells in the newborn.

Lancet. 1970;1 7659:1293-4.

21.   van Schravendijk MR, Handunnetti SM, Barnwell JW and Howard RJ. Normal human

erythrocytes express CD36, an adhesion molecule of monocytes, platelets, and

endothelial cells. Blood. 1992;80 8:2105-14.

22.   Oldenborg PA, Zheleznyak A, Fang YF, Lagenaur CF, Gresham HD and Lindberg FP.

Role of CD47 as a marker of self on red blood cells. Science. 2000;288 5473:2051-4.

23.   Dore LC and Crispino JD. Transcription factor networks in erythroid cell and

megakaryocyte development. Blood. 2011;118 2:231-9.

doi:10.1182/blood-2011-04-285981.

24.   Bresnick EH, Hewitt KJ, Mehta C, Keles S, Paulson RF and Johnson KD. Mechanisms

of erythrocyte development and regeneration: implications for regenerative medicine

and beyond. Development. 2018;145 1 doi:10.1242/dev.151423.

25.   Liu N, Hargreaves VV, Zhu Q, Kurland JV, Hong J, Kim W, et al. Direct Promoter

Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell. 2018;173

2:430-42 e17. doi:10.1016/j.cell.2018.03.016.

26.   Ohneda K and Yamamoto M. Roles of hematopoietic transcription factors GATA-1

and GATA-2 in the development of red blood cell lineage. Acta haematologica.

2002;108 4:237-45. doi:10.1159/000065660.

27.   Kingsley PD, Greenfest-Allen E, Frame JM, Bushnell TP, Malik J, McGrath KE, et al.

Ontogeny of erythroid gene expression. Blood. 2013;121 6:e5-e13.

doi:10.1182/blood-2012-04-422394.

28.   Ishibashi T, Yokota T, Satoh Y, Ichii M, Sudo T, Doi Y, et al. Identification of MS4A3 as

a reliable marker for early myeloid differentiation in human hematopoiesis.

Biochemical and biophysical research communications. 2018;495 3:2338-43.

doi:10.1016/j.bbrc.2017.12.117.

29.   Iwasaki H and Akashi K. Myeloid lineage commitment from the hematopoietic stem

cell. Immunity. 2007;26 6:726-40. doi:10.1016/j.immuni.2007.06.004.

30.   Stone KD, Prussin C and Metcalfe DD. IgE, mast cells, basophils, and eosinophils.

The Journal of allergy and clinical immunology. 2010;125 2 Suppl 2:S73-80.

doi:10.1016/j.jaci.2009.11.017.

31.   Hausmann OV, Gentinetta T, Fux M, Ducrest S, Pichler WJ and Dahinden CA. Robust

expression of CCR3 as a single basophil selection marker in flow cytometry. Allergy.

2011;66 1:85-91. doi:10.1111/j.1398-9995.2010.02431.x.

32.   Chirumbolo S, Ortolani R and Vella A. CCR3 as a single selection marker compared to

CD123/HLADR to isolate basophils in flow cytometry: some comments. Cytometry

Part A : the journal of the International Society for Analytical Cytology. 2011;79

2:102-6. doi:10.1002/cyto.a.21008.

33.   Gurish MF, Tao H, Abonia JP, Arya A, Friend DS, Parker CM, et al. Intestinal mast cell

progenitors require CD49dbeta7 (alpha4beta7 integrin) for tissue-specific homing. The

Journal of experimental medicine. 2001;194 9:1243-52.

34.   Arinobu Y, Iwasaki H, Gurish MF, Mizuno S, Shigematsu H, Ozawa H, et al.

Developmental checkpoints of the basophil/mast cell lineages in adult murine

hematopoiesis. Proceedings of the National Academy of Sciences of the United

States of America. 2005;102 50:18105-10. doi:10.1073/pnas.0509148102.

35.   Iwasaki H, Mizuno S, Arinobu Y, Ozawa H, Mori Y, Shigematsu H, et al. The order of

expression of transcription factors directs hierarchical specification of hematopoietic

lineages. Genes & development. 2006;20 21:3010-21. doi:10.1101/gad.1493506.

36.   Iwasaki H, Mizuno S, Mayfield R, Shigematsu H, Arinobu Y, Seed B, et al.

Identification of eosinophil lineage-committed progenitors in the murine bone marrow.

The Journal of experimental medicine. 2005;201 12:1891-7.

doi:10.1084/jem.20050548.

37.	Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489 7414:57-74. doi:10.1038/nature11247.

38.	Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR and Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics. 2010;26 19:2438-44. doi:10.1093/bioinformatics/btq466.

