ABC portal: a single-cell database and web server for blood cells

Xin Gao [ID]*, Fang Hong, Zhenyu Hu, Zilong Zhang, Yang Lei, Xiaoyun Li and Tao Cheng*

State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China

ABSTRACT

ABC portal (http://abc.sklehabc.com) is a database and web portal containing 198 single-cell transcriptomic datasets of development, differentiation and disorder of blood/immune cells. All the datasets were re-annotated with a manually curated and unified single-cell reference, especially for the haematopoietic stem and progenitor cells. ABC portal provides web-based interactive analysis modules, especially a comprehensive cell-cell communication analysis and disease-related gene signature analysis. Importantly, ABC portal allows customized sample selection based on a combination of several metadata for downstream analysis and comparison analysis across datasets. ABC portal also allows users to select multiple cell types for analysis in the modules. Together, ABC portal provides an interactive interface of single-cell data exploration and re-analysis with customized analysis modules for the researchers and clinicians, and will facilitate understanding of haematopoiesis and blood/immune disorders.

INTRODUCTION

Blood/immune cells are involved in the normal and dys-regulated functions of all human tissues. Haematopoiesis and blood/immune disorder research has been significantly driven by the breakthrough of single-cell technology (1,2). High-resolution discrimination of the haematopoiesis lineage and differentiation trajectories were discovered (3–8). The development of the haematopoietic system was meticulously depicted from fetus to adult (8,9), covering many haematopoietic lineages, such as haematopoietic stem cells (10–12), macrophages (13), dendritic cells (14–16), megakaryocytes (17), innate lymphoid cells (18) and T cells (19). Meanwhile, heterogeneity of many types of blood/immune cells was further unveiled with single-cell

sequencing, such as neutrophils (20), natural killer cells (21,22), T cells (23) and myeloid cells (24,25). On the other hand, novel cellular and regulatory mechanisms of blood disease progression and therapy have been revealed with single-cell analysis. It was demonstrated that the heterogeneous of malignant blood cells and stem cells underlie the disease initiation and progression of acute myeloid leukemia (AML) (26,27), chronic myeloid leukemia (CML) (28), acute lymphoid leukemia (ALL) (29), myeloproliferative neoplasms (MPN) (30–32), aplastic anaemia (AA) (33,34), lymphoma (35), and multiple myeloma (MM) (36,37). Sub-cell-type specific drug response and resistance mechanisms were uncovered for AML (38), CLL (39) and MM (37,40–42). Together, the massive studies using single-cell methods become a large valuable resource for haematology and immunology research. The effective re-access and re-analysis of the resource will benefit the experimental researchers and clinicians with weak computational skills in the field of both haematology and immunology.

Many general single-cell databases were developed, including Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) of Broad Institute, Single Cell Expression Atlas (https://www.ebi.ac.uk/gxa/sc/home) of EMBL-EBI, SCPortalen, PanglaoDB, scRNASeqDB, SCDevDB, SC2disease, cancerSCEM and TISCH (43–49). BloodSpot is a database of gene expression of haematopoietic cells in bulk samples. It was updated with several single-cell/purified bulk datasets recently (50). However, a haematology specific, interactive re-analysis allowed database of single-cell transcriptome is still lacking. As non-solid tumor malignancies, blood cancer was seldom collected by cancer immune databases (47,49,51), due to the entanglement of the malignant cells and immune cells. In addition, lacking cell-type annotation of many downloaded datasets hinders the effective reuse of the published datasets.

Here, we built ABC portal, a blood/immune cell specific database and tool with curated annotation for blood cell type, especially hematopoietic stem/progenitor cells (HSPCs), at single-cell resolution. ABC portal provides four interactive modules for the exploration and analysis of

*To whom correspondence should be addressed. Tel: +86 22 2390 9006; Fax: +86 22 2390 9006; Email: gaoxin1@ihcams.ac.cn
Correspondence may also be addressed to Tao Cheng. Email: chengtao@ihcams.ac.cn

gene expression, cellular composition, gene signatures and cell-cell communication. 111 datasets of haematopoiesis and development and 87 datasets of blood disease were collected, all of which passed through a uniform analysis pipeline. Curated cell type reference was generated from integrating four and six datasets for human and mouse, respectively. All the datasets in ABC were re-annotated uniformly by the two references, which facilitates cross-dataset comparison. More importantly, ABC allows the re-analysis of user-selected subsets of data in one dataset or comparison across datasets. Thus, ABC provides a comprehensive and professional database and tool for the haematology and immunology community.

MATERIALS AND METHODS

Data collection

We searched PubMed for the literature related to single-cell RNA-seq and haematology. The search query was 'Search: ((('single-cell RNA') OR (scRNA-seq)) AND (hemato* OR leukemia OR blood) AND (human OR mouse OR patient OR murine) NOT (review) filters: Associated data'. The literature was then manually confirmed if the data were publicly available. We downloaded the data from the Gene Expression Omnibus, ArrayExpress, FigShare (52) and Zendo (https://zenodo.org/). Through manually reviewing each literature and supplemental materials, we curated the meta-information of each dataset, including species, cancer type, technology, tissue source, flow cytometry gates and perturbations. Curated meta-information and processing steps of each dataset were elaborated in the database.

Data quality control and batch effect correction

We downloaded the quantification matrix (count or TPM) of gene expression. If the quantification matrix is not available, we downloaded the raw FASTQ files and processed them with Cell Ranger 3.0.2. Cells with a total count of >500, gene numbers >200 and a percentage of mitochondrial gene counts $<20\%$ were kept. Human genes were lifted over to hg38 and Ensembl v105, and mouse genes were lifted over to mm10 and Ensembl v105 using biomaRt v2.38.0 (53). Each dataset was normalized and integrated across samples using Harmony v0.1.0 (54).

Reference generation

Four datasets from Atlas of Human Blood Cells (ABC) (3), fetal bone marrow (8) and Human Cell Atlas (HCA) were integrated to generate the reference for human, including nine samples derived from fetal bone marrow, adult bone marrow and cord blood (Table S1). Six bone marrow datasets (4,5,55–57) were integrated to generate the reference for mouse (Table S1). Cell filter criteria for each dataset was shown in Table S1. DoubletFinder v2.0.3 (58) was used to remove doublets for 10X platform generated datasets. After quality control (Table S1), mitochondrial genes and genes expressed in <10 cells were removed. To integrate datasets, 2000 highly variable genes were selected for each sample by FindVariableFeatures from Seurat v3.1.1 (59,60)

and 2000 highly variable genes were selected for each dataset and then total datasets by SelectIntegrationFeatures from Seurat. Datasets were integrated using Harmony v0.1.0 (54) with 50 PCs to remove batch effects of samples and platforms.

After integration, cells expressing two types of canonical marker genes were also removed as doublets. In detail, data were first clustered into >200 small clusters using Seurat with parameters 'dims = 1:60, reduction = 'harmony', k.param = 10, resolution = 15'. If any one of the canonical marker genes of a lineage (Table S2) was expressed in $>50\%$ of cells (30% for mouse) in a small cluster and its mean expression was more than a heuristic value, one-tenth of 99% quantile of the marker gene expression level in all cells, the small cluster was tagged by the lineage. A small cluster with two types of lineage tags was recognized as a doublet cluster. Clusters expressed both T and natural killer (NK) cell markers, both megakaryocyte and myeloid cell markers or both NK and myeloid cell markers were kept.

Doublet removed datasets were roughly clustered by lineages and then each lineage was finely clustered by Seurat using Louvain algorithm. The cell type annotations were determined based on the marker genes of each cluster identified by Seurat and the original cell annotation from the publications. Cluster similarity and cell cycle were checked to confirm the cell clustering and cell-type annotation using function BuildClusterTree and CellCycleScoring from Seurat.

