**Methods**

*Bulk mRNA data analysis*

The quality of raw FASTQ files were mapped to mouse reference GRCm38 using STAR two-pass alignment (v2.7.0), and transcript abundance estimates were performed using Kallisto (1), aligned to the same (GRCm38) reference genome. Kallisto transcript count data for each sample were concatenated, and transcript per million (TPM) data were reported for each gene after mapping gene symbols to ensemble IDs using R packages “tximport”, “tximportData”, and “ensembldb” Differential gene expression was estimated using DESeq2 (2). For data visualization and gene ranking, log fold changes were adjusted using lfcShrink in DESeq2, to minimize the effect size of poorly expressed genes. GSEA (v4.2.2)(3) was performed on preranked gene sets from differential expression between control and treated groups.

Single-cell mRNA read alignments and quality control

Single cells were isolated from patients, before receiving and after treatment. Samples were captured in a droplet emulsion using a Chromium Single-Cell instrument (10x Genomics) and libraries were prepared using the 10x Genomics 3' single-cell protocol. 10x libraries were pooled and sequenced on Illumina instruments. Reads from single cells were aligned to GRCh38 (GENCODE v32/Ensembl 98) human reference using Cell Ranger (version 6.0.0, 10x Genomics).  Low-quality cells with fewer UMIs or high mitochondria percentage were excluded from further analysis.

Cell clustering, doublet calling, and annotation

Gene expression counts were normalized, log-transformed, and scaled by 'SCTransform' in the R software package Seurat (4.0.0) (4). The principal components were calculated from the top 2000 variable genes and to generate Uniform Manifold Approximation and Projection (UMAP). Clusters of similar cells are detected using the original Louvain algorithm and UMAP coordinates to construct a shared nearest neighbor graph by 'FindNeighbors' function to obtain markers from each cluster.

Identify cell-type signatures

To find markers for each cell type, we performed differential expression analysis between each cluster against all other clusters, using the pairwise Wilcoxon test implemented in the 'FindMarkers' function. We considered genes with a Bonferroni-corrected q-value < 0.05. We also leveraged Blueprint (5), Encode (6), and azimuth reference dataset (<https://azimuth.hubmapconsortium.org/>) to annotate cell types. Identify doublets by searching cells with substantial and consistent expression profiles from two or more tissue compartments and/or cell types.

Pathway enrichment analysis of scRNA-seq data

Pre-ranked genes were obtained using Differential expression analysis in Seurat 4 using the Wilcoxon test. Statistical analyses were performed using the fgseaMultilevel function in the fgsea (1.16.0) R package (7). Visualization of gene sets was conducted using ggpubr (0.4.0) and ggplot2 (3.3.5) (8) R package. Differential analysis results were presented in heatmaps and volcano plots.

*Constructing cell development trajectories and RNA velocity*

To investigate the developmental trajectories of cells across multiple time points, and order them in pseudotime, we used the algorithms implemented in the Monocle 3 package (9-11). We can infer the source of malignant cells and the developmental progress among different sub-clusters. RNA velocity was analyzed by python tool velocyto (v0.17.17) (12). The ratio between intronic (unspliced) and exonic (spliced) mRNA in scRNA-seq data was used to estimate the “speed” of change in transcript abundances. We then used scVelo (0.2.5-dev) (13) to estimate per-cell velocities and visualized the finally averaged vector fields as a stream plot overlaid on the CCA integrated UMAP as described above.

1. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature biotechnology. 2016;34(5):525-7.

2. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.

3. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

4. Butler JM, Nolan DJ, Vertes EL, Varnum-Finney B, Kobayashi H, Hooper AT, et al. Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. Cell Stem Cell. 2010;6(3):251-64.

5. Stunnenberg HG, International Human Epigenome C, Hirst M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell. 2016;167(5):1145-9.

6. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

7. Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2019:060012.

8. Wickham H. ggplot2 - Elegant Graphics for Data Analysis | Hadley Wickham | Springer: Springer-Verlag New York; 2016.

9. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature biotechnology. 2014;32(4):381-6.

10. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nature methods. 2017;14(10):979-82.

11. Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. Nature methods. 2017;14(3):309-15.

12. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.

13. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nature Biotechnology. 2020;38(12):1408-14.