Supplementary Material

**Functional Subsets of Plasma Cells Associated with Amyloid Production and Venetoclax Sensitivity in Light Chain Amyloidosis**

Xuezhu Wang1,2,3,†, Hongxiao Han1,3,†, Jian Sun3,4,†, Qi Wang2, Xue-min Gao1,3, Kai-ni Shen1,3, Lu Zhang1,3, Yushan Zhao2, Xin-xin Cao1,3, Minping Qian5, Yang Chen2,\*, Jian Li1,3,\*

\* Correspondence: Jian Li, e-mail: lijian@pumch.cn

Yang Chen, e-mail: yc@ibms.pumc.edu.cn

**Table of Contents**

|  |  |
| --- | --- |
| Fig S1-S7 | Page 2-10 |
| Table S1 | Page 11-12 |
| Table S2 | Page 13 |
| Supplementary Methods | Page 14-16 |



**Supplementary Fig. S1**

1. Cell doublets identified by DoubletFinder in the bone marrow single-cell transcriptome.
2. UMAP embedding and clustering of the bone marrow single cells.
3. Up-regulated genes in each cluster of bone marrow single cells.
4. Immunoglobulin production in each cluster of bone marrow single cells.
5. Marker gene expression in each cluster of bone marrow single cells.
6. Cell type of each cluster of bone marrow single cells, inferred by SingleR.



**Supplementary Fig. S2**

1. UMAP embedding of the 14,081 single BMPCs, integrated and then labeled by patient origin.
2. Dendrogram showing the clustering of 14,081 single BMPCs at different resolutions.



**Supplementary Fig. S3**

1. Image flow cytometry analysis of BMPCs showing different APOE, STMN1, and B4GALT1 intensity.
2. Subcellular distribution of APOE, STMN1, and B4GALT1 in U-2 OS cell line demonstrated by the Human Protein Atlas.



**Supplementary Fig. S4**

**A.** Regulon specificity of transcription factors in each subset, inferred by SCENIC.



**Supplementary Fig. S5**

1. UMAP embedding of the merged transcriptome of BMPCs, split by subsets.
2. Inferred copy number variation of BMPCs in AL and MGUS, compared with healthy controls. Column: genomic region; Row: subsets and patient origin.



**Supplementary Fig. S6**

1. Volcano plot showing the differentially expressed genes in subset 2 in CCND1hi and CCND1low AL, compared with healthy controls.
2. Volcano plot showing the differentially expressed genes in subset 2 in CCND1hi and CCND1low AL, compared with MGUS.



**Supplementary Fig. S7**

1. Principal component analysis (PCA) of the bulk transcriptome of BMPCs in AL (n=32).
2. PCA of the bulk transcriptome of BMPCs in AL (n=29), after removing the outlier samples (n=3).
3. Heatmap showing the proportion of subsets in AL, clustered into 3 subgroups

**Table S1**

Overview of patient samples analyzed by scRNA-seq and image flow cytometry in this study.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Identifier** | **Age/Sex** | | **Description** | | **M protein (g/L)** | | **Light chain** | | **Differential free**  **light**  **chain (mg/L)** | | **iFish t(11;14)%** | | **BMPC%** | | **No. of single cells** | |
| pALPC1 | 76/M | pAL, kidney, *Mayo12-1* | | 10.8 | | lambda | | 168.6 | | 0.0 | | 3.5 | | 3675 | |
| pALPC2 | 52/M | pAL, heart, *Mayo12-4* | | 3.1 | | lambda | | 461.7 | | 8.5 | | 4.5 | | 1806 | |
| pALPC3 | 65/M | pAL, heart, *Mayo12-3* | | 11 | | lambda | | 104.6 | | 30.0 | | 2.5 | | 4190 | |
| MGUSPC1 | 73/M | MGUS | | 4.9 | | N.A. | | N.A. | | N.A. | | 3.5 | | 1550 | |
| MGUSPC2 | 55/F | MGUS | | 11.3 | | N.A. | | N.A. | | N.A. | | 12.5 | | 1789 | |
| normalPC | 35/F | Healthy controls | | N.A. | | N.A. | | N.A. | | N.A. | | N.A. | | 354 | |
| publicPC | (avg=55, sd=15.6)/M, (avg=46.3, sd=11.9)/F | Healthy controls | | N.A. | | N.A. | | N.A. | | N.A. | | N.A. | | 717 | |
| APOE,STMN1,B4GALT1 | 53/M | pALheart, *Mayo12-3* | | 1.03 | | lambda | | 818.0 | | N.A. | | 7.5 | | 6633,13285,3765 | |
| APOE\_rep1,STMN1\_rep1,B4GALT1\_rep1 | 40/F | pAL | | 19.03 | | lambda | | 582.5 | | 76.0 | | 16.5 | | 9742,22032,18406 | |
| APOE\_rep2 | 59/M | pAL, heart, kidney, *Mayo12-3* | | 1.68 | | lambda | | 658.5 | | 7.0 | | 3.0 | | 1624 | |
| STMN1\_rep2 | 53/M | pAL, heart, *Mayo12-3* | | N.A. | | lambda | | 366.0 | | 68.0 | | 8.0 | | 8835 | |
| B4GALT1\_rep2 | 54/M | pAL, heart, *Mayo12-4* | | 10.5 | | kappa | | 104.0 | | NA | | 6.0 | | 1564 | |

