Dynamics and Plasticity of Immune Cells within Tumor Microenvironment

Chuijin Wei[1], Liaoliao Dong[1], Shumin Xiong[1], Ping Yu[1], Ren Zhou[1], Lin Cheng[1,*]

[1]Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China.

[*]Corresponding author: lincheng@shsmu.edu.cn (L. Cheng)

Abstract

Extensive research has been conducted on the heterogenicity of immune cells within the tumor microenvironment, like cancer cell heterogenicity, particularly with the emergence of single cell analysis. While inducing factors have been used to artificially alter immune cell fate *in vitro* and dynamic cancer cell plasticity has been recently discovered, it remains unknown whether tumor infiltrating immune cells acquire plasticity and dynamics that contribute to heterogenicity. In this study, we explored mitochondrial DNA mutation combining with chromosome single nucleotide polymorphism to construct phylogenetic trees of immune cells within multiple solid tumors, together with precise cell type and subtype definition based on single cell RNA sequencing data. Based on these lineage tracing landscapes, we systematically identified cell state transitions and fate changes among different immune cell subtypes and types within multiple solid tumors. Interestingly, immune cells demonstrated a high level of plasticity for transitioning between different states, transdifferentiating from one type to another or dedifferentiating to a progenitor stage, in varying frequencies across different cancers. Moreover, most of these cell state transitions and cell fate changes discovered here were previously unknown. The cell changes may

arise from extrinsic growth factors and cytokines secreted by tumor microenvironment cells, but intrinsic genetic mutations, particularly those related to ribosomes, may also be involved. Our data reveal that immune cell complexity extends beyond heterogenicity and also encompasses plasticity similar to that of cancer cells. Understanding the underlying mechanism of these cell changes will help elucidate the role of immune cells in cancer development and manipulating the cell change direction may ultimately enhance the efficiency of current immunotherapy.

Introduction

The immune cells present within the tumor microenvironment are known to play a crucial role in cancer development. Targeting these cells as a part of immunotherapy against cancer has yielded significant success in recent years; however, several questions still remain unanswered. With the advent of single-cell analysis, an increasing number of immune cell types and subtypes have been identified, contributing to the heterogeneity of immune cells within the tumor[1]. Different immune cell types have varying functions as pro-inflammatory or anti-inflammatory agents during cancer progression[2]. Furthermore, different subtypes of the same immune cell type have contradictory roles in promoting or inhibiting cancer cell proliferation, invasion, and metastasis, such as type 1 and 2 macrophages[3], and type 1 and 2 neutrophils[4,5].

The generation of heterogenous immune cells has traditionally been attributed to canonical hematopoiesis, a top-down hierarchical developmental process[6]. Once generated, immune cells typically acquire and maintain a specific function and phenotype, a widely accepted and well-known notion . However, the emergence of cell reprogramming techniques has challenged

this stereotype[7]. Mature immune cells can now be transdifferentiated from one type to another or dedifferentiated into immature progenitors or stem cells *in vitro* by overexpressing exogenous transcription factors or through chemical compound treatment[8-12]. Under pathological conditions, some immune cells can acquire plasticity and transition from one subtype to another[13]. In cancer, it has been reported that both natural killer (NK) cells and erythrocytes can transdifferentiate into myeloid-derived suppressor cells by cytokines[14,15], B-cell precursors can transdifferentiate into macrophage-like cells[16]. Given the complexity of the microenvironment and immune cell heterogeneity in cancer, whether other immune cells undergo dynamic state transitions or fate changes remains largely unknown.

To decode cell dynamics and construct trajectories of cell state transitions and fate changes, lineage tracing based on exogenous barcodes or endogenous genetic codes is the gold standard protocol. The former has been widely utilized in animal models or *in vitro* assays, while the latter method has been applied in analyzing human samples. In this paper, we combine two genetic tracing methods, mitochondrial DNA (mtDNA) mutation and chromosome single nucleotide polymorphism (SNP), to construct phylogenetic trees of immune cells in the tumor microenvironment across multiple cancers, based on single-cell RNA sequencing data. Combining precise cell type definition inferred from the same single cell, we establish a comprehensive landscape mapping the dynamics of immune cells within the tumor microenvironment (Fig. 1a). Based on these maps, we report numerous immune cell state transitions and fate changes for the first time, and analyze the underlying mechanisms potentially manipulating these dynamics, including intrinsic and extrinsic factors. This study not only provides the first comprehensive atlas

of the dynamics of tumor microenvironment immune cells, but also enhances our understanding of the generation of cell heterogeneity in cancer and can potentially improve the efficiency of immunotherapy by manipulating the direction of immune cell changes.

Results

Single Cell Type Definition Paired with Genetic Mutation to Construct Phylogenetic Tree

Accurately defining cell types and subtypes is essential for further analysis. In scRNA-seq (10×) data from human renal cell carcinoma (RCC) samples, $CD45^+$ cells were isolated and clustered into meta-clones and sub-clones. Using U-MAP analysis and classical immune cell markers (Extended Data Fig. 1a), we identified major immune cell types, including T cells, B cells, NK cells, monocytes, macrophages, neutrophils, dendritic cells (DCs), and common lymphoid progenitors (CLPs) (Fig. 1b). Furthermore, major clusters were subdivided into more precise sub-clusters, which have been reported to exhibit diverse and specific functions[17-19].

RNA velocity analysis revealed that the primary developmental direction followed differentiation from progenitors into either lymphoid or myeloid cells[20]. However, further analysis demonstrated that many immune cells did not follow this main direction. Velocity vectors for some cells in each cluster deviated in different directions, and even opposing directions (Extended Data Fig. 1b). Additionally, pseudotime analysis revealed that many cells clustered as the same cell type were not deposited together, but rather scattered in different branches defined as another cell type (Extended Data Fig. 1c). Unsupervised clustering of mtDNA mutations and SNP of these cells individually or jointly further confirmed that diverse cell types could be closely or even jointly clustered (Extended Data Fig. 1d and Fig. 1c), although most cells of the same cell types stayed

together. These findings indicate that mature immune cells within the tumor microenvironment may acquire plasticity and transition from one state to another or convert from one cell type to another, as previously reported.

To gain a better understanding of the precise transition or conversion direction between each immune cell pair, mtDNA mutations and SNP for each individual immune cell were inferred from scRNA-seq (10×) data[21,22]. A phylogram with lineage alleles of each cell was then constructed. Based on the lineage tracing protocol, the offspring cell should inherit the ancestor's mutation points and accumulate additional mutations. Using this logic, 4351 or 5047 cells in total 7792 cells were verified with a ten-level up-down hierarchy respectively (Extended Data Fig. 2 and Extended Data Fig. 3). To improve the accuracy of the hierarchies, two genetic trees were overlapped, and pair cells with high correlation appearing in both phylogenetic trees were retained and reconstructed (Fig. 1d,e). A five-level up-down hierarchy was obtained, containing 596 (P1 pRCC) or 806 (P2 cRCC) cells (Fig. 1f).

We generated a circus plot that represented every single immune cell type and their lineage relationships (Fig. 1g). It was apparent that cell expansion accounted for more than half of the relationships in most cell types, especially those with relatively higher cell numbers. State transitions between different cell subtypes within the same cell type were also observed. Surprisingly, cell transdifferentiation between different cell types occurred at a higher frequency compared to cell differentiation. Both state transitions and transdifferentiations were unidirectional or bidirectional. Additionally, a few mature immune cells could dedifferentiate into progenitors,

although at a low frequency. All of the above findings illustrate that the function and phenotype of immune cells within renal carcinoma are dynamic, and the single-cell endogenous genetic mutations can help to decisively depict the lineage tracing direction of these dynamics.

Dynamics of Immune Cells within Tumor microenvironment

It has been reported that diverse single cell RNA sequence methods can provide information on both cell type and genetic mutation. In this study, we applied our protocol to SMART-seq and scATAC-seq data and compared the cell type definition and phylogenetic tree among these two sources and scRNA-seq data[21]. We found that mtDNA coverage was comparable between SMART-seq and scATAC-seq, and both were better than scRNA-seq. Furthermore, the SNP counts from scATAC-seq over thousands were higher than those from scRNA-seq, with SMART-seq about hundreds having a lower count number (Extended Data Fig. 4a-h). When combining the genetic codes for each data type, we observed a few cell pairs in the phylogenetic tree from SMART-seq and several dozen cell pairs in scATAC-seq (Extended Data Fig. 4i-l). However, only a few cells with precise definition could be identified in the cell pairs from SMART-seq, and cells in the cell pairs from scATAC-seq were difficult to define, even though they could be clustered separately (Extended Data Fig. 4m-t). Therefore, we focused on analyzing the scRNA-seq data further.