39.	Trop-Steinberg S and Azar Y. Is Myc an Important Biomarker? Myc Expression in Immune Disorders and Cancer. The American journal of the medical sciences. 2018;355 1:67-75. doi:10.1016/j.amjms.2017.06.007.

40.	Amati B, Littlewood TD, Evan GI and Land H. The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. The EMBO journal. 1993;12 13:5083-7.

41.	Lin CY, Tuan J, Scalia P, Bui T and Comai L. The cell cycle regulatory factor TAF1 stimulates ribosomal DNA transcription by binding to the activator UBF. Current biology : CB. 2002;12 24:2142-6.

42.	North TE, Stacy T, Matheny CJ, Speck NA and de Bruijn MF. Runx1 is expressed in adult mouse hematopoietic stem cells and differentiating myeloid and lymphoid cells, but not in maturing erythroid cells. Stem cells. 2004;22 2:158-68. doi:10.1634/stemcells.22-2-158.

43.	Dakic A, Metcalf D, Di Rago L, Mifsud S, Wu L and Nutt SL. PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. The Journal of experimental medicine. 2005;201 9:1487-502. doi:10.1084/jem.20050075.

44.	Van Kaer L, Parekh VV and Wu L. Invariant natural killer T cells: bridging innate and

adaptive immunity. Cell and tissue research. 2011;343 1:43-55.

doi:10.1007/s00441-010-1023-3.

45.   Bendelac A, Savage PB and Teyton L. The biology of NKT cells. Annual review of

immunology. 2007;25:297-336. doi:10.1146/annurev.immunol.25.022106.141711.

46.   Godfrey DI, MacDonald HR, Kronenberg M, Smyth MJ and Van Kaer L. NKT cells:

what's in a name? Nature reviews Immunology. 2004;4 3:231-7. doi:10.1038/nri1309.

47.   Kronenberg M. Toward an understanding of NKT cell biology: progress and paradoxes.

Annual review of immunology. 2005;23:877-900.

doi:10.1146/annurev.immunol.23.021704.115742.

48.   Van Kaer L. NKT cells: T lymphocytes with innate effector functions. Current opinion in

immunology. 2007;19 3:354-64. doi:10.1016/j.coi.2007.03.001.

49.   Van Der Vliet HJ, Nishi N, Koezuka Y, Peyrat MA, Von Blomberg BM, Van Den

Eertwegh AJ, et al. Effects of alpha-galactosylceramide (KRN7000), interleukin-12

and interleukin-7 on phenotype and cytokine profile of human Valpha24+ Vbeta11+ T

cells. Immunology. 1999;98 4:557-63.

50.   Vivier E and Anfossi N. Inhibitory NK-cell receptors on T cells: witness of the past,

actors of the future. Nature reviews Immunology. 2004;4 3:190-8.

doi:10.1038/nri1306.

51.   Gumperz JE, Miyake S, Yamamura T and Brenner MB. Functionally distinct subsets of

CD1d-restricted natural killer T cells revealed by CD1d tetramer staining. The Journal

of experimental medicine. 2002;195 5:625-36.

52.   Lee PT, Benlagha K, Teyton L and Bendelac A. Distinct functional lineages of human

V(alpha)24 natural killer T cells. The Journal of experimental medicine. 2002;195
5:637-41.

53.   Coquet JM, Chakravarti S, Kyparissoudis K, McNab FW, Pitt LA, McKenzie BS, et al.

Diverse cytokine production by NKT cell subsets and identification of an

IL-17-producing CD4-NK1.1- NKT cell population. Proceedings of the National

Academy of Sciences of the United States of America. 2008;105 32:11287-92.

doi:10.1073/pnas.0801631105.

54.   Michel ML, Keller AC, Paget C, Fujio M, Trottein F, Savage PB, et al. Identification of

an IL-17-producing NK1.1(neg) iNKT cell population involved in airway neutrophilia.

The Journal of experimental medicine. 2007;204 5:995-1001.

doi:10.1084/jem.20061551.

55.   Brennan PJ, Brigl M and Brenner MB. Invariant natural killer T cells: an innate

activation scheme linked to diverse effector functions. Nature reviews Immunology.

2013;13 2:101-17. doi:10.1038/nri3369.

56.   Constantinides MG and Bendelac A. Transcriptional regulation of the NKT cell lineage.

Current opinion in immunology. 2013;25 2:161-7. doi:10.1016/j.coi.2013.01.003.

