BRIEF REPORT

Single-cell RNA-seq data have prevalent blood contamination but can be rescued by Originator, a computational tool separating single-cell RNA-seq by genetic and contextual information

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

*Correspondence:
lgarmire@umich.edu

[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] Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA
[4] Department of Statistics, University of Michigan, 1085 S University Ave, Ann Arbor, MI 48109, USA

Abstract

Single-cell RNA sequencing (scRNA-seq) data from complex human tissues have prevalent blood cell contamination during the sample preparation process. They may also comprise cells of different genetic makeups. We propose a new computational framework, Originator, which deciphers single cells by genetic origin and separates immune cells of blood contamination from those of expected tissue-resident cells. We demonstrate the accuracy of Originator at separating immune cells from the blood and tissue as well as cells of different genetic origins, using a variety of artificially mixed and real datasets, including pancreatic cancer and placentas as examples.

Keywords: ScRNA-seq, Blood cell, Contamination, Genetic origin, Tissue heterogeneity, Placenta, Cancer

Background

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.

Unjitwattana *et al. Genome Biology*   (2025) 26:52

Page 2 of 13

However, currently there is no dedicated tool to perform 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.

Results and discussion

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 [3]. Then, Originator takes advantage of the differences in the genotype information from the scRNA-seq reads and utilizes
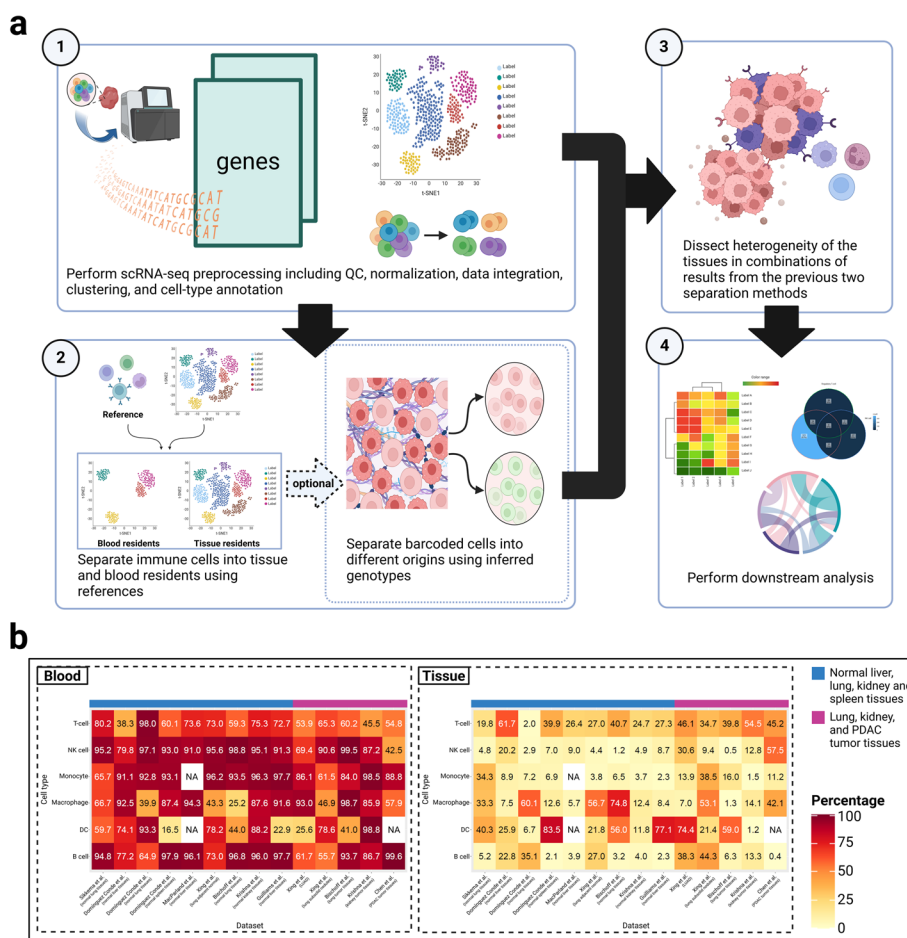

Fig. 1 Illustration of Originator framework and the prevalence of blood immune cells in tissue samples. **a** The input data are the scRNA-seq experiment on tissue sections. (1) Data preprocessing and cell type annotation. (2) Separating barcoded cells into different origins by blood vs. tissue residents 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, kidney, and 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

Unjitwattana *et al. Genome Biology* (2025) 26:52

Page 3 of 13

freemuxlet, a reference-free version of the popsicle suite [4] to separate the barcoded cells into different (default $N=2$) genetic origins, for example, maternal vs. fetal origins. Freemuxlet is chosen, as it showed better accuracies at recovering the genetic origins in comparison to scSplit [5], in the scRNA-seq data of two placentas and the paired clear cell renal cell carcinoma (ccRCC) and PBMC cells from the same patient (Additional file 1: Table S1). Next, given that many tissues have blood contamination, the subsequent step separates immune cells by blood vs. expected tissue-resident context using the publicly available whole blood scRNA-seq data as the reference [6]. After dissecting the heterogeneity of the scRNA-seq data, various downstream functional analyses can be performed.