To confirm our unexpected finding, we collected single cell RNA sequence data from various solid tumors, such as lung adenocarcinoma (LUAD), liver hepatocellular carcinoma (LIHC), intrahepatic cholangiocarcinoma (ICC), skin cutaneous melanoma (SKCM), bladder urothelial

carcinoma (BLCA), prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA), esophageal carcinoma (ESCA), breast invasive carcinoma (BRCA), pancreatic adenocarcinoma (PAAD), stomach adenocarcinoma (STAD), colon adenocarcinoma (COAD), and ovarian serous cystadenocarcinoma (OV) (Fig. 3a). In total, there were approximately 0.5 million immune cells (Extended Data Fig. 5a,b). Ten major cell clusters were identified, consisting of $CD4^+$ T cells, $CD8^+$ T cells, NK cells, B cells, monocytes, neutrophils, macrophages, mast cells, erythrocytes, and DCs. Additionally, most clusters were further divided into several sub-clusters (Extended Data Fig. 5c), as previously reported.

Individual dynamic landscapes of immune cell changes were generated for each of the tumors (Extended Data Fig. 6). Notably, the frequencies of immune cell transdifferentiation, state transition, and dedifferentiation were significantly higher in SKCM, COAD, LIHC, ICC, OV and LUAD compared to other tumors, which had a higher frequency of immune cell expansion (Fig 2b). The low frequency of immune cell differentiation and dedifferentiation could be attributed to the limited presence of hematopoietic stem or progenitor cells, including a few CLPs, in these tumors.

Immune cell expansion was observed across almost all immune cell types and subtypes, while differentiation was a well-known phenotype. Therefore, we focused on analyzing the other three immune cell fate changes in greater detail (Fig. 2c). It was evident that state transitions among $CD4^+$ T cell subtypes, $CD8^+$ T cell subtypes, macrophage subtypes, neutrophil subtypes, and monocyte subtypes occurred at a high frequency. Similarly, the frequencies of transdifferentiation

between $CD4^+$ T cells and $CD8^+$ T cells, as well as between macrophages and neutrophils, were also high (Fig. 2d and Extended Data Fig. 7a). Generally, both cell fate changes occurred at a higher frequency within lymphoid cells or myeloid cells. For example, transdifferentiation from $CD4^+$ T cells to $CD8^+$ T cells and vice versa was predominantly observed in SKCM, LIHC and BLCA, with a frequency of more than 10%. Transdifferentiations between macrophages and monocytes, macrophages and neutrophils, were mainly observed in LUAD, RCC and PAAD. However, transdifferentiation between these two major cell clusters occurred at a low frequency and was limited to minor cell subtypes, such as $CD4^+$ T Tm cells and monocytes, B cells and macrophages.

Interestingly, while conventional DCs hardly underwent transitions among their subtypes and transdifferentiated from or to other immune cell types, except for macrophages and monocytes, B cells exhibited higher cell fate change activity. Bidirectional cell fate changes between B cells and $CD4^+$ T Tm cells, $CD8^+$ T KIR^+ NK-like cells, $CD4^+$ T Treg, $CD4^+$ T Th17, and $CD8^+$ T $GZMK^+$ Tm cells were mainly observed in COAD. The change between B cells and $CD4^+$ T $TNFRSF9^-$ Treg cells occurred mainly in BRCA (Fig. 2e,f). In LUAD, the cell fate changes within myeloid cells, such as monocytes and macrophages, constituted the major part. Neutrophil transdifferentiation mainly occurred from or into macrophages in RCC, PRAD, and OV (Extended Data Fig. 7b).

Tumor Microenvironment Extrinsic Factors Contributing to Immune Cell Plasticity

Cells undergoing a fate change typically require a level of plasticity before the transformation can

occur[23]. In order to determine if immune cells possess this plasticity, we conducted a comparison of transcriptional signatures between lymphoid and myeloid cells. Through analysis of discrete values for each cell type and subtype, we observed that certain lymphoid cells, such as B cells, $CD4^+$ T cells, and $CD8^+$ T cells, had partially acquired myeloid cell signatures. Conversely, many macrophage and neutrophil subtypes had acquired lymphoid cell signatures (Fig.3a). These mixed transcriptional signatures suggest that these cells may possess developmental plasticity. The highest discrete value was observed in CLPs, while most DCs had the lowest value. These findings are consistent with the cell fate change pattern depicted in Figure 2c.

To further validate the generation and characteristics of plastic cells, we calculated the intermediate co-expression score of the starting and ending cells based on the cell change pairs identified earlier[24]. Specifically focusing on the cell changes between B cells and macrophages, we discovered intermediate cells labeled as intermediate cell (IC) 1 and IC2 from B cells to macrophage-APOE and macrophage-C1QA, respectively. Conversely, intermediate cells labeled as IC3 and IC4 were identified from macrophages to B cells (Fig. 3b). When compared to the original cells, B cell-specific gene expression was significantly suppressed in IC1 and IC2, which highly expressed *TYROBP*, *FCER1G*, *S100A11*, and *CST3*. No significantly expressed genes were found in IC3 and IC4, although macrophage-specific gene expression was inhibited (Fig. 3c). Principal component analysis demonstrated that there was no significant difference between diverse starting B and macrophage subtypes and IC subtypes (Fig. 3d). Ligand-receptor analysis revealed that these cells selectively interacted with microenvironmental cells, including malignant epithelial cells, fibroblasts, endothelial cells, and other immune cells, and were influenced by the

soluble factors secreted by these cells. Compared to B and macrophage cells that underwent expansion, the cells that underwent fate changes were generally influenced by TGFβ1, AGT, CCL3, INFγ, and TNF (Fig. 3e).

For cell change between $CD4^+$ T cells and $CD8^+$ T cells, some $CD4^+$ T cells acquire $CD8^+$ T cell characteristics and vice versa (Extended Data Fig.8a). Similarly, when NK cells transdifferentiate into $CD4^+$ T Tm cells or $CD8^+$ T $GZMK^+$ Tem cells, ICs are present in all transdifferentiating pairs and directions (Extended Data Fig. 8b,c). Intriguingly, the gene expression pattern of these ICs closely resembles that of natural killer T (NKT) cells[25], which express both NK cell markers (KLRD1, KLRF1, DUSP1) and T cell markers (CD3D, CD3E, IL32) (Extended Data Fig. 8d). It appears that TGFβ1, FGF2, and JAG1, combined with other factors, may play a role in generating NKT ICs from different cell types[26] (Extended Data Fig. 8e). Moreover, our data confirms previous report that TGFβ1 induces the state transition from $CD4^+$ T Th17 cells into $CD4^+$ T Treg cells (Extended Data Fig. 8f-h).

We also conducted a similar analysis to explore the transdifferentiation between neutrophils and macrophages/monocytes. Notably, monocytes displayed a limited capacity for plasticity (Extended Data Fig. 9a). In contrast, diverse macrophage subtypes exhibited significant activation of neutrophil-related genes, while simultaneously suppressing macrophage-specific genes in their initial ICs. Interestingly, ICs derived from neutrophils did not display any significant changes in gene expression patterns (Extended Data Fig. 9b,c). The top factors influencing these cell transitions were APOE and CSF1, along with FGF2, TGFβ1, INFγ, TNF, and CCL3 (Extended

Data Fig. 9d).

Intrinsic Genetic Mutation Affecting Immune Cell Fate Changes

Our next objective was to determine whether intrinsic genetic mutations played a role in modulating immune cell fate changes. Upon analyzing the top ten genes with high mutation frequencies, we discovered that they mainly consisted of human leukocyte antigen (HLA) genes, regardless of the immune cell fate change types (Extended Data Fig. 10a). After excluding gene mutations that appeared during cell expansion and differentiation, we observed that none of the previously detected HLA genes were present in the remaining cell fate changes (Fig. 4a). These findings suggest that although HLA gene mutations are frequent in immune cells within solid tumors as reported[27], they do not specifically contribute to immune cell plasticity.

Although some gene mutations related to clonal hematopoiesis and tumorigenesis were still present, such as *NPM1*, *STAG2*, *ARID5B*, and *NCOR1*[28,29], they appeared specially in starting immune cells underwent fate change but were at low frequencies (Fig. 4a). Gene ontology analysis of all identified mutated genes at three levels, including biological processes, molecular functions, and cellular components, revealed that ribosome-related gene mutations had a significant impact on immune cell fate changes (Fig. 4b). Of the approximately 80 ribosome protein genes[30], 32 were detected here, including RPLs and RPSs, with *RPLP2*, *RPL35A*, and *RPL34* appearing in multiple cell type fate changes (Fig. 4c). Ribosomes have been shown to regulate P53 function[31], which significantly impact cell fate change[32]. Although P53 mutation was not detected, its expression level was found to change in the cells with ribosome gene mutations. Moreover, the expression

levels of P53-targeted genes, including *ISCU*, *TRIM22*, and *PROM1*, were all positively or negatively correlated with diverse ribosome gene mutations (Fig. 4d).