57.   Lam PY, Nissen MD and Mattarollo SR. Invariant Natural Killer T Cells in Immune

Regulation of Blood Cancers: Harnessing Their Potential in Immunotherapies.

Frontiers in immunology. 2017;8:1355. doi:10.3389/fimmu.2017.01355.

58.   Bagnara D, Ibatici A, Corselli M, Sessarego N, Tenca C, De Santanna A, et al.

Adoptive immunotherapy mediated by ex vivo expanded natural killer T cells against

CD1d-expressing lymphoid neoplasms. Haematologica. 2009;94 7:967-74.

doi:10.3324/haematol.2008.001339.

59. Exley MA, Friedlander P, Alatrakchi N, Vriend L, Yue S, Sasada T, et al. Adoptive

Transfer of Invariant NKT Cells as Immunotherapy for Advanced Melanoma: A Phase

I Clinical Trial. Clinical cancer research : an official journal of the American Association

for Cancer Research. 2017;23 14:3510-9. doi:10.1158/1078-0432.CCR-16-0600.

60. Motohashi S, Ishikawa A, Ishikawa E, Otsuji M, Iizasa T, Hanaoka H, et al. A phase I

study of in vitro expanded natural killer T cells in patients with advanced and recurrent

non-small cell lung cancer. Clinical cancer research : an official journal of the

American Association for Cancer Research. 2006;12 20 Pt 1:6079-86.

doi:10.1158/1078-0432.CCR-06-0114.

61. Kunii N, Horiguchi S, Motohashi S, Yamamoto H, Ueno N, Yamamoto S, et al.

Combination therapy of in vitro-expanded natural killer T cells and

alpha-galactosylceramide-pulsed antigen-presenting cells in patients with recurrent

head and neck carcinoma. Cancer science. 2009;100 6:1092-8.

doi:10.1111/j.1349-7006.2009.01135.x.

62. Yamasaki K, Horiguchi S, Kurosaki M, Kunii N, Nagato K, Hanaoka H, et al. Induction

of NKT cell-specific immune responses in cancer tissues after NKT cell-targeted

adoptive immunotherapy. Clinical immunology. 2011;138 3:255-65.

doi:10.1016/j.clim.2010.11.014.

63. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively

parallel digital transcriptional profiling of single cells. Nature communications.

2017;8:14049. doi:10.1038/ncomms14049.

64.   Satija R, Farrell JA, Gennert D, Schier AF and Regev A. Spatial reconstruction of

single-cell gene expression data. Nature biotechnology. 2015;33 5:495-502.

doi:10.1038/nbt.3192.

FIGURE LEGEND

Figure 1: Different cell types and proportion in PB and UCB.

A. Two-dimensional t-SNE distributions of UCB cells and downloaded PB cells in the same space and each cluster was assigned with their respective identities according the known markers; UCB cells were colorized on the left and PB cells were colorized on the right; All clusters were dyed at the same color according to the corresponding cell types.
B. Scaled average gene expression of the major known markers (rows) detected in different cell types in UCB (U) and PB (P) (column).
C. Cell type proportions of PB1, PB2, UCB1 and UCB2, which dyed using the same color in Figure 1A

Figure 2: NRBCs in the UCB samples displayed a pronounced polarity.
A. The ordering of NRBCs along pseudotime in a two-dimensional space determined by Monocle2. Each dot represents a single cell, and the color represents the order of pseudotime.
B. Expression of genes that encoding surface markers and proteins critical to the function of red blood cells in NRBCs ordered by pseudotime of three states: Primitive stage(top), Intermediate stage(middle), Mature stage(bottom).
C. Expression of transcriptional regulators in NRBCs ordered by pseudotime time of three states: Primitive stage(top), Intermediate stage(middle), Mature stage(bottom).
D. Detected UMI numbers and Gene numbers in each cell ordered by pseudotime, each dot represents a cell.

Figure 3: Progenitor cells in UCB samples displayed heterogeneous molecular signatures.
A. The zoomed t-SNE projection of progenitor cells from UCB and PB samples, showing the formation of two main clusters shown in different colors, blue: uIBC and red: HSC. The functional description of each cluster is determined by the gene expression characteristics of each cluster.
B. Heatmap of progenitor cell clusters with the differentially expressed signature genes. Cells along lateral axis were ranked by t-SNE1 in figure 3A. The panel above the heatmap identifies the different clusters.
C. Violin plots show the single-cell gene expression pattern of two clusters, blue: uIBC and red: HSC.
D. Transcription factor enrichment analysis of the two clusters using the HSC signature genes (469, top left) and the uIBC signature genes (450, bottom left) revealed enriched transcription factors of HSC (top middle) and uIBC (bottom

middle) and corresponding enrichment score (-log(FDR), top right and bottom right).