Current public scRNA-seq data lack paired PBMC (or blood) and complex primary tissue completely free of blood contamination from the same person. We therefore first validated the pipeline using the 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 (Additional file 2: Supplementary Notes) [7–9]. Before creating the artificial mixture scRNA-seq data, we removed the batch effect between the PBMC and the breast cancer cell lines (Additional file 2: Fig. S1), using Harmony [10] which showed good performance for this task previously [11]. As shown by the UMAP plot in Additional file 2: Fig. S1, Originator separates the two compartments highly accurately, with AUC of 0.96, F-1 score of 0.97, and area under the precision-recall curve (AUCPR) of 0.96 (Additional file 1: Table S2). Next, we tested Originator on the dataset of Krishna et al. [12], which contains paired ccRCC tissue and PBMC from the same patients. To generate the "ground truth" cell types for blood-eliminated ccRCC tissue, we applied Originator (or Seurat, for comparison) to the ccRCC to remove the potential blood immune cells from those residing in the tissue. We then integrated the cleaned ccRCC tissues and PBMC to generate an artificial mixture. Next, we ran Originator on this pre-cleaned mixture dataset for five iterations and obtained averaged F-scores of 0.98, 0.93, 0.91, and 0.99 for B cell, CD4 T-cell, CD8 T-cell, and NK cell respectively on the pre-cleaned, mixed dataset (Additional file 2: Fig. S2). We thus conclude that Originator is effective at removing blood contamination from the results of both datasets above.

To signify 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, kidney, and pancreas, with normal (or adjacent normal) and/or tumor tissue samples [12–19]. 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 Additional file 2: Figs. S3–S4. 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 demonstrate the potential biases in the downstream analyses due to contamination of the immune cells from the blood, we first applied Originator to a pancreatic ductal adenocarcinoma (PDAC) scRNA-seq dataset [19], to demonstrate its utility

Unjitwattana *et al. Genome Biology*   (2025) 26:52

Page 4 of 13

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 those expected from the tumor tissues, and detected significant differences in gene expression in each immune cell type, especially for macrophage and T-cells (Fig. 2c, Additional file 1: Tables S3–S6) [20, 21]. For example, CCL4 and CCL5 expression is higher in T-cells in the tumor tissue compared to those in the blood (Additional file 2: Fig. S5a), consistent with the previous study that these genes are directly associated with T-cell-inflamed phenotype and antigen-presenting cell-mediated processes in PDAC [22]. We also examined the genes common in some immune cell types, but showed significant differences between the blood vs. tumor compartments (Additional file 2: Supplementary Notes) [23–43]. 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 expected 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 [44]. 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 [45]. Particularly, MIF signaling between T-cells/regulatory T-cells and other cell types (e.g., epithelial cells) is absent after removing blood immune cells, indicating a compromised anti-tumor immune response in PDAC [46].

(See figure on next page.)