To find out whether there were specific gene mutation or combination of these mutations could lead to generate specific cell types or subtypes, we further analyzed the ending cells after immune cell fate changes. Consistent with the analyzing results of starting cells, HLA genes were still at top mutate efficiency without discriminating cell fate changes (Extended Data Fig. 10a). For the cells generated through transdifferentiation, dedifferentiation and state transition, mutated ribosome genes were still as the main intrinsic factors, except for common cancer gene mutations, such as *BRCA1*, *EP300*, *EZH2* at low efficiency (Extended Data Fig. 10b). Comparing with other ribosome gene mutations, *RPL35A* and *RPLP2* mutation were not only at relatively higher frequency but also highly related with P53 target genes, such as *ENC1*, *RLK2*, *PHLDA3*, *TP53I3* etc positively and *TLR3*, *CPE*, *EPHA2*, *TP53INP1* negatively (Extended Data Fig. 10c-e).

Discussion

In this study, we generated phylogenetic trees for immune cells within multiple solid tumors based on SNP and mtDNA mutation inferred from single-cell RNA sequencing data, combined with cell type definitions. Our results demonstrate that immune cells could acquire plasticity and exist at highly dynamic state or stage (Fig. 4e). Generally, not only the cell subtypes within lymphoid cells or myeloid cells could transit to each other, but also the cell types could switch between each other. Comparing with other cells, B cells showed a relatively higher dynamic potential to change into multiple cell types, including T cells, macrophages and neutrophils, and can also be derived from these lineages. Conversely, DCs appear to maintain a more stable identity. Almost all these cell

state transitions or fate changes require the initial acquisition of plasticity becoming ICs with suppression of starting cell-specific genes and gradual establishment of an ending cell transcriptional signature. These changes were highly influenced by intrinsic genetic mutations, such as ribosome gene mutation with regulating P53-targeted genes, and mixed tumor microenvironment extrinsic factors, such as TGFβ, TNF and IFNγ, which have been reported to affect cell fate, especially epithelial-mesenchymal-transition. Further investigation is required to determine whether these influencing factors act independently or synergistically.

Previous studies have reported that immune cells within tumor or under pathological conditions could acquire plasticity, with a primary emphasis on subtype transition, particularly for neutrophils and macrophages with pro- and anti-inflammatory properties, classified as type I or type II[3-5]. Th17 could transdifferentiate into Treg cells by TGFβ modulating Smad3/AhR signaling during immune response[33]. Only a few reports have described cell changes between meta clusters. In a breast mouse model, $CSF1R^{+}Pax5^{Low}$ B-cell precursors transdifferentiated into macrophage-like cells, coaxed by cancer-secreted macrophage colony-stimulating factor[16]. In advanced cervical carcinoma and multiple animal tumor models, $CD45^{+}$ erythroid precursor cells gradually acquired myelopoietic potential and finally transdifferentiate into erythroid differentiated myeloid cells under the influence of granulocyte-macrophage colony-stimulating factor[14]. Both induced cells lead to a decline in T cell-mediated tumor suppression. Beyond these sporadic cases, our study provides a systemic view of all possible immune cell phenotype changes in pan-cancer, not only verifying previous reports but also discovering previously unknown changes.

Cell fate changes have been widely, artificially achieved both *in vitro* and *in vivo* by modulating transcription factor expression for not just solid tissue cells, but also immune cells. Additionally, microRNAs and small molecules that regulate epigenetics and canonical stemness-related signaling pathways have been used successfully to induce cell changes[34]. In addition to these key driving forces, cell culture conditions, such as oxygen pressure[35], and contextual stiffness[36], as well as microenvironment soluble factors, such as growth factors and cytokines[37], all affect cell fate change direction and efficiency. Given all these inducing factors existed in solid tumor as complex combination, it is unsurprising that so many immune cells acquire plasticity, undergo state transition and fate change, combining with widespread gene mutation in cancer. Investigating each cell fate change pairs and the mechanism controlling the changing direction not only helps understand the cell heterogenicity and plasticity in cancer but also enhances current immunotherapy[38].

The current gold standard for lineage tracing relies on genetic labelling through either fluorescent protein or genetic barcodes. Unfortunately, this method is limited to *in vitro* assays or animal models, making it challenging to utilize for human sample analysis[39]. Single-cell RNA and DNA sequencing analysis of the same cell is considered to be the optimal technique for lineage tracing analysis. However, this method is often expensive and low-throughput, which prevents widespread use[40]. Only genomic sequencing can provide a detailed genetic code that is necessary for constructing accurate phylogenetic trees, but losing cell type information. Multiple bioinformatic algorithms based on RNA sequencing, such as RNA velocity and pseudotime analysis, can help predict cell fate changes[41]. These methods are only prediction but not surely

establishing ancestor and decedent relationship among cells. SNP and mtDNA mutation inferred from single cell RNA sequence have been previously, individually explored to define cell genetic relationship, we combine them together for the first time. The immune cell phylogenetic trees generated here were based on strict filter criteria, and many cell change pairs reported before were also verified in our results. Nonetheless, future efforts to confirm other cell changes through canonical lineage tracing methods are still necessary.

Acknowledgements

This work is supported by National Natural Science Foundation of China (92068101), Shanghai Municipal Health Commission Grant Support (2022XD050) and Innovative Research Team of High-level Local Universities in Shanghai.

References

1 Ren, X. *et al.* Insights Gained from Single-Cell Analysis of Immune Cells in the Tumor Microenvironment. *Annual review of immunology* **39**, 583-609, doi:10.1146/annurev-immunol-110519-071134 (2021).

2 Nguyen, P. H. D. *et al.* Intratumoural immune heterogeneity as a hallmark of tumour evolution and progression in hepatocellular carcinoma. *Nature communications* **12**, 227, doi:10.1038/s41467-020-20171-7 (2021).

3 Cassetta, L. & Pollard, J. W. A timeline of tumour-associated macrophage biology. *Nature reviews. Cancer* **23**, 238-257, doi:10.1038/s41568-022-00547-1 (2023).

4 Hedrick, C. C. & Malanchi, I. Neutrophils in cancer: heterogeneous and multifaceted.

Nature reviews. Immunology **22**, 173-187, doi:10.1038/s41577-021-00571-6 (2022).

5 Xiong, S., Dong, L. & Cheng, L. Neutrophils in cancer carcinogenesis and metastasis. *Journal of hematology & oncology* **14**, 173, doi:10.1186/s13045-021-01187-y (2021).

6 Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science (New York, N.Y.)* **367**, doi:10.1126/science.aaw3381 (2020).

7 Sánchez Alvarado, A. & Yamanaka, S. Rethinking differentiation: stem cells, regeneration, and plasticity. *Cell* **157**, 110-119, doi:10.1016/j.cell.2014.02.041 (2014).

8 Riddell, J. *et al.* Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell* **157**, 549-564, doi:10.1016/j.cell.2014.04.006 (2014).

9 Zhang, M. *et al.* Transcription factor Hoxb5 reprograms B cells into functional T lymphocytes. *Nature immunology* **19**, 279-290, doi:10.1038/s41590-018-0046-x (2018).

10 Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663-676, doi:10.1016/s0092-8674(04)00419-2 (2004).

11 Zhou, Y. *et al.* Chemical Cocktail Induces Hematopoietic Reprogramming and Expands Hematopoietic Stem/Progenitor Cells. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)* **7**, 1901785, doi:10.1002/advs.201901785 (2020).

12 Qin, J. *et al.* Direct chemical reprogramming of human cord blood erythroblasts to induced megakaryocytes that produce platelets. *Cell stem cell* **29**, 1229-1245.e1227, doi:10.1016/j.stem.2022.07.004 (2022).

13 Wei, C., Yu, P. & Cheng, L. Hematopoietic Reprogramming Entangles with Hematopoiesis. *Trends in cell biology* **30**, 752-763, doi:10.1016/j.tcb.2020.07.006 (2020).

14 Long, H. *et al.* Tumor-induced erythroid precursor-differentiated myeloid cells mediate immunosuppression and curtail anti-PD-1/PD-L1 treatment efficacy. *Cancer cell* **40**, 674-693.e677, doi:10.1016/j.ccell.2022.04.018 (2022).