**Table S2**

Cell type determination was according to the following cell markers.

|  |  |  |
| --- | --- | --- |
| **Cell type** | **Cell marker** | |
| Plasma cells | CD138 (SDC1) (+), CD19 (-), CD38 (+), CD20 (MS4A1) (-), PRDM1(+), XBP1 (+), PAX5 (-), IRF4 (+), CXCR4 (+) |
| Immature B-cells | MS4A1(+), CXCR4 (+) |
| Monocytes | FCGR3A (+), TYROBP (+) |
| Macrophages | HMOX (+), TYROBP (+) |
| NK cells | NKG7(+), GNLY (+), TRAC (+), FCGR3A (+), IL7R (-) |
| CD8+ T-cells | NKG7 (+), IL7R (+), TRAC (+), MAL (-) |
| CD4+ T-cells | IL7R (+), MAL (+), TRAC (+) |

**Supplementary Methods**

**scRNA-seq data pre-processing**

We used fastp to perform quality control and trimming of the raw RNA sequencing data, then used CellRanger to align the sequence reads to the GRCh38 (hg38) reference genome. In total, 8 953 cells were successfully demultiplexed after pre-processing and filtering using CellRanger. The estimated number of cells in individual samples was as follows: normal PC (n=2 407), MGUSPC1 (n=4 280), MGUSPC2 (n=2 847), pALPC1 (n=4 328), pALPC2 (n=2 747), and pALPC3 (n=5 831).

**scRNA-seq data analysis using Seurat**

Seurat was used to process the single-cell data matrix. Genes expressed in <3 cells were removed. Low-quality cells were filtered out by selecting those expressing >100 and <2 500 genes, <5% mitochondria, and <1% haemoglobin. The mitochondrial and ribosomal genes were removed. For each sample, the single-cell transcriptome data matrix was log normalised under variable selection (top 3 000 variable genes), principal component analysis, and then clustered using the Louvain algorithm to identify cell clusters.

**Identification of BMPCs from bone marrow single-cell transcriptome**

We designed a pipeline to identify BMPCs in the bone marrow single-cell transcriptome. Cell clusters with doublets removed by DoubletFinder(1), determined as plasma cells based on cell markers (Table S2), identified as B-cells by SingleR(2), and expressing a significantly high percentage of immunoglobulin, were determined to be bone marrow plasma cells.

**Identification of functional subsets of single BMPC transcriptome**

We used the “FindIntegrationAnchors” function, wrapped in the Seurat R package, to integrate the samples based on the canonical correlation analysis algorithm. All BMPCs were clustered into seven clusters at a resolution of 0.4 by the Louvain algorithm.

**Pseudo-time and signalling entropy analysis**

Slingshot(3), an algorithm used to identify multiple pseudo-time trajectories in a single-cell transcriptome, was used to analyse the trajectories involving the seven functional subsets. The starting cluster was Cluster 6. We used the “CompSRana” function, wrapped in the LandSCENT(4), to calculate the signalling entropy of the seven BMPC subsets. Signalling entropy is a robust measure of differentiation potency.

**Inferred copy number variations**

InferCNV(5) was used to estimate copy number variation (CNV) based on a single-cell data matrix. normalPC and public PC were set as healthy controls. Random sampling of each BMPC group generated a cell subset for inferCNVs. Chromosomal expression patterns were estimated from the average gene expression with a moving 100 gene window. Parameter settings: denoise=TRUE, HMM=TRUE, cluster-by-groups=TRUE, cutoff value=0.1.

**Gene regulatory network analysis**

SCENIC(6) was used to infer regulons and transcription factors from the single-cell data matrix. Regulon activity and specificity scores were assigned to individual cells in each BMPC subset. Cytoscape was used to visualise the gene regulatory network composed of transcription factors and regulated genes in each regulon.

All 14 081 single BMPCs were included in one gene regulatory network for a general description of gene regulation in BMPC subsets. The BMPCs in healthy controls, MGUS, and AL were placed in separate gene regulatory networks to derive differences in transcription factors and regulons between these diseases.

**Differential analysis and protein–protein interaction network**

Differentially expressed genes of BMPC subgroups were generated by comparing the specific cluster with all other BMPC clusters, using the Wilcoxon test (BH adjustment), wrapped in the “FindAllMarkers” function in Seurat. Differentially expressed genes between the two BMPC groups were computed using the Wilcoxon test (BH adjustment), wrapped in the “FindMakers” function in Seurat.

STRING (<https://string-db.org/>) was used to infer protein–protein interactions (PPI) among the differentially expressed genes. Cytoscape was used to visualise the PPI network.

**Reference**

1. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell systems. 2019;8(4):329-37.e4.

2. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature immunology. 2019;20(2):163-72.

3. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC genomics. 2018;19(1):477.

4. Teschendorff AE, Enver T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. Nature communications. 2017;8:15599.

5. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science (New York, NY). 2014;344(6190):1396-401.

6. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nature methods. 2017;14(11):1083-6.