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 expected 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 expected tumor resident tissue. **d** GSEA results comparing blood and expected 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. CCC of common cell types in fetal (**k**) and maternal (**l**) tissues
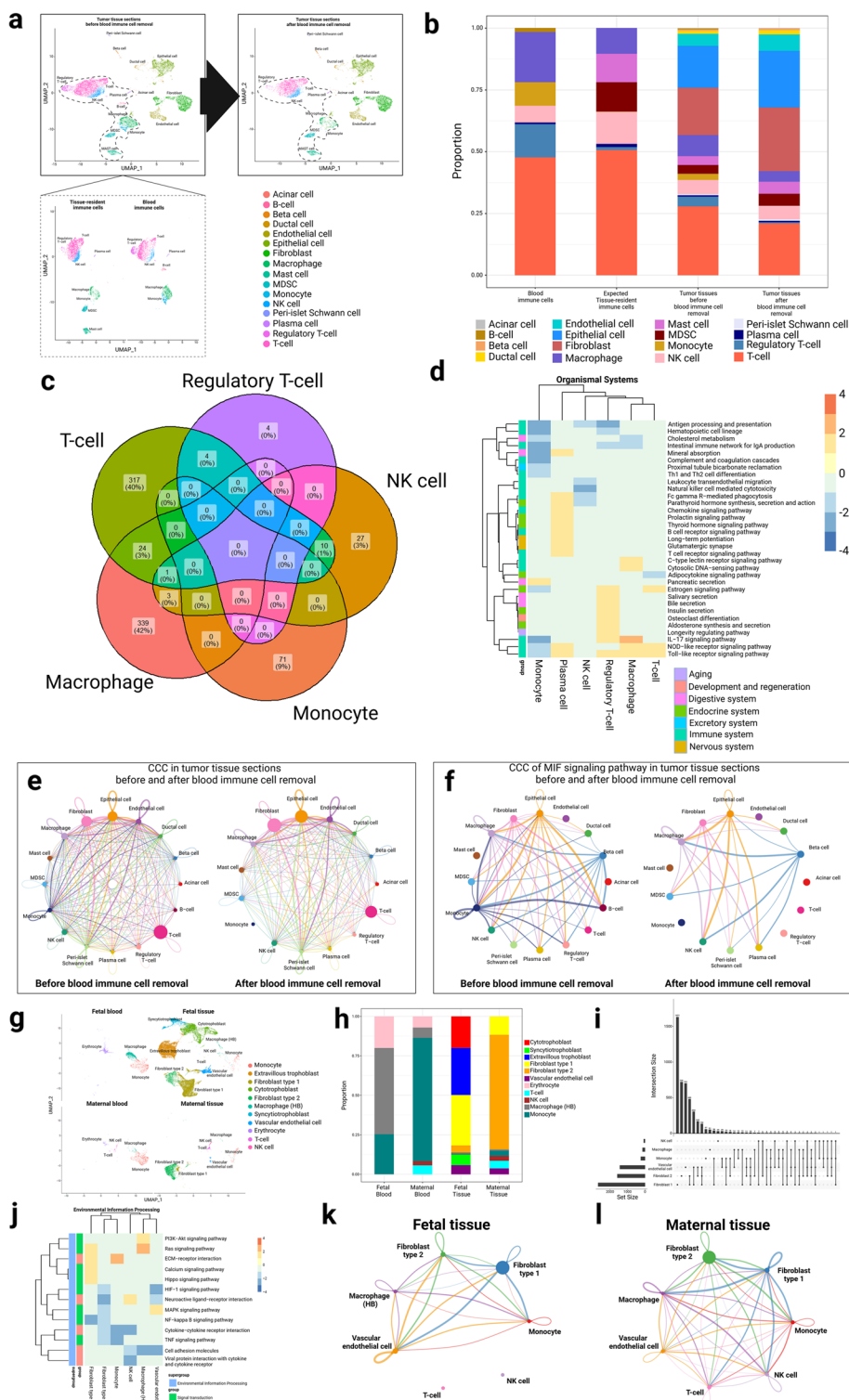
Fig. 2 (See legend on previous page.)

Unjitwattana *et al. Genome Biology*   (2025) 26:52

Page 6 of 13

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 [47] and separated the single cells into fetal and maternal origins as well as blood vs. expected tissue-resident cells. The cells from the fetal origin account for the majority of the cell populations, and the expected 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, h) [48]. 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, Additional file 1: Tables S7–S9). 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 (Additional file 2: Fig. S6a) [49]. However, it is much lower in fibroblast subtype 1 or mostly absent in fibroblast subtype 2 in maternal tissues, similar to a previous report [50]. Among the DE genes from macrophage cells (Additional file 1: Table S9), SEPP1 is expressed at much higher levels in Hofbauer cells originating from the fetal tissue compared to the maternal tissue (Additional file 2: Fig. S6b), in accordance with a previous study [51]. 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 [52]. 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 [53]. We also compared the CCC among the common cell types between the fetal and maternal tissues (Fig. 2k, l). 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 [54]. On the contrary, higher interactions between type 2 fibroblast and macrophages in maternal tissue may assist trophoblast invasion through growth factors and cytokines [55]. Thus, teasing apart the cell types by their genetic and local tissue context helps to refine the molecular analysis in placenta scRNA-seq data.

Conclusions

Originator is the first dedicated systematic tool to decipher scRNA-seq data by genetic origin and blood/tissue contexts in heterogeneous tissues. 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. Future work will be focused on generating pan-tissue immune cell

Unjitwattana *et al. Genome Biology*   (2025) 26:52

Page 7 of 13

atlas, which is free of immune cells originating from blood contamination, which will better annotate the expected tissue-resident immune cells truthfully from the tissues of interest.

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 [13], as well as lung cancer datasets from Xing et al. and Bischoff et al. [17, 18]. For liver tissues, we used the single-cell liver landscape dataset from five individuals [15] and another healthy liver dataset from Guilliams et al. [16]. Additionally, we used lung, spleen, and liver scRNA-seq datasets from Domínguez Conde et al. [14]. We obtained the paired blood and kidney cancer dataset from Krishna et al. [12]. For pancreatic ductal adenocarcinoma (PDAC), we use a scRNA-seq dataset from GSE212966, which includes two tumor tissues [19]. 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) [47]. Cell types in the placenta and PDAC tissues are annotated using the cell-type-specific marker genes [56] (Additional file 1: Tables S10–S11).