15 Park, Y. J. *et al.* Tumor microenvironmental conversion of natural killer cells into myeloid-derived suppressor cells. *Cancer research* **73**, 5669-5681, doi:10.1158/0008-5472.Can-13-0545 (2013).

16 Chen, C. *et al.* Cancer co-opts differentiation of B-cell precursors into macrophage-like cells. *Nature communications* **13**, 5376, doi:10.1038/s41467-022-33117-y (2022).

17 Cheng, S. *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792-809.e723, doi:10.1016/j.cell.2021.01.010 (2021).

18 Kourtis, N. *et al.* A single-cell map of dynamic chromatin landscapes of immune cells in renal cell carcinoma. *Nature cancer* **3**, 885-898, doi:10.1038/s43018-022-00391-0 (2022).

19 Zheng, L. *et al.* Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science (New York, N.Y.)* **374**, abe6474, doi:10.1126/science.abe6474 (2021).

20 Li, S. *et al.* A relay velocity model infers cell-dependent RNA velocity. *Nature biotechnology*, doi:10.1038/s41587-023-01728-5 (2023).

21 Ludwig, L. S. *et al.* Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* **176**, 1325-1339.e1322, doi:10.1016/j.cell.2019.01.022

(2019).

22 Xu, J. *et al.* Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *eLife* **8**, doi:10.7554/eLife.45105 (2019).

23 Zhao, T. *et al.* Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. *Cell stem cell* **23**, 31-45.e37, doi:10.1016/j.stem.2018.05.025 (2018).

24 Raghavan, S. *et al.* Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119-6137.e6126, doi:10.1016/j.cell.2021.11.017 (2021).

25 Bendelac, A., Savage, P. B. & Teyton, L. The biology of NKT cells. *Annual review of immunology* **25**, 297-336, doi:10.1146/annurev.immunol.25.022106.141711 (2007).

26 Flavell, R. A., Sanjabi, S., Wrzesinski, S. H. & Licona-Limón, P. The polarization of immune cells in the tumour environment by TGFbeta. *Nature reviews. Immunology* **10**, 554-567, doi:10.1038/nri2808 (2010).

27 Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nature biotechnology* **33**, 1152-1158, doi:10.1038/nbt.3344 (2015).

28 Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-339, doi:10.1038/nature12634 (2013).

29 Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine* **371**, 2477-2487, doi:10.1056/NEJMoa1409405 (2014).

30 Elhamamsy, A. R., Metge, B. J., Alsheikh, H. A., Shevde, L. A. & Samant, R. S.

Ribosome Biogenesis: A Central Player in Cancer Metastasis and Therapeutic Resistance. *Cancer research* **82**, 2344-2353, doi:10.1158/0008-5472.Can-21-4087 (2022).

31 Luan, Y. *et al.* Deficiency of ribosomal proteins reshapes the transcriptional and translational landscape in human cells. *Nucleic acids research* **50**, 6601-6617, doi:10.1093/nar/gkac053 (2022).

32 Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460**, 1140-1144, doi:10.1038/nature08311 (2009).

33 Gagliani, N. *et al.* Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* **523**, 221-225, doi:10.1038/nature14452 (2015).

34 Rao, S. *et al.* Inactivation of ribosomal protein L22 promotes transformation by induction of the stemness factor, Lin28B. *Blood* **120**, 3764-3773, doi:10.1182/blood-2012-03-415349 (2012).

35 Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T. & Yamanaka, S. Hypoxia enhances the generation of induced pluripotent stem cells. *Cell stem cell* **5**, 237-241, doi:10.1016/j.stem.2009.08.001 (2009).

36 Zhang, X. *et al.* Harnessing matrix stiffness to engineer a bone marrow niche for hematopoietic stem cell rejuvenation. *Cell stem cell* **30**, 378-395.e378, doi:10.1016/j.stem.2023.03.005 (2023).

37 Tang, Y., Xiong, S., Yu, P., Liu, F. & Cheng, L. Direct Conversion of Mouse Fibroblasts into Neural Stem Cells by Chemical Cocktail Requires Stepwise Activation of Growth Factors and Nup210. *Cell reports* **24**, 1355-1362.e1353,

doi:10.1016/j.celrep.2018.06.116 (2018).

38 Xiong, S., Feng, Y. & Cheng, L. Cellular Reprogramming as a Therapeutic Target in Cancer. *Trends in cell biology* **29**, 623-634, doi:10.1016/j.tcb.2019.05.001 (2019).

39 Yang, D. *et al.* Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* **185**, 1905-1923.e1925, doi:10.1016/j.cell.2022.04.015 (2022).

40 Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nature methods* **12**, 519-522, doi:10.1038/nmeth.3370 (2015).

41 Kester, L. & van Oudenaarden, A. Single-Cell Transcriptomics Meets Lineage Tracing. *Cell stem cell* **23**, 166-179, doi:10.1016/j.stem.2018.04.014 (2018).

Software and algorithms

CellRanger 3.1.0	10× Genomics	https://10xgenomics.com/
CellRanger-atac 2.1.0	10× Genomics	https://10xgenomics.com/
Seurat 4.1.1	Satija*, Farrell*, et al.,2015	https://satijalab.org/seurat/
NicheNet 1.1.1	Browaeys et al., 2020	https://github.com/saeyslab/nichenetr
ANNOVAR	Wang et al., 2010	https://doc-openbio.readthedocs.io/projects/annovar/en/latest/
Jupyter 1.0.0		https://jupyter.org/
Pysam 0.21.0	Li H et al., 2009	https://github.com/pysam-developers/pysam
Samtools 1.13	Petr Danecek et al., 2021	http://www.htslib.org/
RNA velocity 0.17	La Manno et al., 2018	https://github.com/velocyto-team/velocyto.py
scVelo 0.2.2	Bergen et al., 2020	https://github.com/theislab/scvelo/
Monocle 2.22.0	Cole Trapnell*, Davide Cacchiarelli*,et al., 2014	https://github.com/cole-trapnell-lab/monocle-release
Circlize	Gu, Z., 2014	https://jokergoo.github.io/circlize_book/book/
Harmony 0.1.1	Ilya Korsunsky, et al., 2019	https://github.com/immunogenomics/harmony
Hisat2 2.2.1	Kim, D., Paggi, J.M., Park, C. et al., 2019	https://github.com/DaehwanKimLab/hisat2
Htseq 2.0.2	Simon Anders, et al., 2014	https://github.com/htseq/htseq
Signac 1.7.0	Stuart et al., 2021	https://stuartlab.org/signac/
Bcftools 1.17	Li H et al., 2011	https://github.com/samtools/bcftools
clusterProfiler 3.16	G Yu et al., 2012	https://github.com/YuLab-SMU/clusterProfiler

METHOD DETAILS

Single-cell RNA-Seq datasets collected in this study

We collected raw single-cell sequencing data on tumor-infiltrating hematopoietic cells from approximately 150 samples across 19 projects, which were diagnosed as one of the 14 common tumor types defined by the TCGA database (Fig.2a, Table S1). Our team obtained the data by downloading them from the GEO, ENA, BioStudies, and GSA databases using SJTU HPC. The data included 10× scRNA-seq, 10× scATAC-seq, and SMART-seq. We only selected data that met

the following criteria: it had sequenced information on both the tumor cells and infiltrated immune cells, and the raw sequence data or bam files were available for download.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single-cell RNA-seq data processing

All upstream analysis processes for scRNA-seq data were carried out on the SJTU HPC platform. Fastq files or bam files were obtained from the Siyuan No.1 platform using the wget function, SRA toolkit, or Aspera software. Bam files were then converted to the fastq file format using the Cell Ranger (version 3.1.0) bamtofastq function. To obtain bam files and expression matrixs, we processed the fastq files through the 10× scRNA-seq Cell Ranger Count pipeline using the GRCh38 human reference genome. Only quality-passed cells that passed the Cell Ranger filtering step were included in downstream analysis.

Single-cell ATAC-seq data processing

We obtained the 10× scATAC-seq data from the GEO database and analyzed them using the 10× Cell Ranger-atac pipeline (version 2.1.0) against the GRCh38 human reference genome to obtain bam files and the atac peak matrix. Only quality-passed cells that passed the Cell Ranger-atac filtering step were included in downstream analysis.

SMART-seq data processing

The SMART-seq data were aligned using hisat2 software (version 2.2.1), which was indexed with the GRCh38 human reference genome. The output was in the form of sam files, which were then converted to sorted bam files using samtools (version 1.13). The expression matrixs were counted using htseq software (version 2.02). Only the quality-passed cells and matrixs were included in downstream analysis.

