Rescue Blood Contamination in scRNA-Seq Data by Originator, a Computational Deciphering Tool using Genetic and Contextual Information

Thatchayut Unjitwattana[1], Qianhui Huang[2], Yiwen Yang[3], Youqi Yang[4], Mengtian Zhou[5], Yuheng Du[2], Lana X. Garmire[2]

[1]Department of Biomedical Engineering, University of Michigan, 2200 Bonisteel, Ann Arbor, MI, 48109, USA

[2]Department of Computation Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Ave Ann Arbor, MI, 48109, USA

[3]School of Information, University of Michigan, 105 S State St, Ann Arbor, MI, 48109, USA

[4]Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI, 48109, USA

[5]Department of Statistics, University of Michigan,1085 S University Ave, Ann Arbor, MI, 48109, USA

Abbreviations:

CCC: Cell-cell communication

DE: Differential expression

GSEA: Gene set enrichment analysis

PBMC: Peripheral blood mononuclear cells

PDAC: Pancreatic ductal adenocarcinoma

QC: Quality control

scRNA-Seq: Single-cell RNA sequencing

Keywords: scRNA-Seq, blood cell, genetic origin, tissue heterogeneity, placenta, cancer

Abstract

Single-cell RNA sequencing (scRNA-Seq) data from complex human tissues have prevalent blood cell contamination due to the sample preparation process and may comprise cells of different genetic makeups. To reveal such complexity and annotate cells appropriately, we propose the first-of-its-kind computational framework, Originator, which deciphers single cells by genetic origin and separates blood cells from tissue-resident cells. We show that blood contamination is widely spread in scRNA-Seq data from a variety of tissues. We warn of the significant biases in downstream analysis without considering blood contamination and genetic contexts using pancreatic ductal adenocarcinoma and placenta data, respectively.

Main

Single-cell RNA sequencing (scRNA-Seq) data analysis of complex (*e.g.* placenta) or inherently mixed (*e.g.* tumor) tissues pose significant challenges in computational biology. Many tissues contain blood vessels, and the resulting scRNA-seq data often include blood cell types (e.g., T cells) having highly similar expression profiles with the same cell types in the resident tissues. For example, in tumor studies, precisely separating blood and tissue-resident immune cells in tumor tissues is crucial for understanding the tumor microenvironment [1]. In placenta tissues interfacing with the mother and fetus, isolating the maternal and fetal cells is pivotal to reveal cellular and immunological differences between them [2]. Therefore, a rigorous preprocessing step is required to enhance the quality of the data of interest, subsequently improving the downstream analysis. However, currently, there is no dedicated tool that performs such a function. Here, we fill in the void and propose a novel computational pipeline called Originator. It deconvolutes barcoded cells into different origins using inferred genotype information from scRNA-Seq data, as well as separating cells in the blood from those in solid tissues, an issue often encountered in scRNA-Seq experimentation.

Fig. 1. Illustration of Originator framework and the prevalence of blood immune cells in tissue samples. a, The input data are the scRNA-Seq experiment on tissue sections. (1) data preprocessing and cell type annotation. (2) separating barcoded cells into different origins by blood vs tissue residents context and optionally by inferred genotype information. (3) using the results in steps (1) and (2) to dissect tissue heterogeneity. (4) the functional downstream analyses with respect to cells' origins. **b,** Heatmap plots show inferred immune cell type proportions originated from the blood (left) vs. the tissue (right), from scRNA-Seq data of a wide-variety of organs including liver, lung, spleen, pancreas, and in the normal or cancer tissues. The cell types were annotated by the original publications, and blood vs tissue identification was done by Originator. NA: the cell type does not exist in the original publication.

The proposed Originator framework is illustrated in **Fig 1a**. Cells undergo standard scRNA-Seq preprocessing steps, including quality control (QC), normalization, data integration, clustering, and cell-type annotation. Then, Originator utilizes the differences in the genotype information from the scRNA-Seq reads to separate the barcoded cells into different (default N=2) genetic origins, for example, maternal vs. fetal origins. Next, given that many tissues have blood contamination, the subsequent step separates immune cells by blood vs. tissue-resident context using the publicly available whole blood scRNA-Seq data as the reference [3]. We validated the pipeline on artificially mixed data from a PBMCs dataset and a cell mixture dataset containing three breast cancer lines (T47D, BT474, MCF7), monocytes, lymphocytes, and stem cells **(Supplementary Notes)** [4, 5, 6]. As shown by the UMAP plot in **Supplementary Figure 1,** Originator separates the two compartments highly accurately, with AUC of 0.97, F-1 score of 0.97, and Area under the Precision-Recall Curve (AUCPR) of 0.96 (**Supplementary Table 1)**. After dissecting the heterogeneity of the scRNA-seq data, various downstream functional analyses can be performed.