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 $10 \times$ genomics as the input [57]. 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 [6]. 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) [58]. For each immune cell type, the top 10 latent variables from each scRNA-seq sample are used to compute a pairwise Euclidean distance matrix between each query immune cell and the reference whole blood. The latent variables can be obtained by UMAP or PCA-based dimension reduction. We compared UMAP-based and PCA-based dimension reduction on the data of Krishna et al. [12] and found that the former yielded higher accuracies in detecting the immune cell types (Additional file 2: Fig. S7). Thus, we used UMAP-based dimension reduction in this report. To separate each query of immune cells into blood or expected tissue-resident cells, *k*-means

Unjitwattana *et al. Genome Biology*   (2025) 26:52

Page 8 of 13

clustering (default $k=2$) 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 expected tissue-resident immune cells.

If the tissue contains a mixture of cells of different genetic backgrounds, such as maternal and fetal origins, Originator provides an optional step to decipher the barcoded cells by the genetic origin relying on the overlapping genetic variants extracted from the scRNA-seq data. Genetic variants are extracted by *bcftools* (version 1.17) [59]. 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) [60]. 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) [4]. 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 fetal origin, while the other cluster is marked as maternal origin.

Testing the performance of Originator

The artificially mixed tissue-blood-resident data were generated to assess the ability of Originator to separate blood cells and tissue immune residents. We included two datasets: (1) 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 [7–9]. (2) A paired pre-cleaned ccRCC tissues and PBMC from the same patient provided by Krishna et al. [12]. The details of the evaluation are in Additional file 2: Supplementary Notes. Additionally, we also benchmarked freemuxlet [4] (used in Originator) against scSplit [5, 61] in identifying cells from different genetic origins using two different datasets, including (1) scRNA-seq data of two placenta samples and (2) mixed PBMC scRNA-seq data from two ccRCC patients provided by Krishna et al. [12, 47] (Additional file 2: Supplementary Notes).

Downstream analysis

DE analysis was performed using *FindMarkers()* in the package Seurat (4.3.0) [58]. 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 [62, 63]. CCC inference was performed using the package CellChat [64].

Code availability

Originator is freely available to use at URL: https://github.com/lanagarmire/Originator under an MIT license compliant with Open Source Initiative (OSI) (http://opensource.org/licenses) [65]. The source code used in the manuscript is also publically available at https://zenodo.org/records/14750795 (DOI: https://doi.org/10.5281/zenodo.14750794) [66].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03495-9.

Additional file 1: Tables S1–S12.

Additional file 2: Supplementary Notes, Figs. S1-S8.

Peer review information

Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' 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., Y.D. and L.T. assisted the analysis and tested the package. All authors read and approved the final manuscript.

Authors' X handles

X handles: @GarmireLab (Garmire's group); @GarmireGroup (Lana X. Garmire).

Data availability

We use a healthy lung scRNA-seq dataset from the integrated human lung cell atlas (v1.0), containing samples from 107 individuals [13, 67]. Additionally, we use lung, spleen, and liver scRNA-seq datasets from Domínguez Conde et al. [14, 68]. For liver tissues, we used the single-cell liver landscape dataset from five individuals [15, 69], as well as another healthy liver dataset from Guilliams et al. [16, 70]. For cancer tissues, we incorporated lung cancer datasets from Xing et al. and Bischoff et al. [17, 18, 71, 72], and a kidney cancer dataset from Krishna et al. [12, 73]. For pancreatic ductal adenocarcinoma (PDAC), we use a scRNA-seq dataset from GSE212966, which includes two tumor tissues [19, 74]. 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) [47, 75]. For simulation, we generated the data using an 8 k PBMCs dataset from a healthy donor (blood-resident cells) and an in vitro cell mixture (expected tissue-resident cells) containing three breast cancer lines (T47D, BT474, MCF7), monocytes (Thp1), lymphocytes (Jukrat), and stem cells (hMSC) [8, 9]. Originator is freely accessible on Github at https://github.com/lanagarmire/Originator [65] and is distributed under the MIT license, compliant with the Open Source Initiative (OSI) (http://opensource.org/licenses). Additionally, the source code used in this manuscript is publicly available on Zenodo at https://zenodo.org/records/14750795 (DOI: https://doi.org/https://doi.org/10.5281/zenodo.14750794) [66].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Del Prete A, Wu Q. Editorial: tissue-resident immune cells in tumor immunity and immunotherapy. Front Cell Dev Biol. 2022 Nov 22 [cited 2024 May 31];10:1068720. Available from: https://www.frontiersin.org/articles/10.3389/fcell.2022.1068720/full.
2. Hussain T, Murtaza G, Kalhoro DH, Kalhoro MS, Yin Y, Chughtai MI, et al. Understanding the immune system in fetal protection and maternal infections during pregnancy. Xu H, editor. J Immunol Res. 2022 Jun 24 [cited 2024 May 31];2022:1–12. Available from: https://www.hindawi.com/journals/jir/2022/7567708/.
3. Stanojevic S, Li Y, Ristivojevic A, Garmire LX. Computational methods for single-cell multi-omics integration and alignment. Genomics Proteomics Bioinformatics. 2022 Oct 1 [cited 2025 Jan 1];20(5):836–49. Available from: https://academic.oup.com/gpb/article/20/5/836/7230458.
4. Kang HM, Zhang F. popscle: a suite of population scale analysis tools for single-cell genomics data (freemuxlet). 2021. Available from: https://github.com/statgen/popscle.