Dimension Reduction and Unsupervised Clustering

The single-cell RNA expression matrixs underwent dimension reduction and unsupervised clustering using the Seurat pipeline (version 4.1.1) on RStudio at the HPC studio visualization platform. The cells were normalized and scaled for homogeneity, and highly variable genes were selected for downstream analysis. Principal component analysis (PCA) was performed on the scaled data to reveal the main axes of variation and denoise the data. Clustering of single cells by expression matrix was achieved by setting the parameter "resolution" of the FindClusters function to 2, generating more clusters for downstream cell type identification. Data were further reduced using the Uniform Manifold Approximation and Projection (UMAP) implemented in the RunUMAP function with default parameters. Cluster-specific marker genes were identified using the FindAllMarkers function with default parameters. To identify cell types, unsupervised clustering was performed, characterizing blood cells with the high-specific marker *PTPRC*. The second round of dimension reduction and unsupervised clustering was performed on these hematopoietic/immune cells to characterize the subsets of hematopoietic/immune cells, identifying lymphoid cells using *CD3D* and myeloid cells using *CD14* following the above strategy. Subdivided cell types were defined by more specific markers published in the study.

For single-cell ATAC peak matrixs, we used the Signac pipeline (version 1.7.0) on RStudio at the HPC studio visualization platform. First, we processed the chromatin data using the peak matrix and fragment file outputted by Cell Ranger-atac. We then performed data frequency-inverse document frequency (TF-IDF) normalization and selected the top peaks for dimensional reduction. Singular value decomposition (SVD) was performed on the TD-IDF matrix using the selected features (peaks). Data were further reduced using the Uniform Manifold Approximation and

Projection (UMAP) implemented in the RunUMAP function with default parameters, as in the RNA analysis workflow. Gene activity matrixs were created for further cell type identification.

Integration of Multiple scRNA-seq Datasets by Harmony

We performed two rounds of Harmony (version 0.1.1) analysis to integrate multiple scRNA-seq datasets and merge cells of the same type to eliminate batch effects caused by different projects within the 14 tumor types. In the first round, we merged samples from the same tumor type to remove batch effects resulting from diverse protocols and platforms, including the 10× Genomics platform and SMART-seq platform. The second round of Harmony analysis aimed to merge the 14 tumor types into a single dataset, further reducing batch effects. The output of the Harmony integration was used as the input data for downstream analysis.

Detection of Mitochondrial Mutations by Pysam

The bam files generated by Cell Ranger were imported into the Pysam (version 0.21.0) module using a custom Python script. The "pysam.AlignmentFile" function was used to transfer the bam file into the alignment file, allowing us to conveniently iterate over all of the read mappings in a specified region. We paid particular attention to mitochondrial information, including the cell barcodes, position information of mitochondrial reads, sequenced base information, and the number of times the site was sequenced. We detected this information using the "pileup-engine" function, which allowed us to obtain the mitochondrial sequence data for the entire sample. To identify mitochondrial mutations, we aligned this sequence data with the human mitochondrial reference genome downloaded from the Ensemble website at the same mitochondrial position site. If the base we sequenced differed from the reference, we recorded it as a mutation. We then calculated the mitochondrial mutation frequency of the entire sample to facilitate lineage tracing

and variant quality control. Allele frequency was calculated as the reads of a specific base at a position divided by the total reads at that position. The data of cells' mitochondrial mutations were prepared for the next lineage tracing and further analysis.

Detection of SNP by samtools and bcftools

To obtain SNP (Single Nucleotide Polymorphism) information for each cell, a series of steps were taken. First, the whole data bam was split into smaller bams for each cell using samtools. Then, samtools pileup was used to align the split bam files to the GRCh38 human reference genome. Next, bcftools (version 1.17) was employed to filter unreliable SNPs, which included those with a quality score of less than 10, a DP (read depth) score of less than 5, and SNPs near INDELs within 5 sites. The output files from this step were VCF (variant call format) files, which are a common file format for variant analysis. The VCF files for each cell were merged into a single VCF file for further analysis to trace the lineage of cells in the tumor. All of the above steps to obtain SNPs used the array job function on the HPC platform, which allowed scripts to run simultaneously.

Prediction of cell relationships through mitochondrial mutations

To ensure accurate lineage tracing, we filtered out variants with an allele frequency less than 5% to eliminate those that may disrupt heteroplasmy due to sequencing errors or RNA editing. We had tested various cutoff values and found 5% to be appropriate for our analysis. Additionally, variants that occurred too infrequently cannot be used to predict cell relationships, which is our primary aim. After filtering, we obtained new mitochondrial mutation data for each sample, and we used a Python script with the 'multiprocessing' module to analyze the data in parallel and obtained results quickly.

We computed the correlation between every two cells by building a matrix that contains their

mitochondrial mutations and frequencies. To obtain preliminary cell change results, we filtered correlation values less than 0.8, which we found to be a convincing cutoff to produce reasonable and beneficial lineage tracing results. We hypothesize that downstream cells should have more mitochondrial mutations than upstream cells since decedent cells inherit the ancestor's mutation points and accumulate additional mutations. Therefore, we calculated the mitochondrial mutation number of cell pairs and defined the cell with more variants as the downstream cell and the cell with fewer mutations as the upstream cell.

Additionally, we considered traditional cell change rules, which suggest that downstream cells contain mitochondrial mutations from upstream cells. Thus, we converted mitochondrial mutations into sets after filtering unreliable mutations. If one set was a subset of another set, we considered the corresponding cell of the first set as the upstream cell and the corresponding cell of the second set as the downstream cell. We used this cell relationship as a supplement to the results obtained in the previous step, which helped remove false positive cell pairs.

Prediction of Cell Relationships by SNP

The VCF files from each single cell were merged into a single file to obtain a precise result of cell relationships. To reduce interference, we filtered the SNPs a second time based on their quality score, removing those with a score less than 30. Additionally, we filtered out SNPs with insufficient read counts, with cutoff values ranging from about 10 to 600 depending on the sample. We also considered the Variant Allele Fraction (VAF) to filter SNPs below a certain threshold.

Next, we ran the same python script as in the previous step to obtain the correlation values between every two cells. We obtained preliminary cell change results calculated by SNP after filtering out pairs with a correlation value less than 0.6. We arrived at this cutoff value after

extensive testing, as it provided reasonable and beneficial results for lineage tracing.

Finally, we counted the number of SNPs in each cell to determine whether it is the upstream or downstream cell in the cell pair.

Integration of the Two Relationships

To obtain a more accurate and precise result of cell change relationships, we merged the predictions obtained from mitochondrial mutation and SNP data. Since these two types of data provided complementary information, merging them increased the overall accuracy of the result. We also took into account the biological reality that a single cell can give rise to multiple offspring cells, but the reverse is not possible. Therefore, we excluded cases where multiple cells transferred into a single cell. However, we made an exception for the pair that had the highest correlation value calculated by either mitochondrial mutations or SNPs. After this merging process, we arrived at the ultimate relationships of hematopoietic/immune cells in the tumor. Across all samples and tumor types, we obtained a total of 9,137 pairs of hematopoietic/immune cells (Table S2).

Developmental Trajectory Inference by RNA Velocity and scVelo

To evaluate the developmental trajectory inference of hematopoietic/immune cells in the tumor, we utilized velocyto (version 0.17) to process the RCC samples' output files from Cell Ranger, generating a loom file for further analysis. We then implemented the workflow of scVelo, which is the evolutionary version of RNA velocity, to calculate the direction of each cell. The resulting UMAP coordinates and arrow coordinates were combined to determine the cells' change direction, which was classified into eight types for more accurate classification.

Developmental Trajectory Inference by Monocle

Monocle2 (version 2.22.0) is a well-known tool used for pseudotime analysis to infer the developmental trajectory of cells. However, it is typically used to analyze physical cell development. In our study, we have applied the Monocle2 algorithm to characterize the developmental origins of hematopoietic cells in RCC.