Cell type re-annotation

Malignant cells were recognized using inferCNV v1.7.1 (61). Normal control samples in each dataset were used as the control for CNV recognition. If not available, randomly selected 2000 cells from the HCA bone marrow datasets (or mouse dataset (57)) in the reference were used as control. HLA genes and cell cycle-related genes (c5.all.v7.4.cyclegene from MSigDB (62)) were removed to exclude confounding segments. We used default parameters except for cutoff = 0.1 (cutoff = 1 for Smartseq/Smart-seq2 platforms) and scaled_data = FALSE (scaled_data = TRUE for samples using external control) to run inferCNV.

To automate the malignant cell recognition, we clustered cells into five groups based on the CNV scores from inferCNV. For each group, the CNV scores were binarized, i.e. genes with the maximum or minimum CNV scores in each group were set to 1, and the rest scores were set to 0. In each group, genes with an averaged binary score of >0.15 were recognized as copy number changed genes. For a group with $>50\%$ of cells, in which 95% of genes were lowly expressed genes (total counts in all cells of the group <2), we increased the averaged binary score cutoff to 0.6. Cells in a group with 15 continuous copy number changed genes were defined as malignant cells.

The non-malignant cell types were re-annotated using scmap v1.4.1 (63) and the single-cell transcriptome reference generated as described above. scmapCluster mode of scmap was applied with the top 10 marker genes of each cell type in the reference dataset as feature genes and a similarity threshold of 0.4.

Score of signature gene set

Score of a selected gene set in signature module were calculated using AUCell v1.4.1 (64).

Ligand–receptor network

We built the ligand–receptor interaction network module, employing CellPhoneDB v2.0 (65) to predict the communications between cell types. Parameters of '– iterations = 1000 –subsampling –subsampling-log false – subsampling-num-cells 1000 –threshold 0.01' were used. The mean strength was normalized to the total strength predicted in each dataset at a user-defined p-value cutoff, then multiplexed by 10^4.

Web portal

The front end of the server was developed with VueJS v2.6.0 (https://vuejs.org/) and ViewUI v4.0.0 (https://www. iviewui.com/), and the back end was built in Java using the SpringBoot web framework v2.1.13. The server is hosted on a Linux Centos v7.0.1406 server running Apache Tomcat v9.0.31. Files of the raw data matrix, pre-processed data and intermediate results were stored in the local Ext4 file system. Metadata and the analysis results were stored in MySQL v8.0.18 database, which enables more efficient querying, visualizing and archiving of the datasets and analysis results. The interactive visualization diagrams were implemented with the D3.js (https://d3js.org/), Echarts.js (https://echarts.apache.org/) and CanvasXpress.js (http:// canvasxpress.org/).

RESULTS

Scheme of ABC portal

ABC portal currently contains 198 datasets from 150 publications and includes 12 blood disease types of human and mouse (Figure 1). We performed unified quality control, normalization and batch correction, and manually adjusted the meta-information for each dataset. Cell types of all datasets were re-annotated with the single-cell reference built in this study. We retained all types of cells in ABC database, including malignant cells and non-blood cells. Malignancy was labeled by implementing inferCNV (61) for the cancer datasets. ABC portal provides modules for gene expression, signature expression, cell composition and ligand-receptor network analysis (Figure 1). The label transfer, cell composition and ligand-receptor network were pre-calculated for each dataset, allowing fast access to the results in these modules. More importantly, users can select a subset of cells by metadata and/or cell types to perform these analyses and can compare across datasets.

Datasets in ABC portal

There are 122 human and 76 mouse datasets in the database currently, and the source types mainly include bone marrow, PBMC, cord blood, fetal liver and others (Figure 2A and B). There are 111 normal haematopoiesis and 87 disease-related datasets (Figure 2C). The disease-related datasets consist of 12 blood disorder types and 26 datasets of inflammation, infections and other disease related to blood and immune cells. Lymphoma, MM, AML and ALL were the most frequently studied haematological malignancies (Figure 2C).The normal haematopoiesis datasets mainly come from the research of hematopoietic hierarchy and hematopoietic system development, covering different types of hematopoietic lineages (Figure 2D). Meanwhile, our database contains datasets derived from a variety of platforms (Figure 2E).

Reference datasets

Out of total 198 datasets, 64 datasets have an original cell type annotation. To facilitate the reuse of the datasets, we provided a unified cell-type annotation for human and mouse datasets respectively in addition to their original annotations from publications. The unified annotation was label transferred from the single-cell transcriptome reference we built.

For the human reference, we integrated three bone marrow (BM) samples from Atlas of Human Blood Cells (ABC) project (3), two fetal bone marrow (FBM) samples (8) and two BM and two cord blood (CB) samples from Human Cell Atlas (HCA) to generate the reference for human. The integrated reference was partitioned into 44 clusters, and the cell identity of the clusters was determined by combining two types of information (Figure 3A). First, the cell annotation from the literature of ABC dataset, HCA dataset and FBM dataset (Figure S1). Second, the expression of canonical marker genes (Figure 3A–B and Supplementary Figure S2A). Comparing the cellular composition of BM, CB and FBM (Supplementary Figure S2B), GMP was enriched in both BM and FBM, HSC was enriched in CB and adult BM (Figure 3D–E). CD14 monocyte1 was enriched in FBM, CD14 monocyte2 was enriched in CB and adult BM, CD16+ monocyte was enriched in adult BM, and naïve B and naïve T cells were enriched in CB (Figure 3D–E and Supplementary Figure S2C–E).

For the mouse reference, we integrated 30 samples from six bone marrow datasets to generate the reference, and hematopoietic stem and progenitor cells were especially enriched in each dataset (4,5,55–57). The integrated reference was partitioned into 30 clusters and the cell identity of the clusters was determined by combining the cell annotation from the five literature and the expression of marker genes (Supplementary Figure S3A–C and Figure S4). The difference in cellular composition across the datasets were shown in Supplementary Figure S3D–E.

The references were used to pre-annotate the datasets in ABC portal. Users can switch to the consensus cell-type annotation by the button on each dataset page.

Sample selection and filter function

ABC portal allows dataset searching by species, tissue source, disease and publications (Figure 4A). An important feature of ABC portal is that dataset-specific filters and their combinations enable customized sample selection and re-analysis (Figure 4B). Dataset-specific filters can be selectively added using '+ Add' button (Figure 4B and C). Different combinations of sample filters, including tissue type,
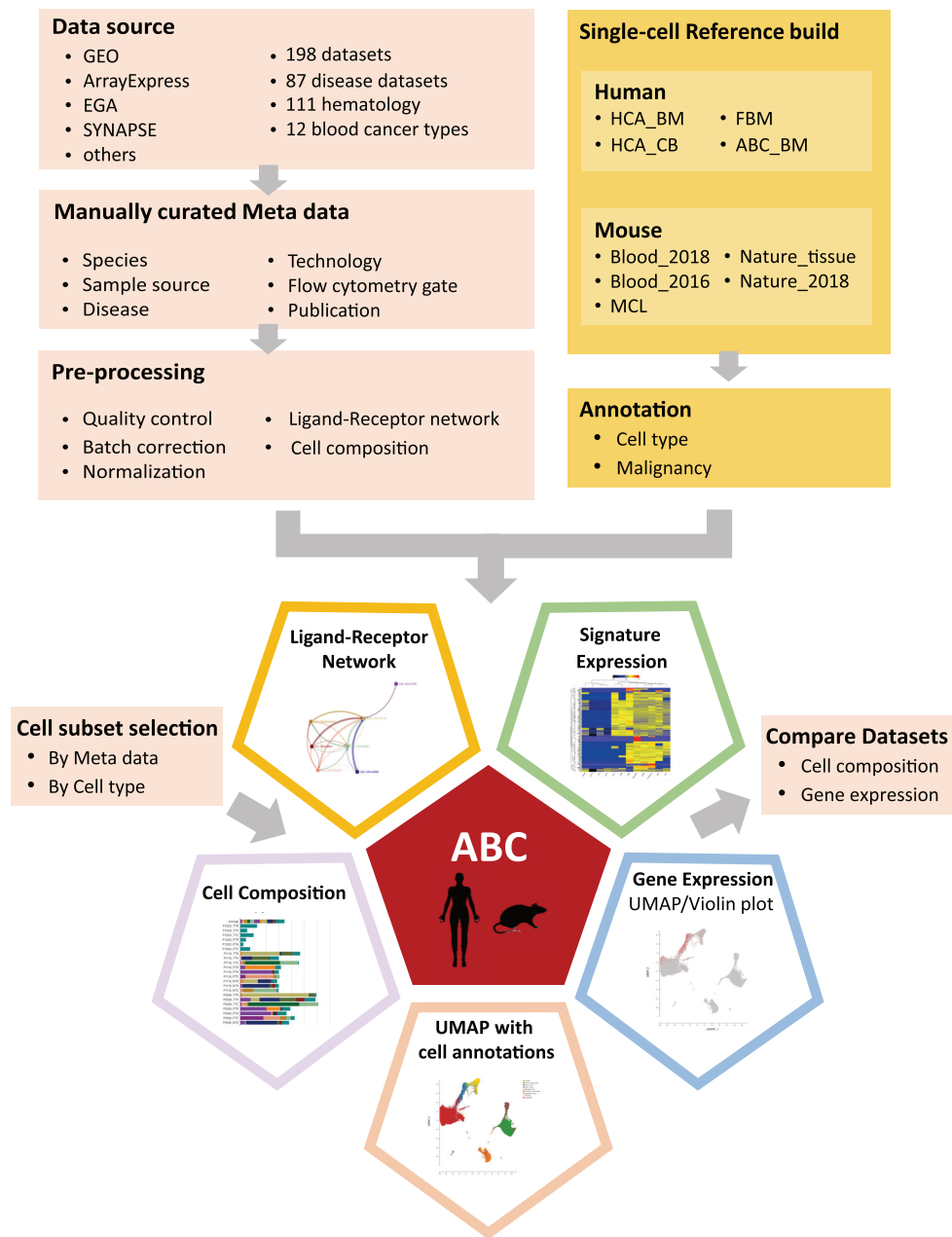
Figure 1. Scheme of ABC portal.