E. Violin plots show the single-cell transcription factors pattern of two clusters, blue: uIBC and red: HSC.

Figure 4: NKT cells and NK cells showed functional heterogeneity in UCB and PB.

A. Two-dimensional t-SNE distributions of all NK cells and the T cells that express NKG7 gene from PB and UCB datasets. Color represents five distinct subgroups (GZMB$^+$ NKT, GZMK$^+$ NK, GZMB$^+$ NK, Non-cytotoxic T, Cytotoxic T) in PB (top) is and six subgroups (GZMB$^+$ NKT, GZMK$^+$ NK, GZMB$^+$ NK, Non-cytotoxic T, Cytotoxic T, Non-cytotoxic NKT) in UCB (bottom) according corresponding cytotoxic signature genes.

B. Violin plots show the single-cell gene expression pattern of signature genes between the distinct subgroups in UCB (right) and PB (left). The shapes represent the distribution of cells based on the scaled expression. Colors represent the different subsets.

C. Violin plots show the scaled single-cell expression of indicated differentially expressed genes between GZMB$^+$ NKT and Non-cytotoxic NKT subsets in UCB.

D. Gene ontology analysis of GZMB$^+$ NKT and Non-cytotoxic NKT subsets in UCB using the differentially expressed genes. X-axis represents the enriched gene percentage in certain pathways. Y-axis shows the enriched pathways ordered by gene percentage. The size of dot represents the gene numbers. Color scale shows the adjusted p-values of each enriched pathway.

Supplementary Figure 1: Sample distribution and gene expression differentiation between UCB and PB.

A. Sample distribution in the same two-dimensional t-SNE space as Figure 1A. Each dot represents a single cell. Color represents the different samples in UCB and PB.

B. Violin plots show the scaled single-cell expression of indicated differentially expressed genes between UCB and PB in B cell, CD14$^+$ Monocyte, NK cell and T cell.

Supplementary Figure 2: Cell subsets express distinct set of transcripts in NK and NKT.

A. Violin plots show the scaled single-cell expression of indicated differentially expressed genes between GZMB$^+$ NK and GZMK$^+$ NK in PB (top) and UCB (bottom).