Lymphoid cell and Myeloid Cells Activity Program

To explore the cellular activity and tendencies in different tumor types, we compiled information on cell types into a single dataset (Extended Data Fig.5), resulting in 286 groups of data with average expression values for the cells they contain. We then selected approximately 1600 genes, including over 900 lymphoid cell genes and over 500 myeloid cell genes, which are the top differential genes of the respective cell types, to calculate lymphoid cell gene scores and myeloid cell gene scores. Variance was then computed to represent the discreteness of each cell, which can provide insight into the cells' tendencies to change. For lymphoid cells, the abscissa represents the lymphoid cell gene change tendency score, where a higher score indicates a greater likelihood of change, while the ordinate represents the lymphoid cell gene score, where a higher score suggests a greater resemblance to a lymphoid cell. For myeloid cells, the abscissa represents the myeloid cell gene score, where a higher score indicates a greater similarity to a myeloid cell, while the ordinate represents the myeloid cell gene change tendency score, where a higher score suggests a greater likelihood of change. For dendritic cells, which express both lymphoid cell genes and myeloid cell genes, the abscissa and ordinate represent their tendency towards lymphoid and myeloid cell gene change, respectively. Finally, for erythrocytes, which do not express lymphoid or myeloid cell genes and are not our focus, the abscissa represents the myeloid cell gene score, while the ordinate represents the lymphoid cell gene score (Fig.3a). We repeated this process with

different gene sets, including neutrophil gene set and macrophage gene set, as well as $CD4^+T$ gene set and $CD8^+T$ gene set, to explore changes in cells within the lymphatic and medullary systems, respectively.

Intermediate co-expressor (IC) program

To investigate the intermediate cells of cell change, our goal is to identify these cells from the tumor hematopoietic/immune cells. Using the cell transdifferentiation from B cell to macrophage-APOE as an example, we defined a B cell gene set and a macrophage-APOE gene set. The intermediate cells must express both *CD19* and *APOE*, and we filtered the cells by requiring them to have both the B cell gene score and the macrophage-APOE gene score. As the B cells are the upstream cells, we set the B cell gene score as a negative number. The filtered cells' abscissa is calculated as the sum of the B cell gene score and the macrophage-APOE gene score, while the ordinate is calculated as the B cell gene score divided by the macrophage-APOE gene score or the macrophage-APOE gene score divided by the B cell gene score, depending on which score is higher. As shown in Fig.3b, cells that are located around the y-axis and have higher y scores are more likely to be intermediate cells. It's important to note that different intermediate cells depend on different gene sets.

NicheNet analysis

NicheNet (version 1.1.1) is a powerful tool for predicting the ligands that drive the transcriptomic changes of target cells. To identify potential ligands that influence cell change, we selected $CD4^+T$ cells, $CD8^+T$ cells, endothelial cells, epithelial cells, fibroblasts, and monocytes as sender cells, respectively, and used intermediate cells as well as new upstream cells as receptor cells. To evaluate the top active ligands, we chose the cells that showed expansion as the negative control

cells and constructed interactions between ligands and receptors. We listed the top 40 ligands of each cell type and assigned scores to these top ligands corresponding to their ranks. The resulting dot plot shows the ligands involved in cell transdifferentiation between B cells and macrophages, the ligands of cell transdifferentiation between macrophages and neutrophils, the ligands involved in cell transdifferentiation between NK cells and T cells, and the ligands associated with the cell state transition of $CD4^+$T Th17 cells and $CD4^+$T Treg cells.

Annovar analysis

Annovar is a valuable tool for annotating somatic mutations. We use the VCF files of the changed cells as input files for Annovar and obtain the annotated genes where the SNPs are located. We specifically focus on two types of gene mutations, exonic mutations (Table S3), and splicing site gene mutations (Table S4). By obtaining a matrix of cell gene mutations per cell, we treat it as a gene expression matrix and use Seurat to identify differential gene mutations of the cell types. These gene mutations can represent different cell types undergoing various changes and mutations (Table S5).

GO analysis

Gene Ontology (GO) is a powerful tool for exploring the biological functions of genes, providing insight into the molecular function, cellular component, and biological process. To better understand the functional implications of the gene mutations we have identified, we employ ClusterProfiler (version 3.16) for analysis. This analysis is performed by comparing the gene output from annovar against the R packages 'org.Hs.eg.db'.

Correlation of the ribosome gene and P53 target gene

To investigate the potential correlation between the mutated ribosome gene and the P53 target

gene, we adopt a multi-step approach. Firstly, we calculate the average expression levels of these two gene sets in different cell types. Specifically, we compare the expression levels of these genes in the changed cell type and the expansion cell type, respectively. Then, we compute the ratio of the mutated ribosome gene and the ratio of the P53 target gene for each cell type.

Next, we examine the correlation between the mutated ribosome gene and the P53 target gene by computing the r values. The r value is a statistical measure of the strength and direction of the linear relationship between two variables. In this context, the r value reflects the correlation between the expression levels of the mutated ribosome gene and the P53 target gene in each cell type. A higher r value indicates a stronger correlation between the two genes.

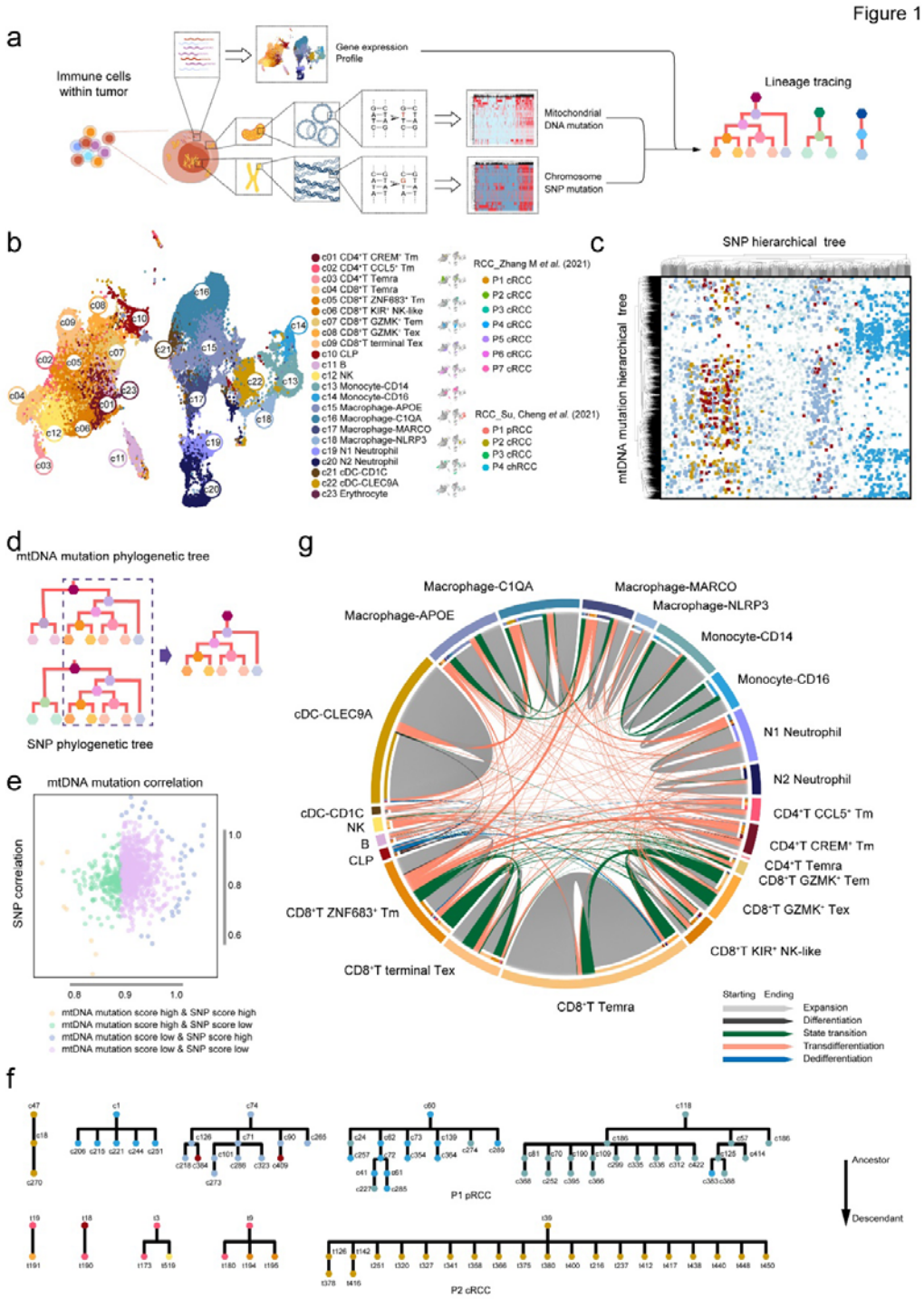

Fig.1 | The lineage tracing method using transcriptome and genetic variation.

a) Schematic of the workflow for lineage tracing in pan-cancer using single-cell transcriptome, mitochondrial DNA mutation, and chromosome SNP mutation of immune cells.