To directly show the prevalence of blood immune cells in tissue samples, we applied Originator to eight publicly available scRNA-Seq data from a variety of organs, including lung, liver, spleen, and pancreas, with normal (or adjacent normal) and/or tumor tissue samples [7, 8, 9, 10, 11, 12, 13, 14]. As summarized in **Fig.1b**, in most datasets, the majority of the immune cells are from the blood, rather than the tissue. The detailed side-by-side comparison of UMAP plots of blood vs tissue immune cells is in **Supplementary Figure 2-3.** These results unambiguously demonstrate the vast amount of immune cells from blood, rather than the tissue microenvironment. Failure to remove these immune cells may yield significantly biased results. To directly show the issues due to mixed immune cells from blood, we first applied Originator to a pancreatic ductal adenocarcinoma (PDAC) scRNA-seq dataset [14]**,** to demonstrate its utility in removing cells that originate from blood. Originator successfully separates immune cells in the blood from those in the tumor tissue **(Fig 2a),** and mast and Myeloid-derived suppressor cells (MDSCs) are exclusive in tissues as expected. The immune cell type proportions of the two compartments are shown in **Fig 2b.** The blood has larger proportions of macrophages and regulatory T-cells, but fewer T-cells. We then performed the differential expression (DE) analysis comparing the common immune cell types between the blood and tumor tissues, and detected significant differences in gene expression in each immune cell type, especially for macrophage and T-cells **(Fig. 2c, Supplementary Table 2-5)**. For example, CCL4 and CCL5 expression is higher in T-cells in the tumor tissue compared to those in the blood (**Supplementary Figure 4a**), consistent with the previous study that these genes are directly associated with T-cell-inflamed phenotype and antigen-presenting cell-mediated processes in PDAC [15]. We subsequently performed Gene Set Enrichment Analysis (GSEA) comparing the immune cells' gene expression difference between the blood vs tissue compartments **(Fig. 2d)**. Toll-like receptor (TLR) signaling pathway and NOD-like receptor signaling pathway are more active in tissue-resident T-cells and macrophages compared to those in blood. However, natural killer cell-mediated cytotoxicity signaling pathway is downregulated in NK cells from the tumor

compared to the blood compartment, consistent with observed NK cell dysfunction in PDAC via the reduced cytotoxic granule components, granzyme B, and perforin [16]. To demonstrate the impact of blood cell contamination in functional interpretation, we inferred cell-cell communications CCC in tumor tissues before and after removing blood cells **(Fig 2e,f).** While the interactions among the cell types are globally less frequent/noisy in the tumor compartment, those between fibroblast/epithelium and other cell types are strengthened in blood cells, as reflected by the increased node size of these two cell types **(Fig 2e)**. Such a trend of change is most drastic for the MIF signaling pathway (**Fig 2f**), which was previously reported to drive the malignant character of pancreatic cancer [17]. Particularly, MIF signaling between T-cells/regulatory T-cells and other cell types (eg. epithelial cells) are absent after removing blood immune cells, indicating a compromised anti-tumor immune response in PDAC [18].

Some tissues may have different genetic lineages that also need to be addressed. For example, the placenta has both maternal and fetal tissues. We then applied Originator to the human placenta [19], and separated the single cells into fetal and maternal origins as well as blood vs. tissue-resident cells. The cells from the fetal origin account for the majority of the cell populations, and the tissue-resident cells significantly over-weigh blood cells, as expected **(Fig 2g)**. The framework correctly and exclusively assigns trophoblast cells, including cytotrophoblasts, syncytiotrophoblasts, and extravillous trophoblasts, to the fetal tissue (**Fig 2g,2h**) [20]. On the contrary, immune cells, vascular endothelial cells, and fibroblasts (type 1 and type 2) appear in both maternal and fetal tissues, also as expected **(Fig 2h)**. We further performed DE analysis on the common cell types between fetal and maternal tissues and discovered drastic differences in gene expression related to the local tissue environment **(Fig 2i, Supplementary Table 6-8).** For example, The top-ranked DE gene EGFL6 between maternal vs. fetal fibroblast cells is highly expressed in fibroblasts subtype 1 in fetal tissues as expected (**Supplementary Figure 5a**) [21]. However, it is much lower in fibroblast subtype 1 or mostly