patient phenotype, flow cytometry gating, and treatment, are restored and can be applied to all analysis modules (Figure 4D). For instance, users can select flow cytometry sorted CD45 + cells from B-ALL samples for subsequent analysis and comparison (Figure 4B). In addition, metadata of each cell are provided in the 'Metainfo' tab and steps of data pre-processing are recorded in the 'Process' tab (Figure 4D).

Application modules

There are four interactive and user-friendly analysis modules developed in ABC portal.

UMAP module. This module contains two exploration functions (Figure 5A). One function is the multi-level cell

annotation, including original sample information, original cell type annotation (if available), unified cell-type annotation and malignancy (Figure 5B). Users can switch easily between original and unified annotations with a slider button (Figure 4B). Specially, we provide a switch button to show the possible cell type of malignant cells, which allows users to explore the characteristics and heterogeneity of malignant cells (Figure 5B). For example, it's shown that the malignant cells in B-ALL samples were mostly pre-pro-B and pro-B cells (Figure 5C). The other function is the gene expression visualization by both UMAP and violin plot (Figure 5D and E). For instance, DNTT, a marker gene of B progenitor cell, was highly expressed in the malignant cells in GSE153697 dataset (Figure 5D). These results con-
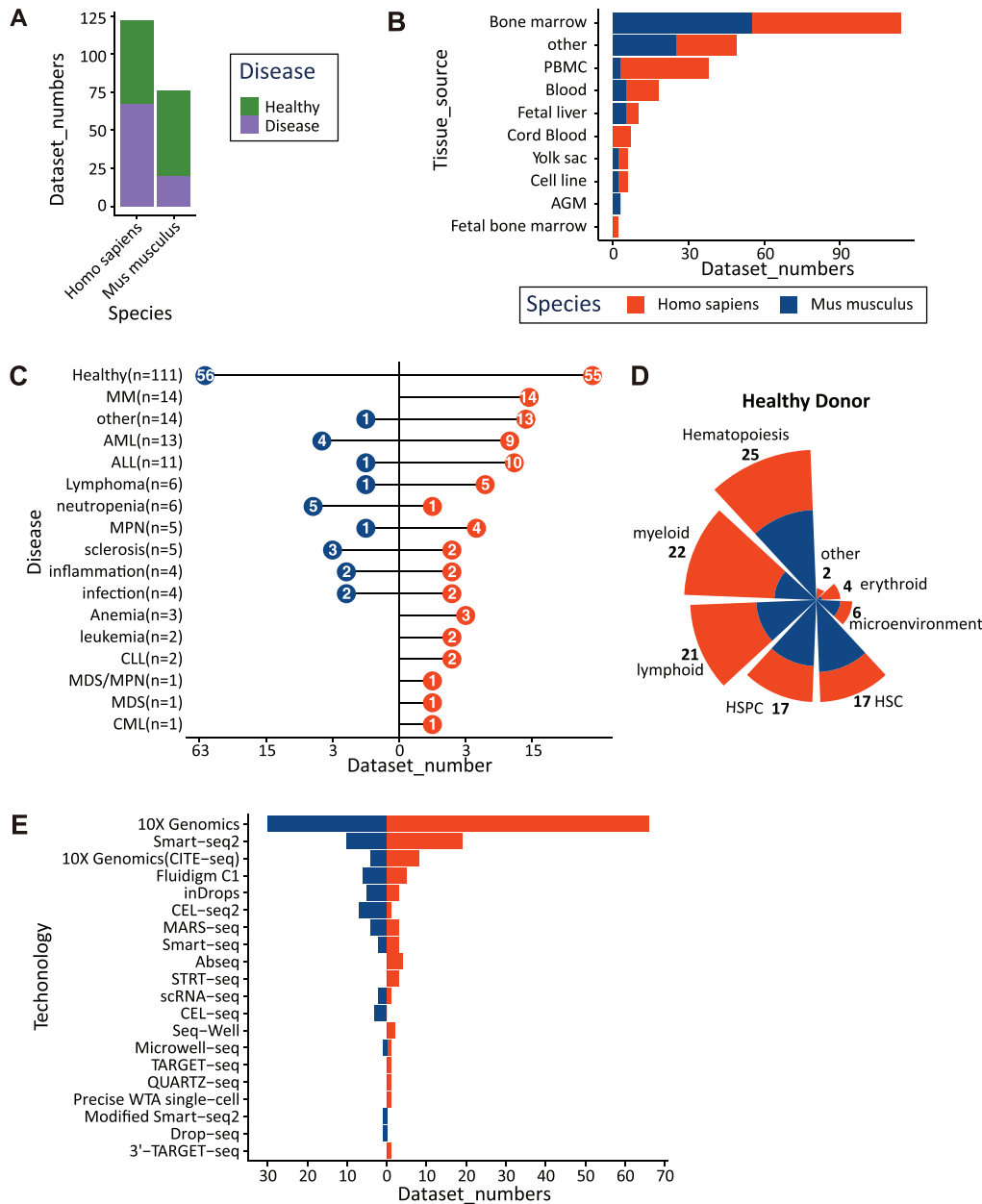
Figure 2. Statistics of datasets in ABC portal. (**A**) Number of datasets summarized by species. (**B**) Number of datasets summarized by tissue source. (**C**) Dataset summary by types of blood disorder. (**D**) Number of datasets from healthy conditions summarized by lineage. (**E**) Dataset summary by sequencing platforms.

firmed the accuracy of cell type and malignant cell annotation. Gene expression can also be compared across samples for a given cell type, across cell types or across malignancy (Figure 5E and Supplementary Figure S5A).

Composition module. The module shows the cellular composition in each sample or each patient (Figure S5B). The cross-sample difference in cellular composition can be compared. For example, in GSE130116 dataset, the proportion of myeloid cells including monocyte and dendritic cell in healthy samples is higher than those in primary B-ALL patients at the diagnosis stage, whereas the proportion of T/NK cells in patients is higher than in healthy samples.

These are consistent with the results reported in the original publication (29) (Supplementary Figure S5B).