5. Xu J, Falconer C, Nguyen Q, Crawford J, McKinnon BD, Mortlock S, et al. Genotype-free demultiplexing of pooled single-cell RNA-seq. Genome Biol. 2019 Dec [cited 2024 Dec 23];20(1):290. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1852-7.

6. The Tabula Sapiens Consortium*, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, et al. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. Science. 2022 May 13 [cited 2024 May 31];376(6594):eabl4896. Available from: https://www.science.org/doi/10.1126/science.abl4896.

7. Hansen K, Risso D, Hicks S. TENxPBMCData: PBMC data from 10X Genomics. 2023. Available from: https://bioconductor.org/packages/TENxPBMCData.

8. 10x Genomics. 8k PBMCs from a healthy donor, single cell gene expression dataset by Cell Ranger 1.3.0. Datasets. 10x Genomics Datasets. 2017. https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/pbmc8k.

9. Sumazin P. GSE220606, Effective methods for bulk RNA-seq deconvolution using scnRNA-seq transcriptomes [cell mixtures scRNA-seq]. Datasets. Gene Expression Omnibus. 2023. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220606.

10. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019 Dec [cited 2024 Apr 29];16(12):1289–96. Available from: https://www.nature.com/articles/s41592-019-0619-0.

11. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, et al. Benchmarking atlas-level data integration in single-cell genomics. Nat Methods. 2022Jan;19(1):41–50.

12. Krishna C, DiNatale RG, Kuo F, Srivastava RM, Vuong L, Chowell D, et al. Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy. Cancer Cell. 2021 May [cited 2024 May 31];39(5):662–677.e6. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1535610821001653.

13. Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE, Zappia L, Madissoon E, et al. An integrated cell atlas of the lung in health and disease. Nat Med. 2023 Jun [cited 2024 May 31];29(6):1563–77. Available from: https://www.nature.com/articles/s41591-023-02327-2.

14. Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science. 2022 May 13 [cited 2024 May 31];376(6594):eabl5197. Available from: https://www.science.org/doi/10.1126/science.abl5197.

15. MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat Commun. 2018 Oct 22 [cited 2024 May 31];9(1):4383. Available from: https://www.nature.com/articles/s41467-018-06318-7.

16. Guilliams M, Bonnardel J, Haest B, Vanderborght B, Wagner C, Remmerie A, et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. Cell. 2022 Jan [cited 2024 May 31];185(2):379–396.e38. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867421014811.

17. Xing X, Yang F, Huang Q, Guo H, Li J, Qiu M, et al. Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. Sci Adv. 2021 Jan 29 [cited 2024 May 31];7(5):eabd9738. Available from: https://www.science.org/doi/10.1126/sciadv.abd9738.

18. Bischoff P, Trinks A, Obermayer B, Pett JP, Wiederspahn J, Uhlitz F, et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. Oncogene. 2021 Dec 16 [cited 2024 May 31];40(50):6748–58. Available from: https://www.nature.com/articles/s41388-021-02054-3.

19. Chen K, Wang Q, Liu X, Tian X, Dong A, Yang Y. Immune profiling and prognostic model of pancreatic cancer using quantitative pathology and single-cell RNA sequencing. J Transl Med. 2023 Mar 21 [cited 2024 May 31];21(1):210. Available from: https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-023-04051-4.

20. Xu J, He B, Carver K, Vanheyningen D, Parkin B, Garmire LX, et al. Heterogeneity of neutrophils and inflammatory responses in patients with COVID-19 and healthy controls. Front Immunol. 2022 Nov 16 [cited 2025 Jan 1];13:970287. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2022.970287/full.

21. He B, Xiao Y, Liang H, Huang Q, Du Y, Li Y, et al. ASGARD is a single-cell guided pipeline to aid repurposing of drugs. Nat Commun. 2023Feb 22;14(1):993.

22. Romero JM, Grünwald B, Jang GH, Bavi PP, Jhaveri A, Masoomian M, et al. A four-chemokine signature is associated with a T-cell–inflamed phenotype in primary and metastatic pancreatic cancer. Clin Cancer Res. 2020 Apr 15 [cited 2024 May 31];26(8):1997–2010. Available from: https://aacrjournals.org/clincancerres/article/26/8/1997/83103/A-Four-Chemokine-Signature-Is-Associated-with-a-T.

23. Liang Z, Yu J, Gu D, Liu X, Liu J, Wu M, et al. M2-phenotype tumour-associated macrophages upregulate the expression of prognostic predictors MMP14 and INHBA in pancreatic cancer. J Cell Mol Med. 2022 Mar [cited 2024 Feb 27];26(5):1540–55. Available from: https://onlinelibrary.wiley.com/doi/10.1111/jcmm.17191.