B. Violin plots show the scaled single-cell expression of indicated differentially expressed genes between GZMB$^+$ NKT and Non-cytotoxic NKT subsets in UCB.
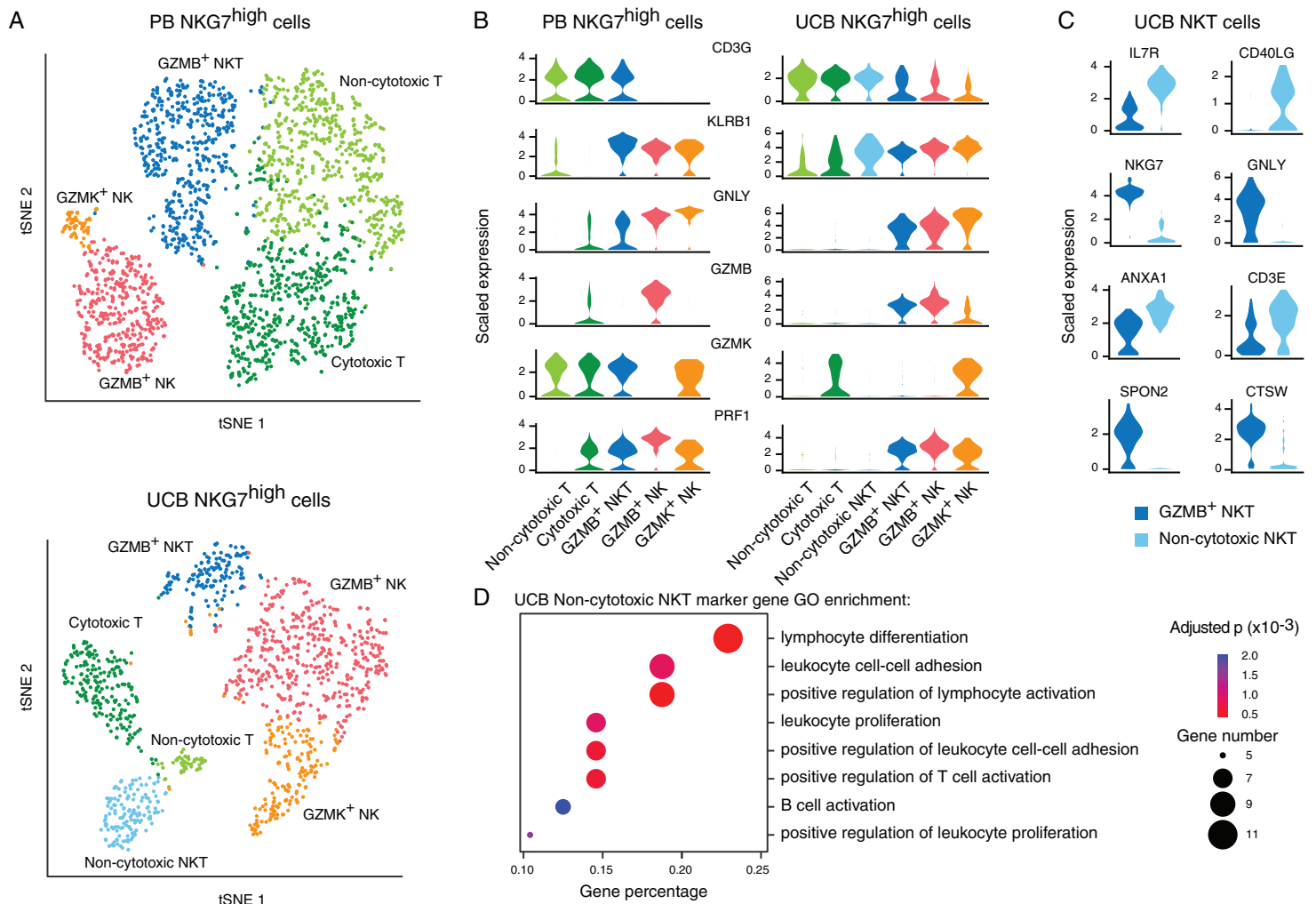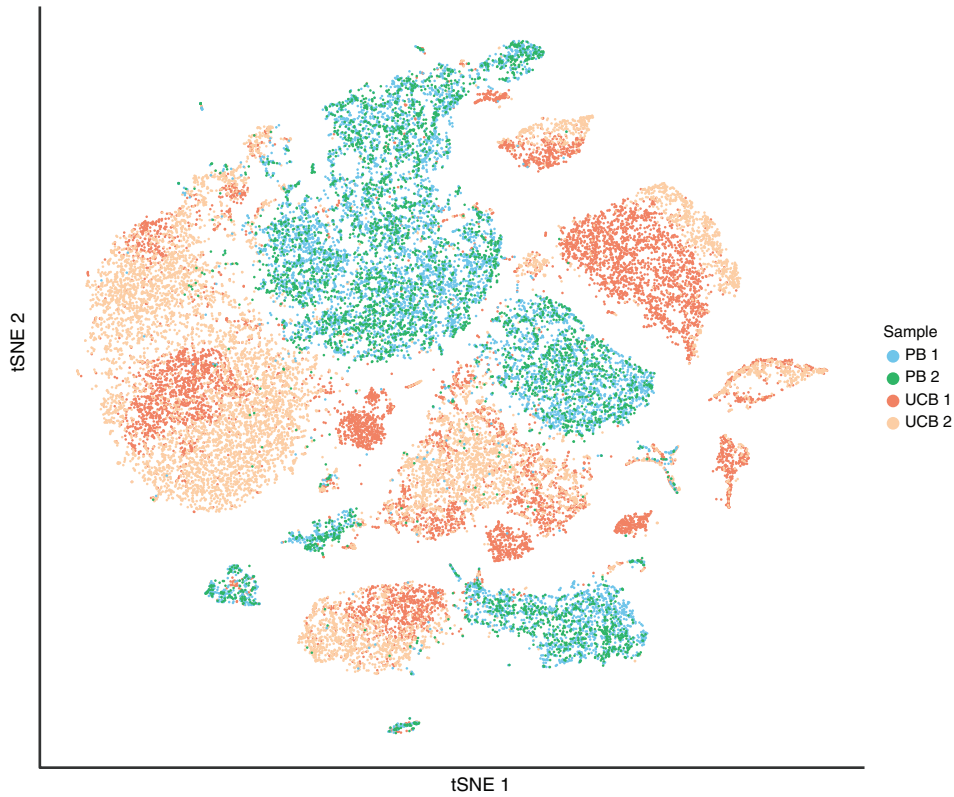
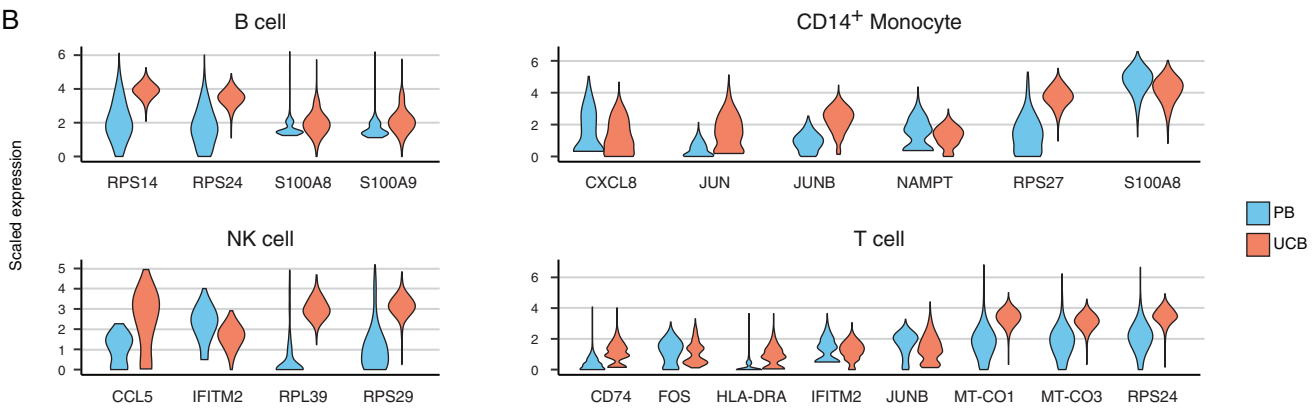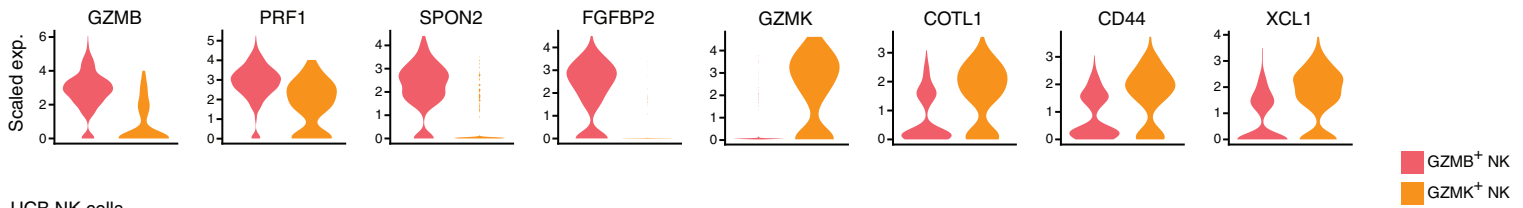