Signature expression module. This module displays the scaled expression of genes in selected signatures and the AUCell score of the selected gene set. Besides the canonical gene sets, we collected 15 blood disorder-related gene sets from the literature to facilitate users' analysis, such as 'AML LSC + gene' signature (Figure 6A). This module allows the comparison of signature expression either across patients in a selected cell type or across cell types, indicating the signature variance across patients or the signature enrichment across cell types (Figure 6B and Sup-
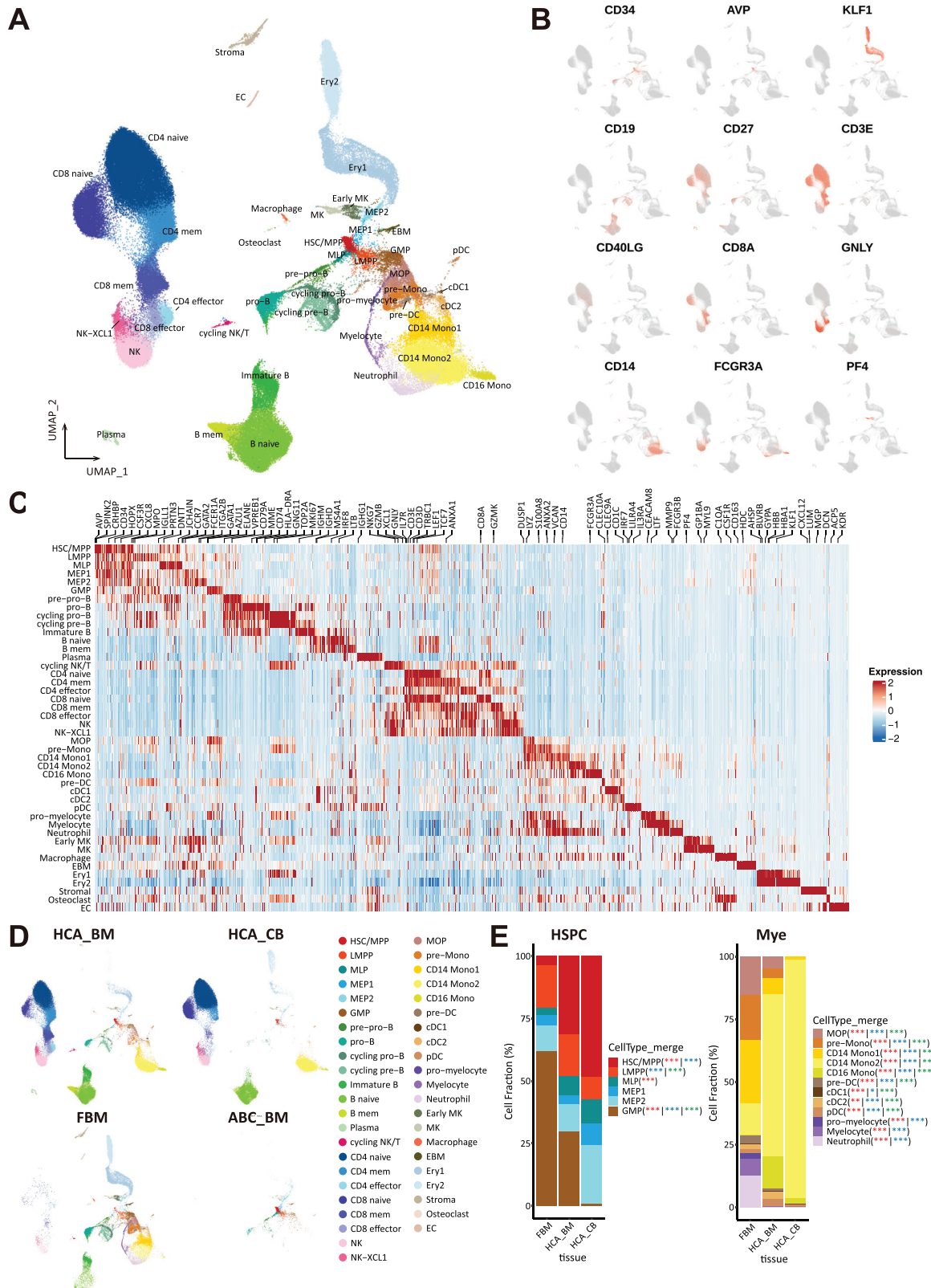
Figure 3. Reference for human datasets. (**A**) UMAP of cell types of the integrated reference dataset. (**B**) Feature plots for marker genes. (**C**) Heat map of marker gene expression for each cell type. (**D**) UMAP of reference split by dataset. (**E**) Bar plot of cell fractions in HSPC and myeloid cells. Statistical significance determined using chi-square test; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Red asterisk, FBM versus HCA_BM; blue asterisk, FBM versus HCA_CB; green asterisk, HCA_BM versus HCA_CB.
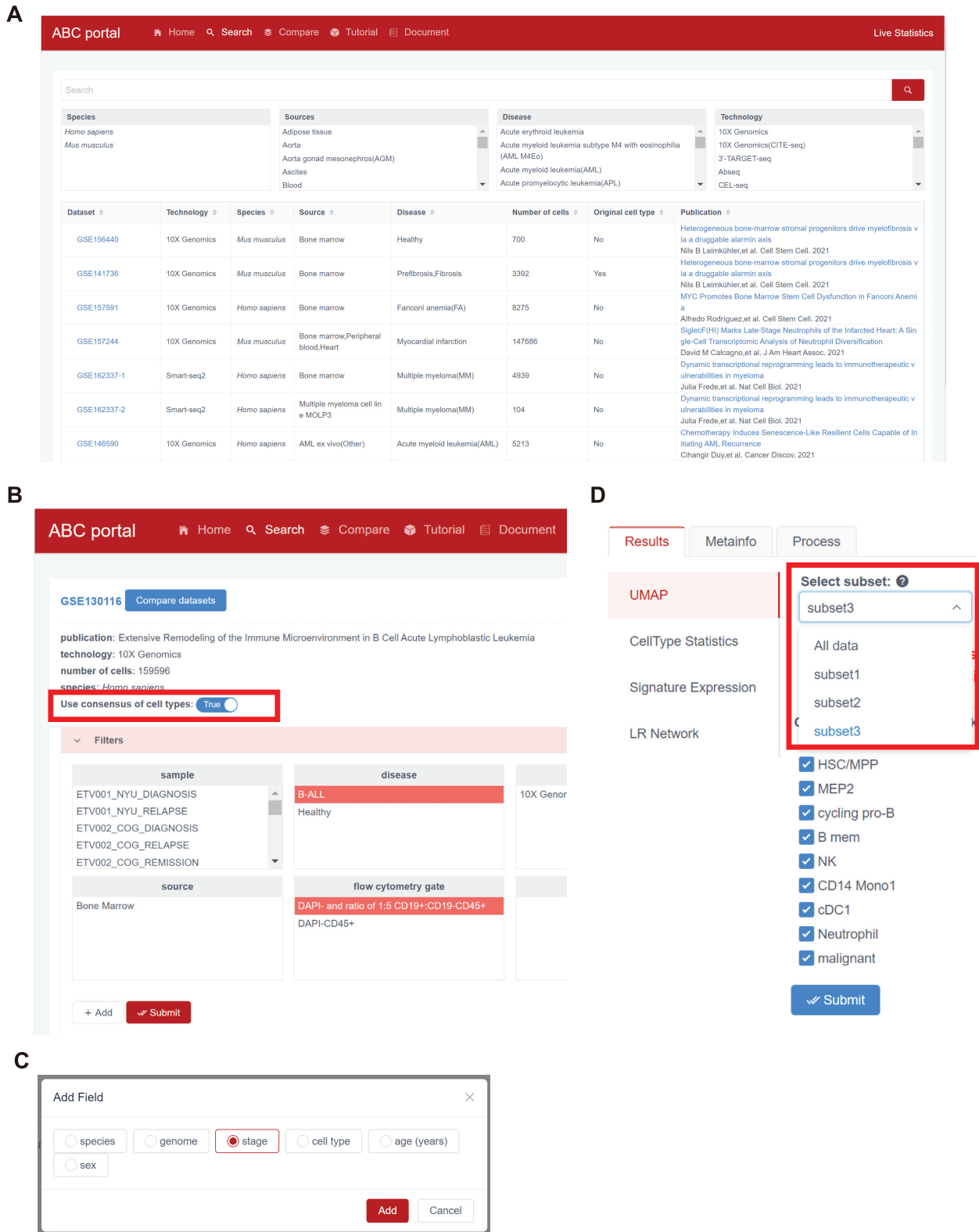