b) UMAP plot showing the integration of two RCC scRNA-seq datasets (RCC_Zhang M et al., 2021 and RCC_Su, Cheng et al., 2021) on the left. UMAP plots displaying the distribution of each sample on the right.

c) Scatter plot displaying the immune cells of the P1 pRCC sample (RCC_Su, Cheng et al., 2021). Each dot represents a cell, with the x-axis indicating the rank of the SNP hierarchical tree and the y-axis indicating the rank of the mtDNA mutation hierarchical tree. A shorter distance between dots indicates a higher similarity in genetic information between cells.

d) Pattern diagram showcasing the phylogenetic tree (left) using mtDNA mutation and SNP, respectively, and the final phylogenetic tree (right) generated by overlapping the two methods.

e) Dot plot illustrating the filtered two-cell pairs in the P1 pRCC sample (RCC_Su, Cheng et al., 2021). The filtration standard requires a mtDNA mutation correlation score of more than 0.8 and a SNP correlation score of more than 0.6. A high mtDNA mutation correlation score indicates a score above 0.9 but below 1.0, while a low mtDNA mutation correlation score indicates a score above 0.8 but below 0.9. A high SNP correlation score indicates a score above 0.8 but below 1.0, while a low SNP correlation score indicates a score above 0.6 but below 0.8.

f) Phylogenetic tree calculated by both mtDNA mutations and SNPs, depicting the process where ancestor cells transfer to descendant cells (from top to bottom) in the P1 pRCC sample and the P2 cRCC sample (RCC_Su, Cheng et al., 2021). Each dot represents a cell, and its color corresponds to the cell type in Fig.2b.

g) Circos plot demonstrating the dynamics of hematopoietic/immune cells in the RCC datasets in Fig.2b. The different colors of arrows indicate different cell change types. The arrows begin with starting cells and end with ending cells. The color of the outer ring corresponds to the cell type of

both starting and ending cells in Fig.2b. The colors of the inner ring indicate the cell type of the ending cells.

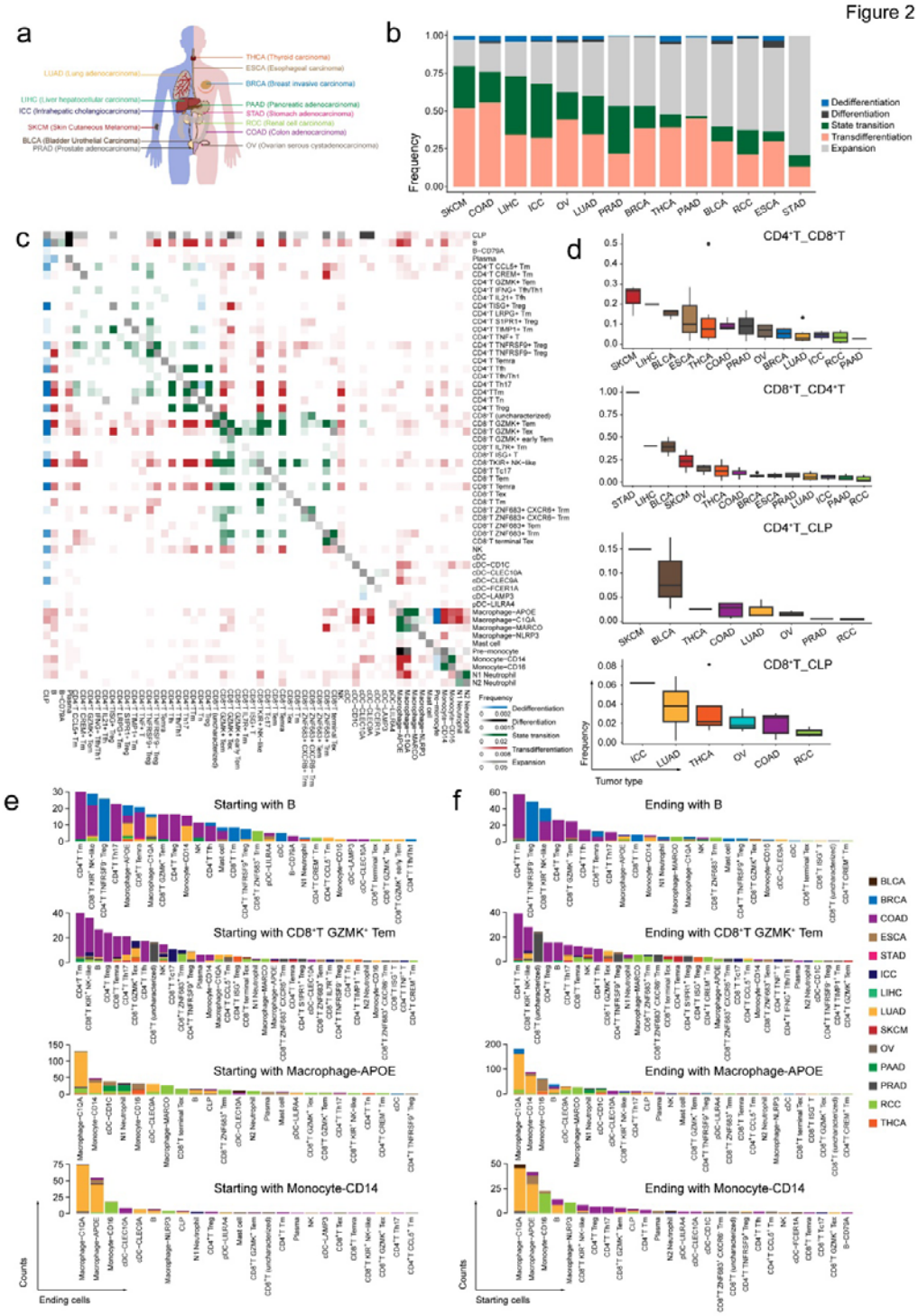

Fig.2 | Dynamics of immune cells in pan-cancer at the single-cell resolution.

a) Pattern diagram showing tumor types involved in the pan-cancer analysis. The abbreviations are according to the TCGA database.

b) Bar plot illustrating the frequency of the five cell change types among each tumor type.

c) Heatmap showing the frequency of cell change types in the pan-cancer analysis. The cell types in the rows are the starting cells, and the cell types in the columns are the ending cells. Different colors represent different cell change types. The frequency is calculated through dividing the counts of each cell change type by the total number of all cell change types.

d) Box plots showing the proportions of the four cell change types in major lineages across various tumor types.

e) Bar plots reporting the counts of ending cells that changed from the labelled cell types across various tumor types.

f) Bar plots reporting the counts of starting cells that changed to the labelled cell types across various tumor types.