absent in fibroblast subtype 2 in maternal tissues, similar to a previous report [22]. Among the DE genes from macrophage cells (**Supplementary Table 8**), SEPP1 is expressed at much higher levels in Hofbauer cells originating from the fetal tissue compared to the maternal tissue (**Supplementary Figure 5b)**, in accordance with a previous study [23]. We additionally performed GSEA on the DE genes of the common cell types between fetal and maternal cells in the placenta tissue **(Fig 2j)**. The MAPK signaling pathway is enriched in vascular endothelial cells from the fetal tissues, which agrees with its role in growth factor-induced fetoplacental angiogenesis [24]. On the other hand, fibroblast type 2 from the fetal origin is enriched with several other signaling pathways, such as calcium, hippo, and Ras signaling pathways as well as ECM-receptor interaction **(Fig 2j),** which indicates extracellular matrix (ECM) remodeling during trophoblast differentiation [25]. We also compared the CCC among the common cell types between the fetal and maternal tissues **(Fig 2k, 2l)**. Type 1 and type 2 fibroblasts both show different degrees of CCC with some other cell types when comparing the maternal and fetal tissue contexts. A higher interaction between fibroblast cells (type 1 and type 2) and vascular endothelial cells in fetal tissues reflects the active formation of the villous stroma underneath the syncytiotrophoblast and surrounding fetal capillaries [26]. On the contrary, higher interactions between type 2 fibroblasts and macrophages in maternal tissue may assist trophoblast invasion through growth factors and cytokines [27]. Thus, teasing apart the cell types by their genetic and local tissue context helps to refine the molecular analysis in placenta scRNA-seq data.

Fig. 2. Applications of Originator to pancreatic ductal adenocarcinoma (PDAC) and placenta scRNA-seq data. a, UMAP plot of cells in PDAC tissues (top-left) separated into blood and tissue-resident immune cells (bottom-left). UMAP plot of cells in PDAC tissues after blood immune cell removal is shown on the top-right. **b,** The cell type proportion barplot shows the immune cell types in blood and PDAC tumor tissue, as well as all cell types in the tumor tissue before and after blood immune cell

removal. **c,** Venn diagram of the significant DE genes identified when comparing immune cells in blood and tumor tissue. **d,** GSEA results comparing blood and tissue-resident immune cell types. **e,** Overall CCC in the tumor tissues before and after blood immune cell removal. **f,** CCC of MIF signaling pathway in tumor tissues before and after blood immune cell removal. **g,** UMAP plot of cells in placenta tissues deciphered into fetal vs. maternal origins and blood vs. placenta-resident cells. **h,** Cell type proportions in the different compartments related to the placenta. **i,** Upset plot of significant DE genes identified in common cell types between fetal and maternal placenta tissues. T-cell is excluded due to the near 0 counts in the fetal tissue. **j,** GSEA results comparing fetal and maternal tissues among common cell types. **k-l,** CCC of common cell types in fetal (**k**) and maternal (**l**) tissues.

In conclusion, Originator is the first dedicated tool to systematically decipher scRNA-Seq data by genetic origin and blood/tissue contexts in heterogeneous tissues, exemplified by PDAC and placenta. It can be used as an effective tool to remove the undesirable blood cells in scRNA-Seq data. It can also provide improved cell type annotations and other downstream functional analyses, based on the genetic background.

Methods

scRNA-Seq case study data sets

scRNA-Seq data from liver, lung, pancreas, and kidney are used in this study. For lung, we used a large scRNA-Seq dataset from the integrated human lung cell atlas (v1.0), containing samples from 107 individuals **[7]**, as well as lung cancer datasets from Xing et al. and Bischoff et al.**[11, 12]**. For liver tissues, we used the single-cell liver landscape dataset from five individuals **[9]**, and another healthy liver dataset from Guilliams et al. **[10]**. Additionally, we used lung, spleen, and liver scRNA-Seq datasets from Domínguez Conde et al. **[8]**. We obtained a kidney cancer dataset from Krishna et al. **[13]**. For pancreatic ductal adenocarcinoma (PDAC), we use a scRNA-Seq dataset from GSE212966, which includes two tumor tissues [14]. The last

dataset is the scRNA-Seq data from placenta tissues, comprising eight placenta samples from the previous study (EGA; https://www.ebi.ac.uk/ega/) hosted by the European Bioinformatics Institute (EBI; accession no. EGAS00001002449) [15]. Cell types in the placenta and PDAC tissues are annotated using the cell-type-specific marker genes (**Supplementary Table 9-10**).

Description of Originator for scRNA-Seq data analysis

Originator is a multi-module framework that can be used to preprocess scRNA-Seq data from heterogeneous tissues. It consists of one mandatory step based on the tissue vs blood compartments and also an optional step based on the genetic origins. Originator takes the gene expression matrix processed by Cell Ranger (version 7.1.0) of 10x genomics as the input [28]. For genetic background based deciphering, it also uses the BAM file. It processes and outputs the R data serialization file (RDS). This output file contains gene expression and related information, including the cell type, blood and tissue immune cell annotation, and annotation by genetic information if this function is needed.