Figure 4. Data searching and exploration page. (**A**) Data searching page. (**B**) Data exploration page for the selected dataset. Red box highlights the switch for using consensus cell annotation. (**C**) Pop-up window for adding more filter conditions. (**D**) Data exploration page with results, metainfo and process tabs. Red box highlights the user-selected subsets of data saved for downstream analysis.
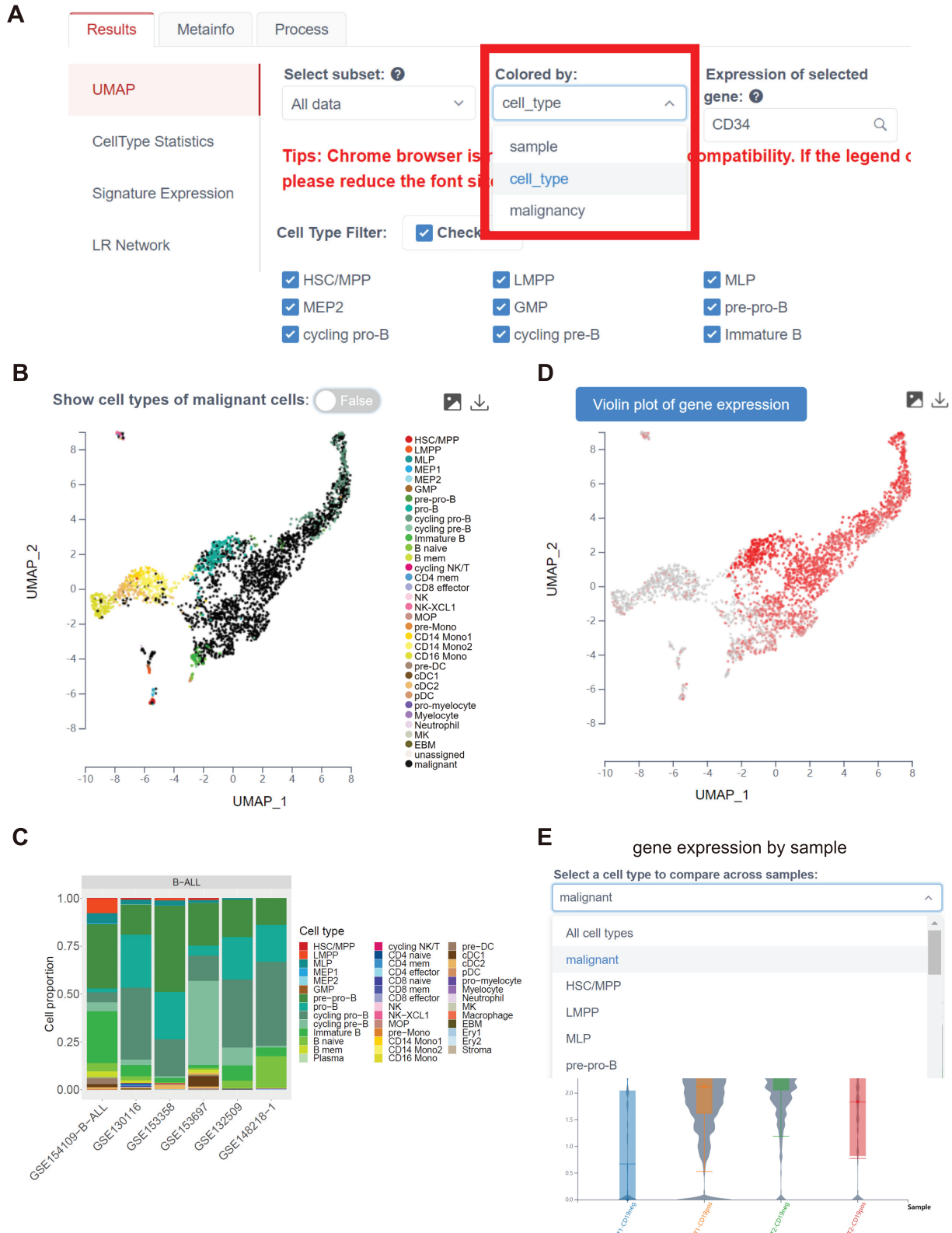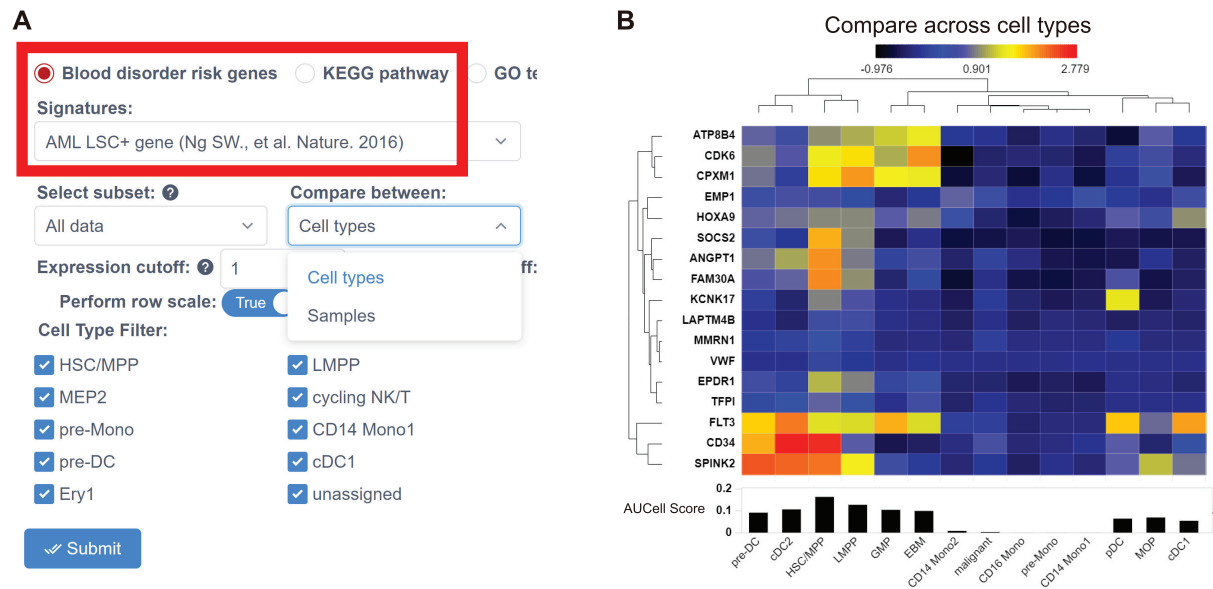
Figure 5. UMAP module. (**A**) UMAP module with different coloring modes (red box). (**B**) UMAP of cells colored by cell types. (**C**) Annotated cell types of malignant cells. (**D**) Gene expression of DNTT highlighted in UMAP. (**E**) Violin plot of gene expression by sample in selected cell type.
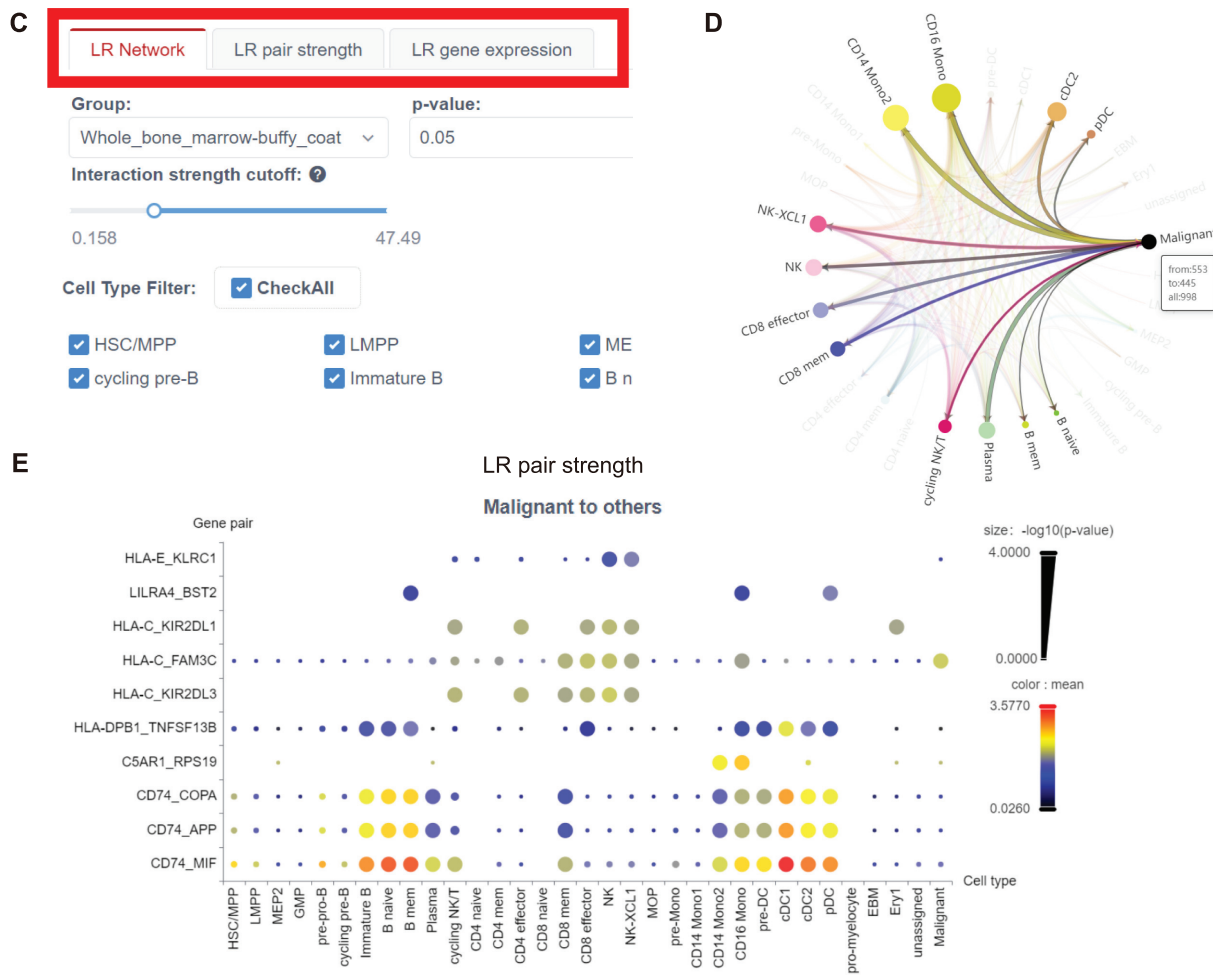
Figure 6. Signature expression module and LR network module. (**A**) Signature expression module with signatures of blood disorder-related genes (red box). (**B**) Signature expression comparing across cell types. (**C**) LR network module with three functional tabs (red box). (**D**) An example of LR network results. (**E**) LR strength from ligands of a selected cell type to receptors of other cell types.