24. De Andrade LF, Lu Y, Luoma A, Ito Y, Pan D, Pyrdol JW, et al. Discovery of specialized NK cell populations infiltrating human melanoma metastases. JCI Insight. 2019 Dec 5 [cited 2024 Feb 27];4(23):e133103. Available from: https://insight.jci.org/articles/view/133103.

25. Mitchell A, Rentero C, Endoh Y, Hsu K, Gaus K, Geczy C, et al. LILRA5 is expressed by synovial tissue macrophages in rheumatoid arthritis, selectively induces pro-inflammatory cytokines and IL-10 and is regulated by TNF-α, IL-10 and IFN-γ. Eur J Immunol. 2008 Dec [cited 2024 Mar 25];38(12):3459–73. Available from: https://onlinelibrary.wiley.com/doi/10.1002/eji.200838415.

26. Lin X, Zhou Y, Xue L. Mitochondrial complex I subunit MT-ND1 mutations affect disease progression. Heliyon. 2024 Apr [cited 2024 Dec 23];10(7):e28808. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2405844024048394.

27. Rivadeneira DB, Delgoffe GM. Antitumor T-cell reconditioning: improving metabolic fitness for optimal cancer immunotherapy. Clin Cancer Res. 2018 Jun 1 [cited 2024 Dec 23];24(11):2473–81. Available from: https://aacrjournals.org/clincancerres/article/24/11/2473/80785/Antitumor-T-cell-Reconditioning-Improving.

28. Field CS, Baixauli F, Kyle RL, Puleston DJ, Cameron AM, Sanin DE, et al. Mitochondrial integrity regulated by lipid metabolism is a cell-intrinsic checkpoint for treg suppressive function. Cell Metab. 2020 Feb [cited 2024 Dec 23];31(2):422–437.e5. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1550413119306655.

29. Chen C, Peng J, Ma S, Ding Y, Huang T, Zhao S, et al. Ribosomal protein S26 serves as a checkpoint of T-cell survival and homeostasis in a p53-dependent manner. Cell Mol Immunol. 2021 Jul [cited 2024 Dec 23];18(7):1844–6. Available from: https://www.nature.com/articles/s41423-021-00699-4.

30. Kamata M, Tada Y. Dendritic cells and macrophages in the pathogenesis of psoriasis. Front Immunol. 2022 Jun 28 [cited 2024 Dec 23];13:941071. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2022.941071/full.

31. Xu L, Chen Y, Liu L, Hu X, He C, Zhou Y, et al. Tumor-associated macrophage subtypes on cancer immunity along with prognostic analysis and SPP1-mediated interactions between tumor cells and macrophages. Tang Y, editor. PLOS Genet. 2024 Apr 22 [cited 2024 Dec 23];20(4):e1011235. Available from: https://dx.plos.org/10.1371/journal.pgen.1011235.

32. Yu Q, Shi X, Wang H, Zhang S, Hu S, Cai T. A novel prognostic signature of comprising nine NK cell signatures based on both bulk RNA sequencing and single-cell RNA sequencing for hepatocellular carcinoma. J Cancer. 2023 [cited 2024 Dec 23];14(12):2209–23. Available from: https://www.jcancer.org/v14p2209.htm.

33. Ye H, Zhou Q, Zheng S, Li G, Lin Q, Wei L, et al. Tumor-associated macrophages promote progression and the Warburg effect via CCL18/NF-kB/VCAM-1 pathway in pancreatic ductal adenocarcinoma. Cell Death Dis. 2018 Apr 18 [cited 2024 Dec 23];9(5):453. Available from: https://www.nature.com/articles/s41419-018-0486-0.

34. Zhong W, Lu Y, Han X, Yang J, Qin Z, Zhang W, et al. Upregulation of exosome secretion from tumor-associated macrophages plays a key role in the suppression of anti-tumor immunity. Cell Rep. 2023 Oct [cited 2024 Dec 23];42(10):113224. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2211124723012366.

35. Khushman M, Patel GK, Laurini JA, Bhardwaj A, Roveda K, Donnell R, et al. Exosomal markers (CD63 and CD9) expression and their prognostic significance using immunohistochemistry in patients with pancreatic ductal adenocarcinoma. J Gastrointest Oncol. 2019 Aug [cited 2024 Dec 23];10(4):695–702. Available from: http://jgo.amegroups.com/article/view/22798/21578.

36. Jewett A, Kos J, Kaur K, Safaei T, Sutanto C, Chen W, et al. Natural killer cells: diverse functions in tumor immunity and defects in pre-neoplastic and neoplastic stages of tumorigenesis. Mol Ther - Oncolytics. 2020 Mar [cited 2024 Dec 23];16:41–52. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2372770519301007.

37. Min Y, Huang R, Zhang H, Yang Q, Zhang Q, Chen D. 884P prognostic value and immune characteristics of LGALS1 in head and neck squamous cell carcinoma. Ann Oncol. 2023 Oct [cited 2024 Dec 23];34:S567. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0923753423028661.