First, Originator separates immune cells in blood vs. those in the tissue using the blood immune cell scRNA-Seq reference constructed from the publicly available scRNA-seq data from the whole blood cells [3]. Annotated immune cells of interest (monocytes, macrophages, T-cells, regulatory T-cells, plasma cells, NK cells, and B-cells) from the datasets of this study are aligned to the whole blood scRNA-Seq reference data using package Seurat (4.3.0) [29]. For each immune cell type, the top 10 embeddings from the UMAP analysis of each scRNA-Seq sample are used to compute a pairwise Euclidean distance matrix between each query immune cell and the reference whole blood. To separate each query of immune cells into blood or tissue-resident cells, k-means clustering (k = 2) is applied using Euclidean distances relative to the whole blood immune cell reference. The cluster more similar to the whole blood immune cell

references is annotated as the blood immune cells, and the other cluster is determined as the tissue-resident immune cells.

If the tissue contains the mixture of cells of different genetic background, such as maternal and fetal origins, Originator provides an optional step to decipher the barcoded cells by the genetic origin relies on the overlapping genetic variants extracted from the scRNA-Seq data. Genetic variants are extracted by *bcftools* (version 1.17) [30]. Variants informative of genetic origins are determined by excluding single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) greater than x percentages (default x=10), using the 1000 Genomes Phase 3 reference panel (release date: 2013/05/02) [31]. The *freemuxlet* package is used to separate the single cells with the parameter '--nsample N' (default N=2), as implemented in popscle (https://github.com/statgen/popscle) [32]. N denotes the number of genetic origins of the single cells (e.g., N = 2 for the maternal and fetal origins in a placenta scRNA-Seq data). Mosaic doublets (a single barcoded cell with a 50% genetic mixture from two individuals) are also identified in this step as part of the quality control. As trophoblast cells in the placenta are of fetal origin, the cluster that contains trophoblast cells is identified with the fetal origin, while the other cluster is marked as the maternal origin.

Simulation

The artificially mixed tissue-blood-resident data was generated to assess the ability of Originator to separate blood cells and tissue immune residents. We included only mixture 1 (GSM6807287) in the simulated data to reduce the computational cost during the origins and tissue-blood deconvolution process (**Supplementary Notes**) [6].

Downstream analysis

DE analysis was performed using *FindMarkers()* in the package Seurat (4.3.0) [22]. GSEA is done using *gseKEGG()* in the package ClusterProfiler (4.8.2) with the gene set information from

the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [33, 34]. CCC inference was performed using the package CellChat [35].

Code availability

Originator is freely available to use at URL: https://github.com/lanagarmire/Originator

Acknowledgments

TU is supported by the Royal Thai Government fellowship from the Thai Government. LXG is supported by grants by NIH/NIGMS, R01 LM012373 and R01 LM012907 awarded by NLM, and R01 HD084633 awarded by NICHD. The authors acknowledge all lab members of Garmire Group for helpful discussions.

Author Contributions

L.X.G. and Q.H. conceived this project and supervised the study. T.U. modified the pipeline, carried the analysis, and wrote the manuscript. Q.H. wrote the Originator pipeline. M.Z., Y.Y., Y.Y., and Y.D. assisted the analysis and tested the package.

Competing Interests

LXG is a member of the Scientific Advisory Boards of Simulations Plus.

References

1. Del Prete, A., & Wu, Q. (2022). Editorial: Tissue-resident immune cells in tumor immunity and immunotherapy. *Frontiers in Cell and Developmental Biology*, *10*, 1068720. https://doi.org/10.3389/fcell.2022.1068720

2. Hussain, T., Murtaza, G., Kalhoro, D. H., Kalhoro, M. S., Yin, Y., Chughtai, M. I., Tan, B., Yaseen, A., & Rehman, Z. U. (2022). Understanding the Immune System in Fetal Protection and Maternal Infections during Pregnancy. *Journal of Immunology Research*, *2022*, 1–12. https://doi.org/10.1155/2022/7567708

3. The Tabula Sapiens Consortium*, Jones, R. C., Karkanias, J., Krasnow, M. A., Pisco, A. O., Quake, S. R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., Harper, W., Hemenez, M., Ponnusamy, R., Salehi, A., Sanagavarapu, B. A., Spallino, E., Aaron, K. A., Concepcion, W., Gardner, J. M., … Wyss-Coray, T. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*, *376*(6594), eabl4896. https://doi.org/10.1126/science.abl4896