Figure 7. Dataset comparison page. (**A**) Compare page (green box). (**B**) Dataset selection Pop-up. (**C**) Cell fraction comparison. (**D**) Gene expression comparison for a selected gene and cell type across samples from different datasets.

plementary Figure S5C). For example, the expression of 'AML LSC + gene' genes across cell types in AML samples shows the leukemia stem cell related genes were expressed highly in HSC/MPP, LMPP and GMP (Figure 6B). Using this module, we also found that the gene expression of glycolysis/gluconeogenesis pathway in yolk sac MK cells was higher than that in the fetal liver in GSE144024 dataset, which is consistent with the result reported in the original article (17) (Supplementary Figure S5C).

Ligand-receptor network module. The 'LR network' module exhibits the predicted ligand-receptor network of selected cell types (Figure 6C). There are three levels of LR analysis in this module: LR network, LR pair strength and LR gene expression. For 'LR network', users can customize and explore the LR network by filtering edge strength or changing network layout (Figure 6C and D). For 'LR pair strength', users can explore the strength of LR interaction pairs between a selected cell type and other cell types by switching the analysis type (Figure 6E and Supplementary Figure S5D). The interaction pairs are oriented from ligands to receptors and the strength plots are separated into two by the orientation. Finally, 'LR gene expression' exhibits the expression levels of the ligand and receptor genes. For example, LR network showed that malignant cells tend to interact with CD14 and CD16 monocytes in GSE161801 dataset, and LR pairs LILRA4-BST2, C5AR1-RPS19, CD74-MIF, LAIR1-LILRB4, CCL4-GPRC5D, TNFRSF17-TNFSF13B and ICAM1-AREG mediated the communication as reported (41). In addition, more interactions, such as GPR37-PSAP and CCL5-CCR1 that were not reported, were also discovered with the analysis and may be worth further study.

Comparison across datasets

ABC portal provides a convenient utility for cross-dataset comparison. Under the 'Compare' page, users can select multiple datasets to compare cell composition and gene expression across samples (Figure 7A and B). Cell counts and fractions for each cell type in each sample can be compared and be highlighted across samples (Figure 7C). Plot of standard deviation versus mean cell fraction indicates the highly variable cell types across samples. Gene expression can be compared across samples in a cell type of interest (Figure 7D). *P* value of Kruskal–Wallis test is shown to indicate the significance of difference.

DISCUSSION

Many single-cell databases were developed to collect various types of single-cell datasets. However, a blood/immune cell specific single-cell database with uniform cell type annotation and analysis utilities is not available. Here, we developed ABC portal, a comprehensive blood/immune single-cell data portal with meticulous cell annotation for researchers to explore and re-analyze the data. There are three special features of ABC. First, ABC is a blood/immune cell specific database with refined cell type annotation and manual curation. Second, ABC provides interactive analysis modules for data re-analyze, especially the cell-cell communication module. Third, ABC allows comparison between datasets and analysis of selected sample subsets.

In addition, as an important project of the consortium of Atlas of Blood Cells, ABC portal will be kept updating. In the future, ABC will integrate other single-cell omics data, such as scATAC-seq, protein data of CITE-seq and scDNA-seq. More analysis modules will be launched. In conclusion, ABC portal is a valuable resource and web tool for researchers to investigate haematopoiesis, blood disease and immunology.

DATA AVAILABILITY

ABC portal is a database and a web server available at http://abc.sklehabc.com.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Watcham,S., Kucinski,I. and Gottgens,B. (2019) New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. *Blood*, **133**, 1415–1426.
2. Brierley,C.K. and Mead,A.J. (2020) Single-cell sequencing in hematology. *Curr. Opin. Oncol.*, **32**, 139–145.
3. Xie,X., Liu,M., Zhang,Y., Wang,B., Zhu,C., Wang,C., Li,Q., Huo,Y., Guo,J., Xu,C. *et al.* (2021) Single-cell transcriptomic landscape of human blood cells. *Natl. Sci. Rev.*, **8**, nwaa180.
4. Nestorowa,S., Hamey,F.K., Pijuan Sala,B., Diamanti,E., Shepherd,M., Laurenti,E., Wilson,N.K., Kent,D.G. and Göttgens,B. (2016) A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood*, **128**, e20–e31.
5. Dahlin,J.S., Hamey,F.K., Pijuan-Sala,B., Shepherd,M., Lau,W.W.Y., Nestorowa,S., Weinreb,C., Wolock,S., Hannah,R., Diamanti,E. *et al.* (2018) A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in kit mutant mice. *Blood*, **131**, e1–e11.
6. Huang,P., Zhao,Y., Zhong,J., Zhang,X., Liu,Q., Qiu,X., Chen,S., Yan,H., Hillyer,C., Mohandas,N. *et al.* (2020) Putative regulators for the continuum of erythroid differentiation revealed by single-cell transcriptome of human BM and UCB cells. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 12868–12876.
7. Dong,F., Hao,S., Zhang,S., Zhu,C., Cheng,H., Yang,Z., Hamey,F.K., Wang,X., Gao,A., Wang,F. *et al.* (2020) Differentiation of transplanted haematopoietic stem cells tracked by single-cell transcriptomic analysis. *Nat. Cell Biol.*, **22**, 630–639.
8. Jardine,L., Webb,S., Goh,I., Quiroga Londoño,M., Reynolds,G., Mather,M., Olabi,B., Stephenson,E., Botting,R.A., Horsfall,D. *et al.* (2021) Blood and immune development in human fetal bone marrow and down syndrome. *Nature*, **598**, 327–331.