38. F Murphy J. Modulation of angiogenesis by tumor associated macrophages in the tumor microenvironment. MOJ Immunol. 2014 Jul 24 [cited 2024 Dec 23];1(3). Available from: https://medcraveonline.com/MOJI/modulation-of-angiogenesis-by-tumor-associated-macrophages-in-the-tumor-microenvironment.html.

39. Schmieder A, Schledzewski K. The role of tumor-associated macrophages (TAMs) in tumor progression. In: Klink M, editor. Interaction of immune and cancer cells. Vienna: Springer Vienna; 2014. p. 49–74. Available from: https://doi.org/10.1007/978-3-7091-1300-4_3.

40. Marquardt N, Kekäläinen E, Chen P, Lourda M, Wilson JN, Scharenberg M, et al. Unique transcriptional and protein-expression signature in human lung tissue-resident NK cells. Nat Commun. 2019 Aug 26 [cited 2024 Dec 23];10(1):3841. Available from: https://www.nature.com/articles/s41467-019-11632-9.

41. Foroutan M, Molania R, Pfefferle A, Behrenbruch C, Scheer S, Kallies A, et al. The ratio of exhausted to resident infiltrating lymphocytes is prognostic for colorectal cancer patient outcome. Cancer Immunol Res. 2021 Oct 1 [cited 2024 Dec 23];9(10):1125–40. Available from: https://aacrjournals.org/cancerimmunolres/article/9/10/1125/665561/The-Ratio-of-Exhausted-to-Resident-Infiltrating.

42. Egelston CA, Guo W, Tan J, Avalos C, Simons DL, Lim MH, et al. Tumor-infiltrating exhausted CD8+ T cells dictate reduced survival in premenopausal estrogen receptor–positive breast cancer. JCI Insight. 2022 Feb 8 [cited 2024 Dec 23];7(3):e153963. Available from: https://insight.jci.org/articles/view/153963.

43. Xiao F, Shen J, Zhou L, Fang Z, Weng Y, Zhang C, et al. ZNF395 facilitates macrophage polarization and impacts the prognosis of glioma.

44. Kim HA, Kim H, Nam MK, Park JK, Lee MY, Chung S, et al. 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 Int. 2023 Sep 27 [cited 2024 Feb 27];23(1):219. Available from: https://cancerci.biomedcentral.com/articles/10.1186/s12935-023-03064-9.

45. Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N, et al. A novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. Cancer Res. 2016 Jul 1 [cited 2024 May 31];76(13):3838–50. Available from: https://aacrjournals.org/cancerres/article/76/13/3838/608326/A-Novel-MIF-Signaling-Pathway-Drives-the-Malignant.

46. Bacher M, Metz CN, Calandra T, Mayer K, Chesney J, Lohoff M, et al. An essential regulatory role for macrophage migration inhibitory factor in T-cell activation. Proc Natl Acad Sci. 1996 Jul 23 [cited 2024 May 31];93(15):7849–54. Available from: https://pnas.org/doi/full/10.1073/pnas.93.15.7849.

47. Tsang JCH, Vong JSL, Ji L, Poon LCY, Jiang P, Lui KO, et al. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. Proc Natl Acad Sci. 2017 Sep 12 [cited 2024 May 31];114(37). Available from: https://pnas.org/doi/full/10.1073/pnas.1710470114.

48. Red-Horse K, Zhou Y, Genbacev O, Prakobphol A, Foulk R, McMaster M, et al. Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface. J Clin Invest. 2004 Sep 15 [cited 2024 Apr 6];114(6):744–54. Available from: http://www.jci.org/articles/view/22991.

49. Suryawanshi H, Morozov P, Straus A, Sahasrabudhe N, Max KEA, Garzia A, et al. A single-cell survey of the human first-trimester placenta and decidua. Sci Adv. 2018 Oct 5 [cited 2024 May 31];4(10):eaau4788. Available from: https://www.science.org/doi/10.1126/sciadv.aau4788.

50. Tang CT, Zhang QW, Wu S, Tang MY, Liang Q, Lin XL, et al. Thalidomide targets EGFL6 to inhibit EGFL6/PAX6 axis-driven angiogenesis in small bowel vascular malformation. Cell Mol Life Sci. 2020 Dec [cited 2024 May 31];77(24):5207–21. Available from: http://link.springer.com/10.1007/s00018-020-03465-3.

51. Liang G, Zhou C, Jiang X, Zhang Y, Huang B, Gao S, et al. De novo generation of macrophage from placenta-derived hemogenic endothelium. Dev Cell. 2021 Jul [cited 2024 May 31];56(14):2121–2133.e6. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1534580721004834.

52. Wang K, Zheng J. Signaling regulation of fetoplacental angiogenesis. J Endocrinol. 2012 Mar [cited 2024 May 31];212(3):243–55. Available from: https://joe.bioscientifica.com/view/journals/joe/212/3/243.xml.

53. MacPhee DJ, Mostachfi H, Han R, Lye SJ, Post M, Caniggia I. Focal adhesion kinase is a key mediator of human trophoblast development. Lab Invest. 2001 Nov [cited 2024 May 31];81(11):1469–83. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0023683722031440.