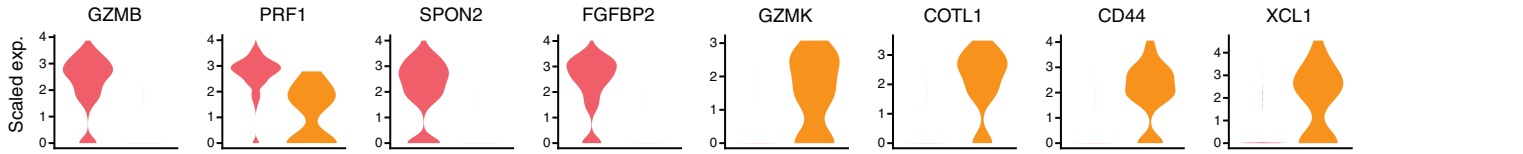Figure 1
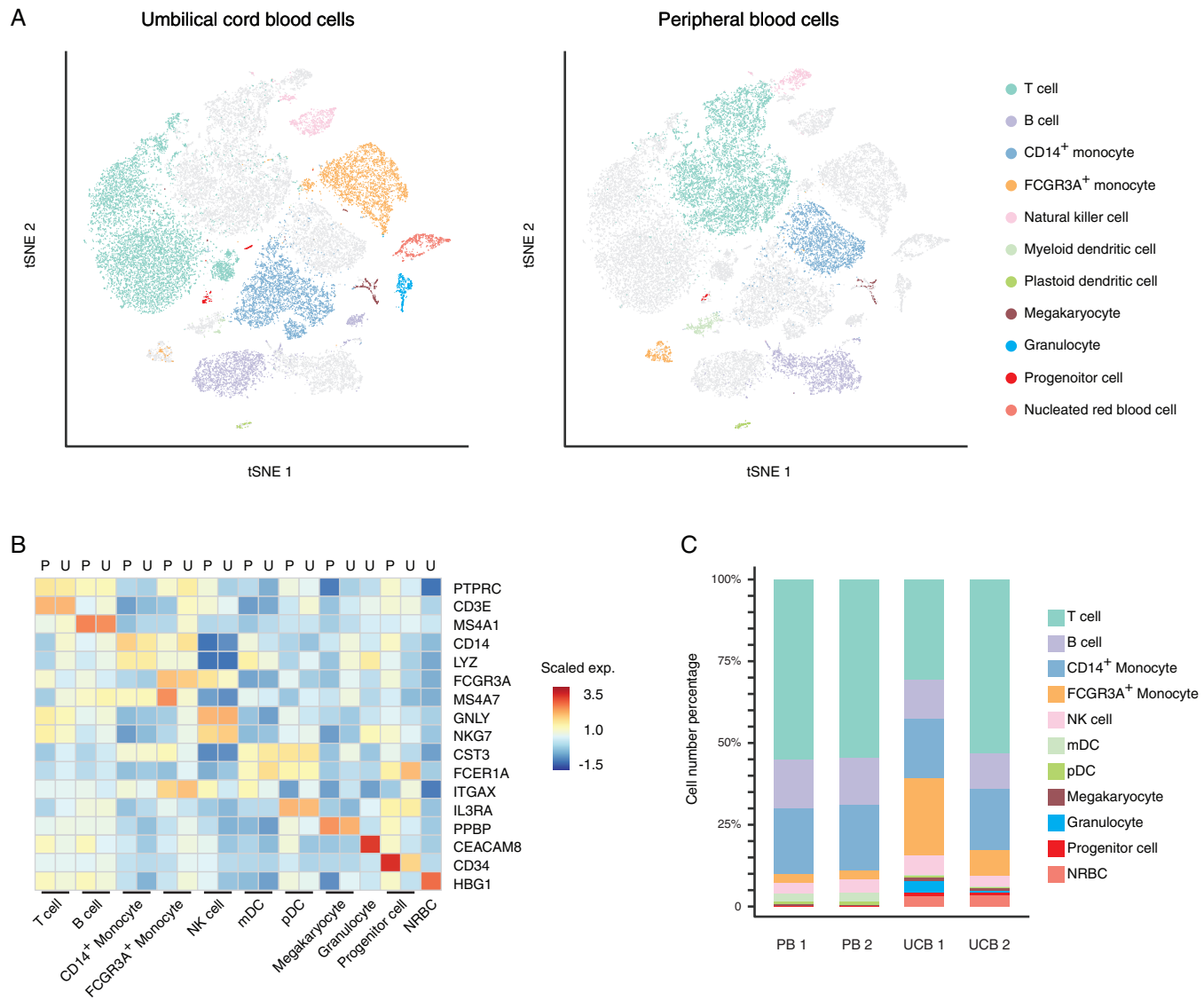
Figure 2
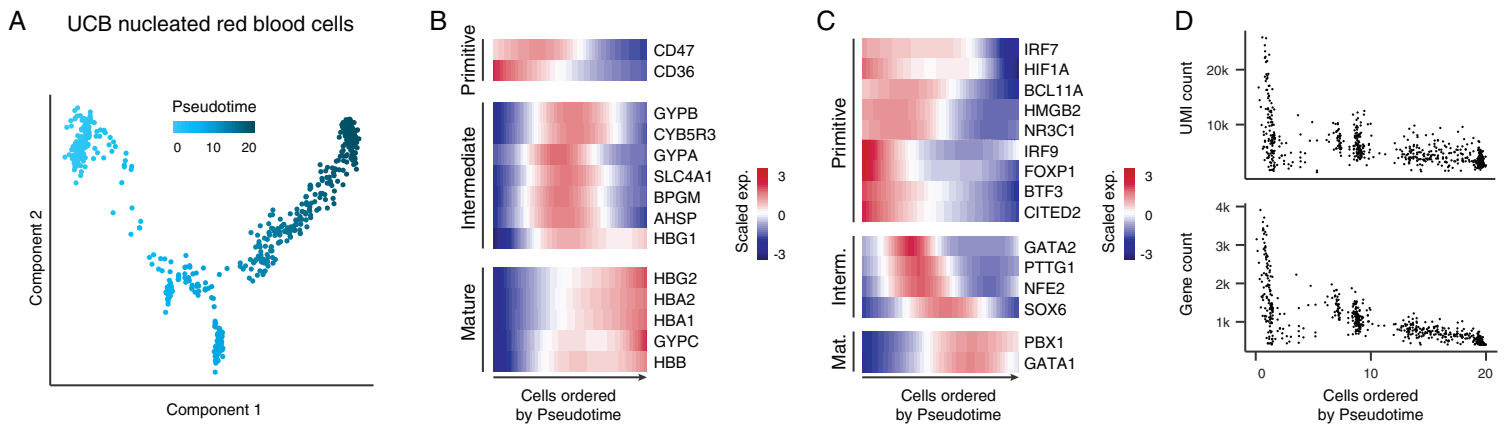