4. Hansen, K. D., Risso, D., & Hicks, S. (2023). *TENxPBMCData: PBMC data from 10X Genomics* [Computer software]. doi:10.18129/B9.bioc.TENxPBMCData

5. 10x Genomics. (2017). *8k PBMCs from a Healthy Donor, Single Cell Gene Expression Dataset by Cell Ranger 1.3.0* [dataset]. https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/pbmc8k

6. Sumazin, P. (2023). *GSE220606, Effective methods for bulk RNA-seq deconvolution using scnRNA-seq transcriptomes [cell mixtures scRNA-seq]* [dataset]. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220606

7. Sikkema, L., Ramírez-Suástegui, C., Strobl, D. C., Gillett, T. E., Zappia, L., Madissoon, E., Markov, N. S., Zaragosi, L.-E., Ji, Y., Ansari, M., Arguel, M.-J., Apperloo, L., Banchero, M., Bécavin, C., Berg, M., Chichelnitskiy, E., Chung, M., Collin, A., Gay, A. C. A., … Theis, F. J. (2023). An integrated cell atlas of the lung in health and disease. *Nature Medicine*, *29*(6), 1563–1577. https://doi.org/10.1038/s41591-023-02327-2

8. Domínguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., Gomes, T., Howlett, S. K., Suchanek, O., Polanski, K., King, H. W., Mamanova, L., Huang, N., Szabo, P. A., Richardson, L., Bolt, L., Fasouli, E. S., Mahbubani, K. T., Prete, M., Tuck, L., … Teichmann, S.

A. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, *376*(6594), eabl5197. https://doi.org/10.1126/science.abl5197

9. MacParland, S. A., Liu, J. C., Ma, X.-Z., Innes, B. T., Bartczak, A. M., Gage, B. K., Manuel, J., Khuu, N., Echeverri, J., Linares, I., Gupta, R., Cheng, M. L., Liu, L. Y., Camat, D., Chung, S. W., Seliga, R. K., Shao, Z., Lee, E., Ogawa, S., … McGilvray, I. D. (2018). Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nature Communications*, *9*(1), 4383. https://doi.org/10.1038/s41467-018-06318-7

10. Guilliams, M., Bonnardel, J., Haest, B., Vanderborght, B., Wagner, C., Remmerie, A., Bujko, A., Martens, L., Thoné, T., Browaeys, R., De Ponti, F. F., Vanneste, B., Zwicker, C., Svedberg, F. R., Vanhalewyn, T., Gonçalves, A., Lippens, S., Devriendt, B., Cox, E., … Scott, C. L. (2022). Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell*, *185*(2), 379-396.e38. https://doi.org/10.1016/j.cell.2021.12.018

11. Xing, X., Yang, F., Huang, Q., Guo, H., Li, J., Qiu, M., Bai, F., & Wang, J. (2021). Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. *Science Advances*, *7*(5), eabd9738. https://doi.org/10.1126/sciadv.abd9738

12. Bischoff, P., Trinks, A., Obermayer, B., Pett, J. P., Wiederspahn, J., Uhlitz, F., Liang, X., Lehmann, A., Jurmeister, P., Elsner, A., Dziodzio, T., Rückert, J.-C., Neudecker, J., Falk, C., Beule, D., Sers, C., Morkel, M., Horst, D., Blüthgen, N., & Klauschen, F. (2021). Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene*, *40*(50), 6748–6758. https://doi.org/10.1038/s41388-021-02054-3

13. Krishna, C., DiNatale, R. G., Kuo, F., Srivastava, R. M., Vuong, L., Chowell, D., Gupta, S., Vanderbilt, C., Purohit, T. A., Liu, M., Kansler, E., Nixon, B. G., Chen, Y.-B., Makarov, V., Blum, K. A., Attalla, K., Weng, S., Salmans, M. L., Golkaram, M., … Hakimi, A. A. (2021). Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to

tumor topology and therapy efficacy. *Cancer Cell*, *39*(5), 662-677.e6. https://doi.org/10.1016/j.ccell.2021.03.007

14. Chen, K., Wang, Q., Liu, X., Tian, X., Dong, A., & Yang, Y. (2023). Immune profiling and prognostic model of pancreatic cancer using quantitative pathology and single-cell RNA sequencing. *Journal of Translational Medicine*, *21*(1), 210. https://doi.org/10.1186/s12967-023-04051-4

15. Romero, J. M., Grünwald, B., Jang, G.-H., Bavi, P. P., Jhaveri, A., Masoomian, M., Fischer, S. E., Zhang, A., Denroche, R. E., Lungu, I. M., De Luca, A., Bartlett, J. M. S., Xu, J., Li, N., Dhaliwal, S., Liang, S.-B., Chadwick, D., Vyas, F., Bronsert, P., … Gallinger, S. (2020). A Four-Chemokine Signature Is Associated with a T-cell–Inflamed Phenotype in Primary and Metastatic Pancreatic Cancer. *Clinical Cancer Research*, *26*(8), 1997–2010. https://doi.org/10.1158/1078-0432.CCR-19-2803