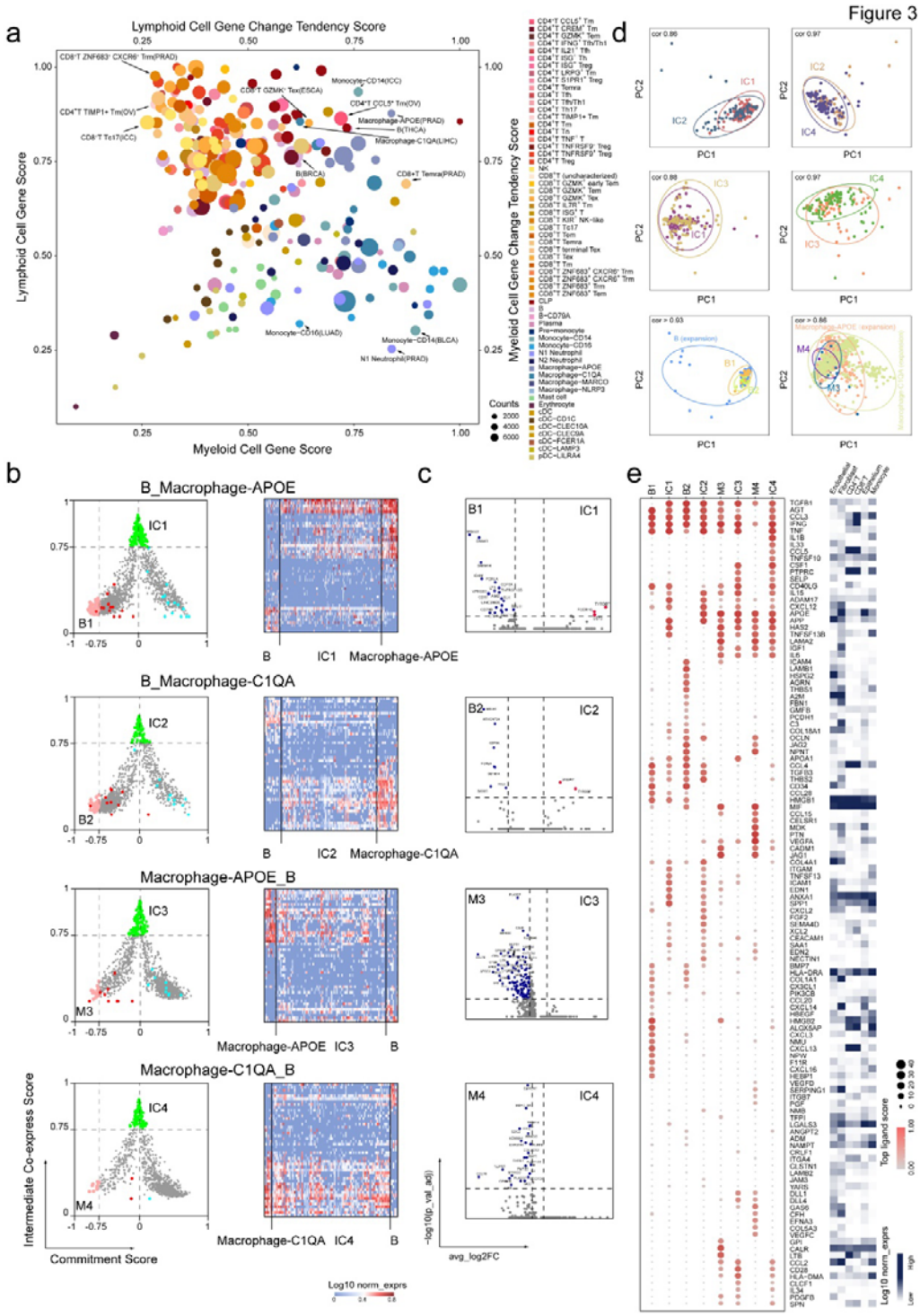

Fig.3 | Tumor microenvironment factors related to immune cell plasticity.

a) Scatter plot illustrating the gene expression of the hematopoietic/immune cells and their potential ability to change. Each dot represents one cell type of the 14 tumor types and its size is

proportional to the number of cells in that particular cell type. For myeloid cell types, the x-axis represents myeloid cell gene score, indicating their gene expression, and the y-axis represents myeloid cell gene change tendency score, indicating whether these cells tend to change. For lymphoid cell types, the y-axis represents lymphoid cell gene score, indicating their gene expression, and the x-axis represents lymphoid cell gene change tendency score, indicating whether these cells tend to change. Dendritic cells' x-axis represents lymphoid cell gene change tendency score, while the y-axis represents myeloid cell gene change tendency score. Erythrocytes' x-axis represents myeloid cell gene score, while the y-axis represents lymphoid cell gene score.

b) Scatter plot showing the change process of transdifferentiation between B and macrophage-APOE or macrophage-C1QA. The abscissa represents B/macrophage commitment score, and the ordinate represents Intermediate Co-express Score. Cells with an x-value less than -0.75 are defined as potential starting cells, and cells with a y-value greater than 0.75 are defined as ICs. Green dots represent intermediate cells and pink dots represent potential starting cells. Red dots and blue dots represent the starting and ending cells of the transdifferentiation which we discovered in pan-cancer using the method we built in this study, respectively. The heatmaps on the right show the specific genes of B cells , macrophages and the intermediate cells.

c) Volcano plots illustrating the differentially expressed genes between potential starting cells and intermediate cells in (b) using Benjamini-Hochberg adjusted two-sided Wilcoxon test.

d) PCA plots showing the distribution of intermediate cells (up and middle) and the distribution of potential starting cells and the expansion B cells or the expansion macrophages in pan-cancer (down).

e) Dot plot illustrating the possible ligands driving the potential starting cells and intermediate

cells against expanding B cells and expanding macrophages, respectively, sent by other immune cells and non-hematopoietic cells. The top 40 ligands are labelled for each cell type and the heatmap shows the expression of ligands in the immune cells and non-hematopoietic cells.

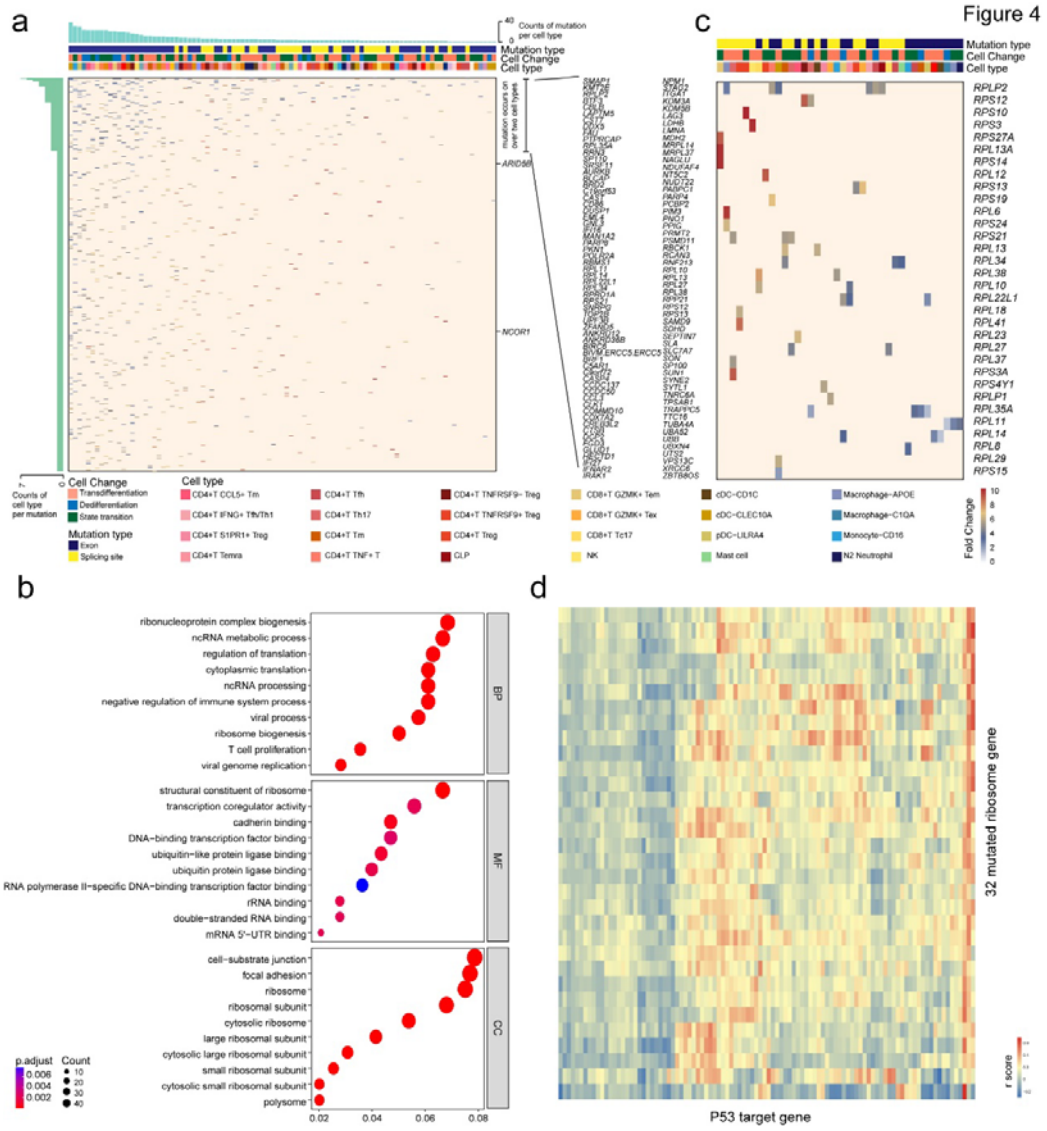

Fig.4 Intrinsic genetic mutations related to immune cell dynamics.

a) Heatmap displaying the gene of SNPs locating on exon site and the splicing site. All the cells are the starting cells of trasdifferentiation, state transition or dedifferentiation. Mutations occurring over two cell types and tumor type-specific mutations are listed. Bar plots show counts of mutation per cell type and counts of cell type per mutation.

b) GO enrichment analysis of the gene related to SNPs in (a).

c) Heatmap showing the differential mutated ribosome protein genes in the starting cells of different cell changes.

d) Heatmap displaying the Pearson correlation of the 32 mutated ribosome gene and P53 target gene. The raw data is a ratio of the expression of cells in (a) against the expansion cells. The r score represents the correlation of the 32 ribosomal protein genes and P53 target gene in the change cells.