Figure 3
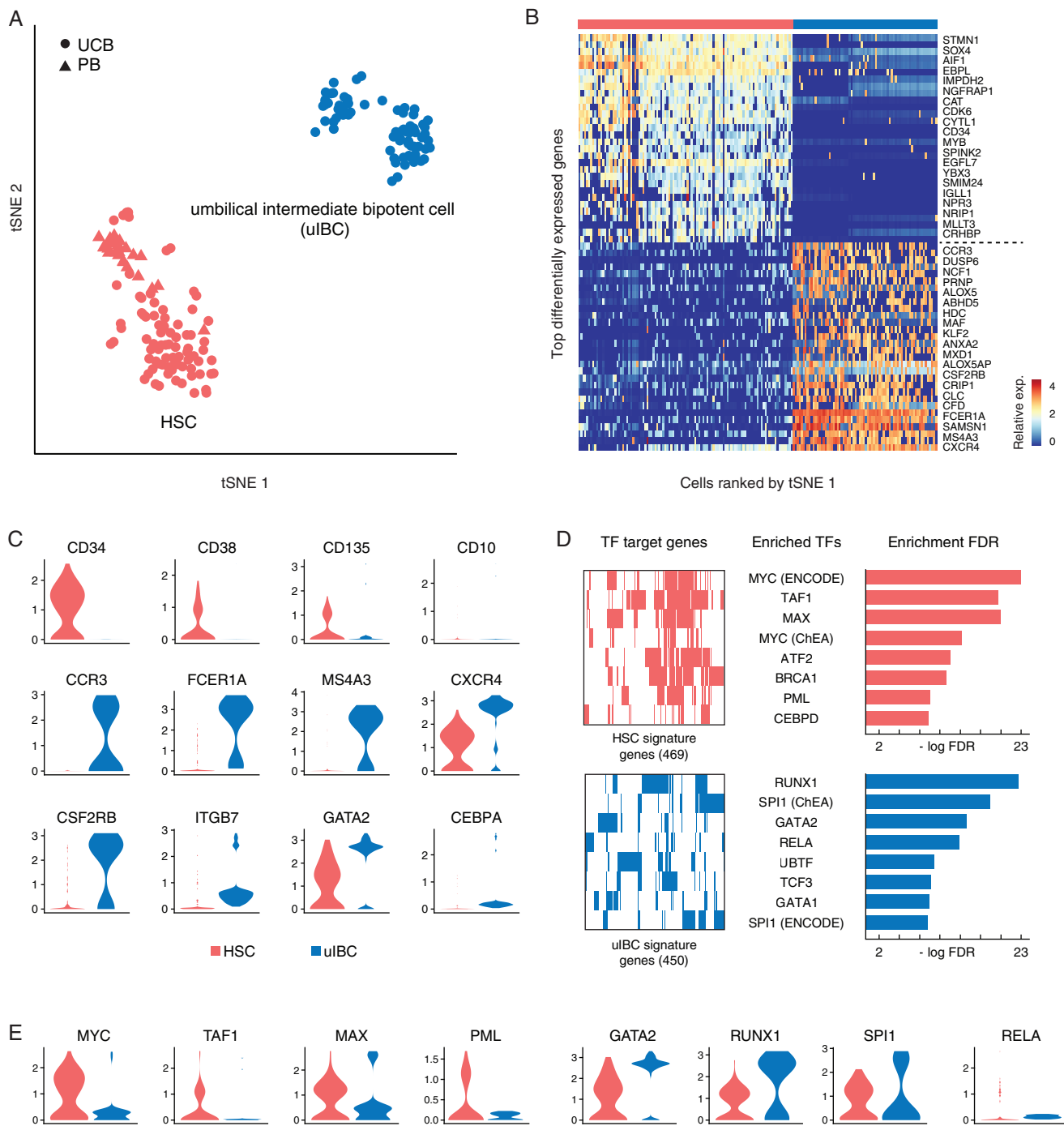
Figure 4

Supplementary Figure 1

A

B

Supplementary Figure 2

A

PB NK cells

UCB NK cells

B