16. Kim, H.-A., Kim, H., Nam, M.-K., Park, J. K., Lee, M.-Y., Chung, S., Lee, K.-M., & Kuh, H.-J. (2023). Suppression of the antitumoral activity of natural killer cells under indirect coculture with cancer-associated fibroblasts in a pancreatic TIME-on-chip model. *Cancer Cell International*, *23*(1), 219. https://doi.org/10.1186/s12935-023-03064-9

17. Yang, S., He, P., Wang, J., Schetter, A., Tang, W., Funamizu, N., Yanaga, K., Uwagawa, T., Satoskar, A. R., Gaedcke, J., Bernhardt, M., Ghadimi, B. M., Gaida, M. M., Bergmann, F., Werner, J., Ried, T., Hanna, N., Alexander, H. R., & Hussain, S. P. (2016). A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2. *Cancer Research*, *76*(13), 3838–3850. https://doi.org/10.1158/0008-5472.CAN-15-2841

18. Bacher, M., Metz, C. N., Calandra, T., Mayer, K., Chesney, J., Lohoff, M., Gemsa, D., Donnelly, T., & Bucala, R. (1996). An essential regulatory role for macrophage migration inhibitory factor in T-cell activation. *Proceedings of the National Academy of Sciences*, *93*(15), 7849–7854. https://doi.org/10.1073/pnas.93.15.7849

19. Tsang, J. C. H., Vong, J. S. L., Ji, L., Poon, L. C. Y., Jiang, P., Lui, K. O., Ni, Y.-B., To, K. F., Cheng, Y. K. Y., Chiu, R. W. K., & Lo, Y. M. D. (2017). Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proceedings of the National Academy of Sciences*, *114*(37). https://doi.org/10.1073/pnas.1710470114

20. Red-Horse, K., Zhou, Y., Genbacev, O., Prakobphol, A., Foulk, R., McMaster, M., & Fisher, S. J. (2004). Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface. *Journal of Clinical Investigation*, *114*(6), 744–754. https://doi.org/10.1172/JCI200422991

21. Suryawanshi, H., Morozov, P., Straus, A., Sahasrabudhe, N., Max, K. E. A., Garzia, A., Kustagi, M., Tuschl, T., & Williams, Z. (2018). A single-cell survey of the human first-trimester placenta and decidua. *Science Advances*, *4*(10), eaau4788. https://doi.org/10.1126/sciadv.aau4788

22. Tang, C.-T., Zhang, Q.-W., Wu, S., Tang, M.-Y., Liang, Q., Lin, X.-L., Gao, Y.-J., & Ge, Z.-Z. (2020). Thalidomide targets EGFL6 to inhibit EGFL6/PAX6 axis-driven angiogenesis in small bowel vascular malformation. *Cellular and Molecular Life Sciences*, *77*(24), 5207–5221. https://doi.org/10.1007/s00018-020-03465-3

23. Liang, G., Zhou, C., Jiang, X., Zhang, Y., Huang, B., Gao, S., Kang, Z., Ma, D., Wang, F., Gottgens, B., Wang, H., Han, J.-D. J., & Liu, F. (2021). De novo generation of macrophage from placenta-derived hemogenic endothelium. *Developmental Cell*, *56*(14), 2121-2133.e6. https://doi.org/10.1016/j.devcel.2021.06.005

24. Wang, K., & Zheng, J. (2012). Signaling regulation of fetoplacental angiogenesis. *Journal of Endocrinology*, *212*(3), 243–255. https://doi.org/10.1530/JOE-11-0296

25. MacPhee, D. J., Mostachfi, H., Han, R., Lye, S. J., Post, M., & Caniggia, I. (2001). Focal Adhesion Kinase Is a Key Mediator of Human Trophoblast Development. *Laboratory Investigation*, *81*(11), 1469–1483. https://doi.org/10.1038/labinvest.3780362

26. Palaiologou, E., Etter, O., Goggin, P., Chatelet, D. S., Johnston, D. A., Lofthouse, E. M., Doherty, R., Pearson Farr, J., Sengers, B. G., Torrens, C., Cleal, J. K., Page, A. M., & Lewis, R. M. (2020). Human placental villi contain stromal macrovesicles associated with networks of stellate cells. *Journal of Anatomy*, *236*(1), 132–141. https://doi.org/10.1111/joa.13082