9. Popescu,D.-M., Botting,R.A., Stephenson,E., Green,K., Webb,S., Jardine,L., Calderbank,E.F., Polanski,K., Goh,I., Efremova,M. *et al.* (2019) Decoding human fetal liver haematopoiesis. *Nature*, **574**, 365–371.

10. Baron,C.S., Kester,L., Klaus,A., Boisset,J.-C., Thambyrajah,R., Yvernogeau,L., Kouskoff,V., Lacaud,G., van Oudenaarden,A. and Robin,C. (2018) Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat. Commun.*, **9**, 2517.

11. Zhu,Q., Gao,P., Tober,J., Bennett,L., Chen,C., Uzun,Y., Li,Y., Howell,E.D., Mumau,M., Yu,W. *et al.* (2020) Developmental trajectory of prehematopoietic stem cell formation from endothelium. *Blood*, **136**, 845–856.

12. Zeng,Y., He,J., Bai,Z., Li,Z., Gong,Y., Liu,C., Ni,Y., Du,J., Ma,C., Bian,L. *et al.* (2019) Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res.*, **29**, 881–894.

13. Bian,Z., Gong,Y., Huang,T., Lee,C.Z.W., Bian,L., Bai,Z., Shi,H., Zeng,Y., Liu,C., He,J. *et al.* (2020) Deciphering human macrophage development at single-cell resolution. *Nature*, **582**, 571–576.

14. Cytlak,U., Resteu,A., Pagan,S., Green,K., Milne,P., Maisuria,S., McDonald,D., Hulme,G., Filby,A., Carpenter,B. *et al.* (2020) Differential IRF8 transcription factor requirement defines two pathways of dendritic cell development in humans. *Immunity*, **53**, 353–370.

15. See,P., Dutertre,C.-A., Chen,J., Günther,P., McGovern,N., Irac,S.E., Gunawan,M., Beyer,M., Händler,K., Duan,K. *et al.* (2017) Mapping the human DC lineage through the integration of high-dimensional techniques. *Science*, **356**, eaag3009.

16. Balan,S., Arnold-Schrauf,C., Abbas,A., Couespel,N., Savoret,J., Imperatore,F., Villani,A.-C., Vu Manh,T.-P., Bhardwaj,N. and Dalod,M. (2018) Large-Scale human dendritic cell differentiation revealing notch-dependent lineage bifurcation and heterogeneity. *Cell Rep.*, **24**, 1902–1915.

17. Wang,H., He,J., Xu,C., Chen,X., Yang,H., Shi,S., Liu,C., Zeng,Y., Wu,D., Bai,Z. *et al.* (2021) Decoding human megakaryocyte development. *Cell Stem Cell*, **28**, 535–549.

18. Liu,C., Gong,Y., Zhang,H., Yang,H., Zeng,Y., Bian,Z., Xin,Q., Bai,Z., Zhang,M., He,J. *et al.* (2021) Delineating spatiotemporal and hierarchical development of human fetal innate lymphoid cells. *Cell Res.*, **31**, 1106–1122.

19. Park,J.-E., Botting,R.A., Domínguez Conde,C., Popescu,D.-M., Lavaert,M., Kunz,D.J., Goh,I., Stephenson,E., Ragazzini,R., Tuck,E. *et al.* (2020) A cell atlas of human thymic development defines t cell repertoire formation. *Science*, **367**, eaay3224.

20. Xie,X., Shi,Q., Wu,P., Zhang,X., Kambara,H., Su,J., Yu,H., Park,S.-Y., Guo,R., Ren,Q. *et al.* (2020) Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. *Nat. Immunol.*, **21**, 1119–1133.

21. Yang,C., Siebert,J.R., Burns,R., Gerbec,Z.J., Bonacci,B., Rymaszewski,A., Rau,M., Riese,M.J., Rao,S., Carlson,K.-S. *et al.* (2019) Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome. *Nat. Commun.*, **10**, 3931.

22. Crinier,A., Milpied,P., Escalière,B., Piperoglou,C., Galluso,J., Balsamo,A., Spinelli,L., Cervera-Marzal,I., Ebbo,M., Girard-Madoux,M. *et al.* (2018) High-Dimensional single-cell analysis identifies organ-specific signatures and conserved NK cell subsets in humans and mice. *Immunity*, **49**, 971–986.

23. Szabo,P.A., Levitin,H.M., Miron,M., Snyder,M.E., Senda,T., Yuan,J., Cheng,Y.L., Bush,E.C., Dogra,P., Thapa,P. *et al.* (2019) Single-cell transcriptomics of human t cells reveals tissue and activation signatures in health and disease. *Nat. Commun.*, **10**, 4706.

24. Mulder,K., Patel,A.A., Kong,W.T., Piot,C., Halitzki,E., Dunsmore,G., Khalilnezhad,S., Irac,S.E., Dubuisson,A., Chevrier,M. *et al.* (2021) Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity*, **54**, 1883–1900.

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

26. van Galen,P., Hovestadt,V., Wadsworth Ii,M.H., Hughes,T.K., Griffin,G.K., Battaglia,S., Verga,J.A., Stephansky,J., Pastika,T.J., Lombardi Story,J. *et al.* (2019) Single-Cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell*, **176**, 1265–1281.

27. Sachs,K., Sarver,A.L., Noble-Orcutt,K.E., LaRue,R.S., Antony,M.L., Chang,D., Lee,Y., Navis,C.M., Hillesheim,A.L., Nykaza,I.R. *et al.* (2020) Single-Cell gene expression analyses reveal distinct self-renewing and proliferating subsets in the leukemia stem cell compartment in acute myeloid leukemia. *Cancer Res.*, **80**, 458–470.

28. Giustacchini,A., Thongjuea,S., Barkas,N., Woll,P.S., Povinelli,B.J., Booth,C.A.G., Sopp,P., Norfo,R., Rodriguez-Meira,A., Ashley,N. *et al.* (2017) Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.*, **23**, 692–702.

29. Witkowski,M.T., Dolgalev,I., Evensen,N.A., Ma,C., Chambers,T., Roberts,K.G., Sreeram,S., Dai,Y., Tikhonova,A.N., Lasry,A. *et al.* (2020) Extensive remodeling of the immune microenvironment in b cell acute lymphoblastic leukemia. *Cancer Cell*, **37**, 867–882.

30. Rodriguez-Meira,A., Buck,G., Clark,S.-A., Povinelli,B.J., Alcolea,V., Louka,E., McGowan,S., Hamblin,A., Sousos,N., Barkas,N. *et al.* (2019) Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol. Cell*, **73**, 1292–1305.

31. Tong,J., Sun,T., Ma,S., Zhao,Y., Ju,M., Gao,Y., Zhu,P., Tan,P., Fu,R., Zhang,A. *et al.* (2021) Hematopoietic stem cell heterogeneity is linked to the initiation and therapeutic response of myeloproliferative neoplasms. *Cell Stem Cell*, **28**, 780.

32. Psaila,B., Wang,G., Rodriguez-Meira,A., Li,R., Heuston,E.F., Murphy,L., Yee,D., Hitchcock,I.S., Sousos,N., O'Sullivan,J. *et al.* (2020) Single-Cell analyses reveal megakaryocyte-biased hematopoiesis in myelofibrosis and identify mutant clone-specific targets. *Mol. Cell*, **78**, 477–492.