54. Palaiologou E, Etter O, Goggin P, Chatelet DS, Johnston DA, Lofthouse EM, et al. Human placental villi contain stromal macrovesicles associated with networks of stellate cells. J Anat. 2020 Jan [cited 2024 May 31];236(1):132–41. Available from: https://onlinelibrary.wiley.com/doi/10.1111/joa.13082.

55. Ueshima C, Kataoka TR, Osakabe M, Sugimoto A, Ushirokawa A, Shibata Y, et al. Decidualization of stromal cells promotes involvement of mast cells in successful human pregnancy by increasing stem cell factor expression. Front Immunol. 2022 Jan 31 [cited 2024 May 31];13:779574. Available from: https://www.frontiersin.org/articles/https://doi.org/10.3389/fimmu.2022.779574/full.

56. Huang Q, Liu Y, Du Y, Garmire LX. Evaluation of cell type annotation R packages on single-cell RNA-seq data. Genomics Proteomics Bioinformatics. 2021 Apr 1 [cited 2025 Jan 1];19(2):267–81. Available from: https://academic.oup.com/gpb/article/19/2/267/7229882.

57. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16 [cited 2024 May 31];8(1):14049. Available from: https://www.nature.com/articles/ncomms14049.

58. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun [cited 2024 May 31];184(13):3573–3587.e29. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867421005833.

59. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov 1 [cited 2024 May 31];27(21):2987–93. Available from: https://academic.oup.com/bioinformatics/article/27/21/2987/217423.

60. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, Abecasis GR, Steering committee, Altshuler DM, et al. A global reference for human genetic variation. Nature. 2015 Oct 1 [cited 2023 Jul 27];526(7571):68–74. Available from: https://www.nature.com/articles/nature15393.

61. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv; 2012 [cited 2024 Dec 24]. Available from: http://arxiv.org/abs/1207.3907.

62. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. The Innovation. 2021 Aug [cited 2023 Sep 5];2(3):100141. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2666675821000667.

63. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes.

64. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. 2023 [cited 2024 Aug 27]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.11.05.565674.

65. Unjitwattana T, Huang Q, Yang Y, Tao L, Yang Y, Zhou M, et al. Originator. Github. https://github.com/lanagarmire/Originator (2025).

66. Unjitwattana T, Huang Q, Yang Y, Tao L, Yang Y, Zhou M, et al. Single-cell RNA-seq data have prevalent blood contamination but can be rescued by Originator, a computational tool separating single-cell RNA-seq by genetic and contextual information. 2025. Zenodo. https://doi.org/10.5281/zenodo.14750794.

67. Sikkema L, Ramírez-Suástegui C, Strobl D, Gillett T, Zappia L, Madissoon E, et al. The integrated Human Lung Cell Atlas (HLCA) v1.0. Datasets. Human Cell Atlas. 2023. https://data.humancellatlas.org/hca-bio-networks/lung/atlases/lung-v1-0.

68. Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Datasets. Human Cell Atlas. 2023. https://explore.data.humancellatlas.org/projects/04e4292c-f62f-4098-ae9b-fd69ae002a90.

69. MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Datasets. Human Cell Atlas. 2024. https://explore.data.humancellatlas.org/projects/4d6f6c96-2a83-43d8-8fe1-0f53bffd4674.

70. Guilliams M, Bonnardel J, Haest B, Vanderborght B, Wagner C, Remmerie A, et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. Datasets. Chan Zuckerberg CELLxGENE. 2022. https://cellxgene.cziscience.com/collections/74e10dc4-cbb2-4605-a189-8a1cd8e44d8c.

71. Xing X, Yang F, Huang Q, Guo H, Li J, Qiu M, et al. Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. Datasets. Curated Cancer Cell Atlas. 2021. https://www.weizmann.ac.il/sites/3CA/lung.

72. Bischoff P, Trinks A, Obermayer B, Pett JP, Wiederspahn J, Uhlitz F, et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. Datasets. Curated Cancer Cell Atlas. 2021. https://www.weizmann.ac.il/sites/3CA/lung.

Unjitwattana *et al. Genome Biology*   (2025) 26:52

Page 13 of 13

73. Krishna C, DiNatale RG, Kuo F, Srivastava RM, Vuong L, Chowell D, et al. Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy. Datasets. Curated Cancer Cell Atlas. 2021. https://www.weizmann.ac.il/sites/3CA/kidney.

74. Chen K, Wang Q, Liu X, Tian X, Dong A, Yang Y. GSE212966: single-cell RNA-seq reveals immune landscape of pancreatic cancer. Datasets. Gene Expression Omnibus. 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE212966.

75. Tsang JCH, Vong JSL, Poon LCY, Jiang P, Lui KO, Ni YB, et al. EGAS00001002449: integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. Datasets. European Genome-phenome Archive. 2017. https://ega-archive.org/studies/EGAS00001002449.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.