27. Ueshima, C., Kataoka, T. R., Osakabe, M., Sugimoto, A., Ushirokawa, A., Shibata, Y., Nakamura, H., Shibuya, R., Minamiguchi, S., Sugai, T., & Haga, H. (2022). Decidualization of Stromal Cells Promotes Involvement of Mast Cells in Successful Human Pregnancy by Increasing Stem Cell Factor Expression. *Frontiers in Immunology*, *13*, 779574. https://doi.org/10.3389/fimmu.2022.779574

28. Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*(1), 14049. https://doi.org/10.1038/ncomms14049

29. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

30. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

31. The 1000 Genomes Project Consortium, Corresponding authors, Auton, A., Abecasis, G. R., Steering committee, Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A.,

Green, E. D., Hurles, M. E., Knoppers, B. M., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

32. Zhang, F., & Min Kang, H. (2021). *popscle: A suite of population scale analysis tools for single-cell genomics data (freemuxlet)* [Computer software]. https://github.com/statgen/popscle

33. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, *2*(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

34. Kanehisa, M., & Goto, S. (n.d.). *KEGG: Kyoto Encyclopedia of Genes and Genomes*.

35. Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M. V., & Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. *Nature Communications*, *12*(1), 1088. https://doi.org/10.1038/s41467-021-21246-9

Fig. 1. Illustration of Originator framework and the prevalence of blood immune cells in tissue samples.