33. Yang,J., Tanaka,Y., Seay,M., Li,Z., Jin,J., Garmire,L.X., Zhu,X., Taylor,A., Li,W., Euskirchen,G. *et al.* (2017) Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors. *Nucleic Acids Res.*, **45**, 1281–1296.

34. Zhu,C., Lian,Y., Wang,C., Wu,P., Li,X., Gao,Y., Fan,S., Ai,L., Fang,L., Pan,H. *et al.* (2021) Single-cell transcriptomics dissects hematopoietic cell destruction and T-cell engagement in aplastic anemia. *Blood*, **138**, 23–33.

35. Milpied,P., Cervera-Marzal,I., Mollichella,M.-L., Tesson,B., Brisou,G., Traverse-Glehen,A., Salles,G., Spinelli,L. and Nadel,B. (2018) Human germinal center transcriptional programs are de-synchronized in b cell lymphoma. *Nat. Immunol.*, **19**, 1013–1024.

36. Ledergor,G., Weiner,A., Zada,M., Wang,S.-Y., Cohen,Y.C., Gatt,M.E., Snir,N., Magen,H., Koren-Michowitz,M., Herzog-Tzarfati,K. *et al.* (2018) Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat. Med.*, **24**, 1867–1876.

37. Cohen,Y.C., Zada,M., Wang,S.-Y., Bornstein,C., David,E., Moshe,A., Li,B., Shlomi-Loubaton,S., Gatt,M.E., Gur,C. *et al.* (2021) Identification of resistance pathways and therapeutic targets in relapsed multiple myeloma patients through single-cell sequencing. *Nat. Med.*, **27**, 491–503.

38. Riether,C., Pabst,T., Höpner,S., Bacher,U., Hinterbrandner,M., Banz,Y., Müller,R., Manz,M.G., Gharib,W.H., Francisco,D. *et al.* (2020) Targeting CD70 with cusatuzumab eliminates acute myeloid leukemia stem cells in patients treated with hypomethylating agents. *Nat. Med.*, **26**, 1459–1467.

39. Rendeiro,A.F., Krausgruber,T., Fortelny,N., Zhao,F., Penz,T., Farlik,M., Schuster,L.C., Nemc,A., Tasnády,S., Réti,M. *et al.* (2020) Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nat. Commun.*, **11**, 577.

40. Waldschmidt,J.M., Kloeber,J.A., Anand,P., Frede,J., Kokkalis,A., Dimitrova,V., Potdar,S., Nair,M.S., Vijaykumar,T., Im,N.G. *et al.* (2021) Single-Cell profiling reveals metabolic reprogramming as a resistance mechanism in BRAF-mutated multiple myeloma. *Clin. Cancer Res.*, **27**, 6432–6444.

41. Tirier,S.M., Mallm,J.-P., Steiger,S., Poos,A.M., Awwad,M.H.S., Giesen,N., Casiraghi,N., Susak,H., Bauer,K., Baumann,A. *et al.* (2021) Subclone-specific microenvironmental impact and drug response in refractory multiple myeloma revealed by single-cell transcriptomics. *Nat. Commun.*, **12**, 6960.

42. Liu,R., Gao,Q., Foltz,S.M., Fowles,J.S., Yao,L., Wang,J.T., Cao,S., Sun,H., Wendl,M.C., Sethuraman,S. *et al.* (2021) Co-evolution of tumor and immune cells during progression of multiple myeloma. *Nat. Commun.*, **12**, 2559.

43. Abugessaisa,I., Noguchi,S., Böttcher,M., Hasegawa,A., Kouno,T., Kato,S., Tada,Y., Ura,H., Abe,K., Shin,J.W. *et al.* (2018) SCPortalen: human and mouse single-cell centric database. *Nucleic Acids Res.*, **46**, D781–D787.

44. Franzén,O., Gan,L.-M. and Björkegren,J.L.M. (2019) PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)*, **2019**, baz046.

45. Cao,Y., Zhu,J., Jia,P. and Zhao,Z. (2017) scRNASeqDB: a database for RNA-Seq based gene expression profiles in human single cells. *Genes (Basel)*, **8**, 368.

46. Wang,Z., Feng,X. and Li,S.C. (2019) SCDevDB: a database for insights into single-cell gene expression profiles during human developmental processes. *Front. Genet.*, **10**, 903.

47. Sun,D., Wang,J., Han,Y., Dong,X., Ge,J., Zheng,R., Shi,X., Wang,B., Li,Z., Ren,P. *et al.* (2021) TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.*, **49**, D1420–D1430.

48. Zhao,T., Lyu,S., Lu,G., Juan,L., Zeng,X., Wei,Z., Hao,J. and Peng,J. (2021) SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res.*, **49**, D1413–D1419.

49. Zeng,J., Zhang,Y., Shang,Y., Mai,J., Shi,S., Lu,M., Bu,C., Zhang,Z., Zhang,Z., Li,Y. *et al.* (2022) CancerSCEM: a database of single-cell expression map across various human cancers. *Nucleic Acids Res.*, **50**, D1147–D1155.

50. Bagger,F.O., Kinalis,S. and Rapin,N. (2019) BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res.*, **47**, D881–D885.

51. Hong,F., Meng,Q., Zhang,W., Zheng,R., Li,X., Cheng,T., Hu,D. and Gao,X. (2021) Single-Cell analysis of the pan-cancer immune microenvironment and scTIME portal. *Cancer Immunol. Res.*, **9**, 939–951.

52. Singh,J. (2011) FigShare. *J. Pharmacol. Pharmacother.*, **2**, 138–139.

53. Durinck,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

54. Korsunsky,I., Millard,N., Fan,J., Slowikowski,K., Zhang,F., Wei,K., Baglaenko,Y., Brenner,M., Loh,P.-R. and Raychaudhuri,S. (2019) Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods*, **16**, 1289–1296.

55. Han,X., Wang,R., Zhou,Y., Fei,L., Sun,H., Lai,S., Saadatpour,A., Zhou,Z., Chen,H., Ye,F. *et al.* (2018) Mapping the mouse cell atlas by microwell-seq. *Cell*, **172**, 1091–1107.

56. Rodriguez-Fraticelli,A.E., Wolock,S.L., Weinreb,C.S., Panero,R., Patel,S.H., Jankovic,M., Sun,J., Calogero,R.A., Klein,A.M. and Camargo,F.D. (2018) Clonal analysis of lineage fate in native haematopoiesis. *Nature*, **553**, 212–216.

57. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group and Principal investigators (2018) Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, **562**, 367–372.

58. McGinnis,C.S., Murrow,L.M. and Gartner,Z.J. (2019) DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.*, **8**, 329–337.

59. Butler,A., Hoffman,P., Smibert,P., Papalexi,E. and Satija,R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, **36**, 411–420.

60. Stuart,T., Butler,A., Hoffman,P., Hafemeister,C., Papalexi,E., Mauck,W.M., Hao,Y., Stoeckius,M., Smibert,P. and Satija,R. (2019) Comprehensive integration of single-cell data. *Cell*, **177**, 1888–1902.

61. Patel,A.P., Tirosh,I., Trombetta,J.J., Shalek,A.K., Gillespie,S.M., Wakimoto,H., Cahill,D.P., Nahed,B.V., Curry,W.T., Martuza,R.L. *et al.* (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, **344**, 1396–1401.

62. Liberzon,A., Subramanian,A., Pinchback,R., Thorvaldsdóttir,H., Tamayo,P. and Mesirov,J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.

63. Kiselev,V.Y., Yiu,A. and Hemberg,M. (2018) scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods*, **15**, 359–362.

64. Aibar,S., González-Blas,C.B., Moerman,T., Huynh-Thu,V.A., Imrichova,H., Hulselmans,G., Rambow,F., Marine,J.-C., Geurts,P., Aerts,J. *et al.* (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*, **14**, 1083–1086.

65. Efremova,M., Vento-Tormo,M., Teichmann,S.A. and Vento-Tormo,R. (2020) CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.*, **15**, 1484–1506.