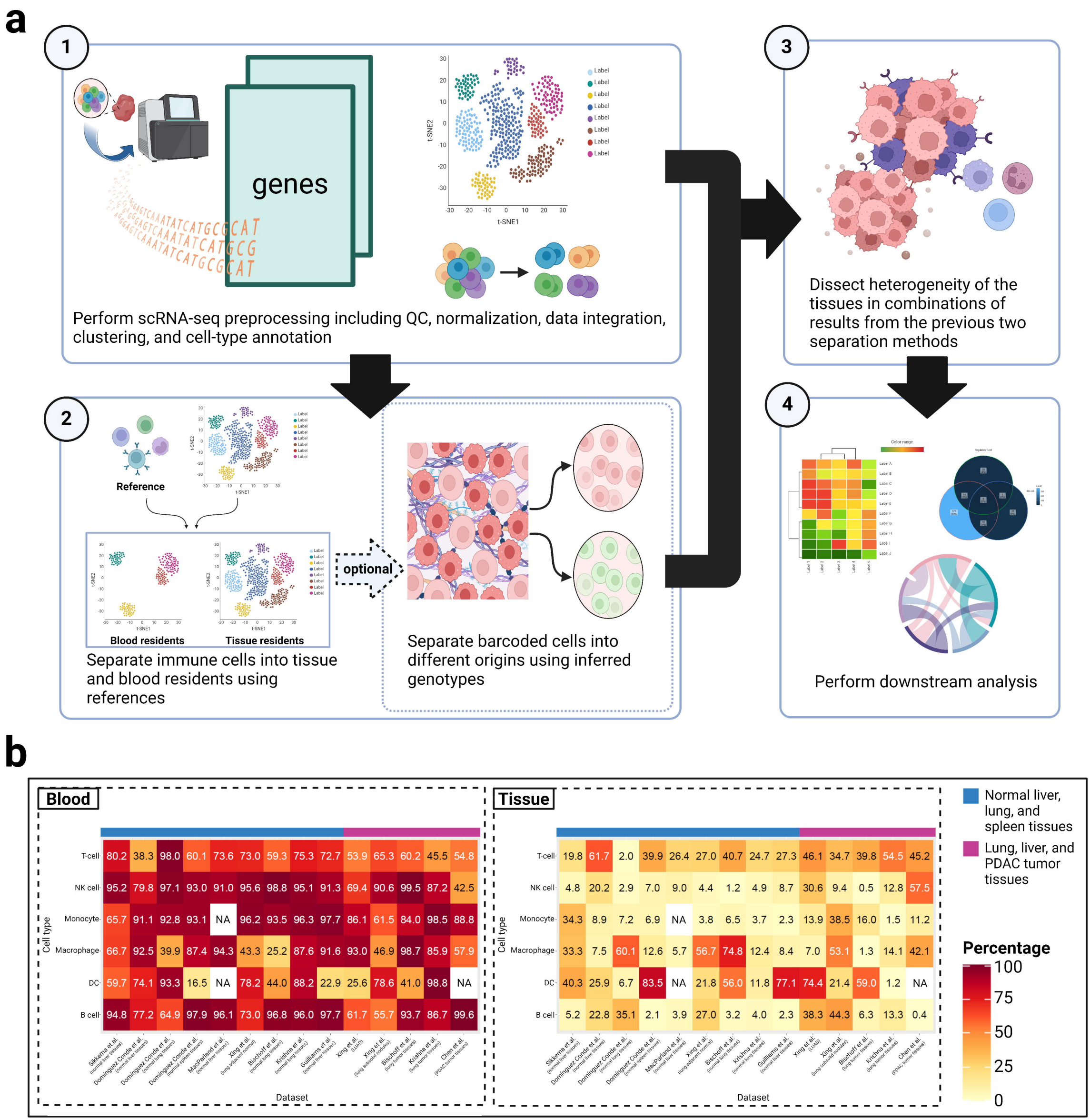
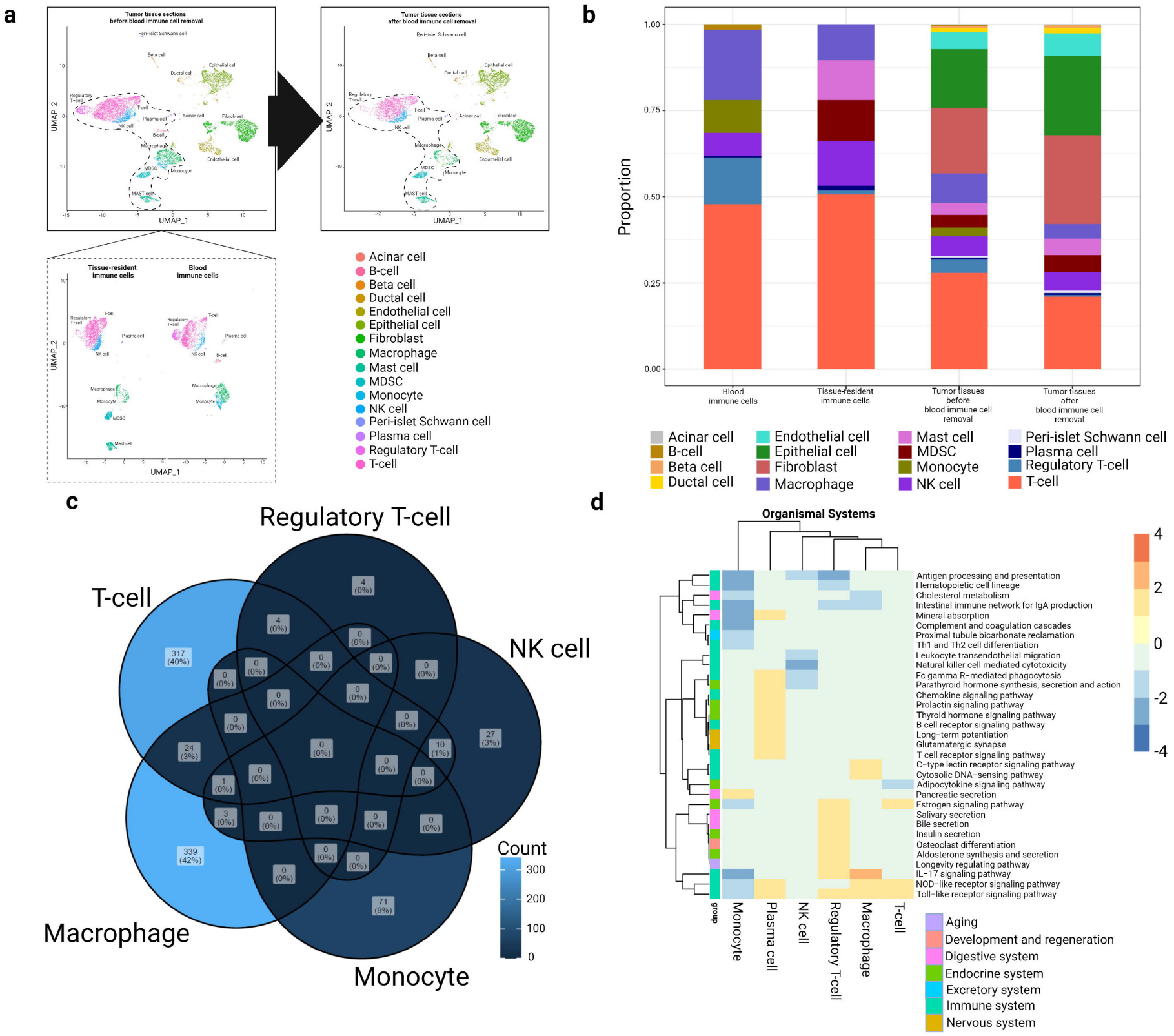
Fig. 2. Applications of Originator to pancreatic ductal adenocarcinoma (PDAC) and placenta scRNA-seq data.