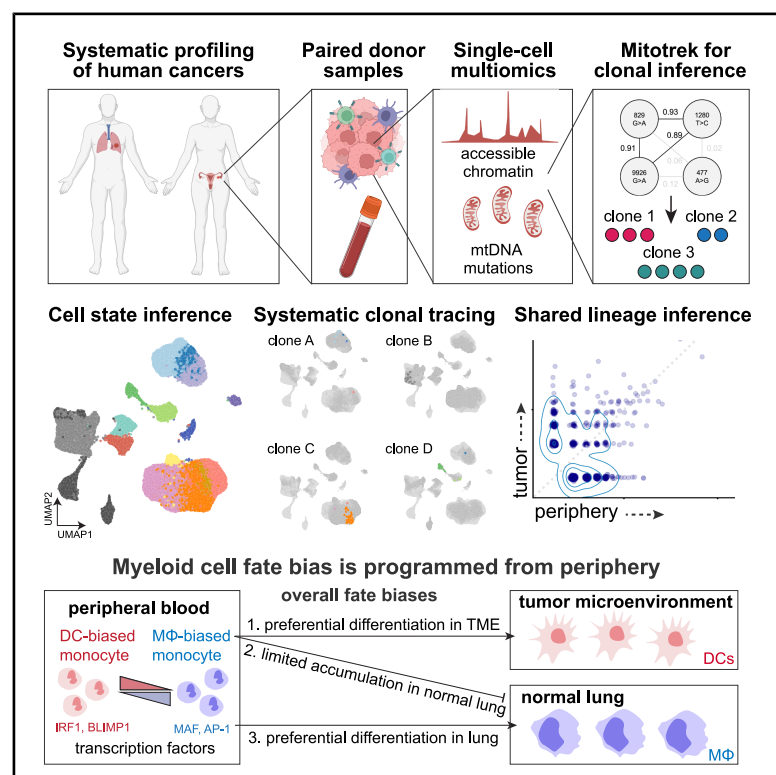


Cancer Cell

Clonal lineage tracing of innate immune cells in human cancer

Graphical abstract



Authors

Vincent Liu, Katalin Sandor, Patrick K. Yan, ..., Brooke E. Howitt, Caleb A. Lareau, Ansuman T. Satpathy

Correspondence

liuv@stanford.edu (V.L.),
lareauc@mskcc.org (C.A.L.),
satpathy@stanford.edu (A.T.S.)

In brief

Liu et al. develop Mitotrek, a method that uses single-cell multi-omics with mitochondrial DNA clonal tracing to map the clonal origins of innate immune cells in human tumors. They reveal that tumor myeloid populations arise from circulating monocytes with pre-existing epigenetic fate biases, suggesting that tumor immunity is shaped before tumor infiltration.

Highlights

- mtDNA lineage tracing enables clonal analysis of innate immune cells in tumors
- Tumor myeloid cells are clonally linked to circulating monocyte precursors
- DC3s derive from monocytes and retain monocytic epigenetic features
- Monocyte fate bias is epigenetically programmed before tumor infiltration

Report

Clonal lineage tracing of innate immune cells in human cancer

Vincent Liu,^{1,2,3,*} Katalin Sandor,^{2,3} Patrick K. Yan,^{2,3} Zhuang Miao,^{1,2,3} Yajie Yin,^{2,3} Robert R. Stickels,^{2,3} Andy Y. Chen,^{2,3,4} Kamir Hiam-Galvez,^{2,3} Jacob Gutierrez,⁵ Wenxi Zhang,^{2,3} Sairaj M. Sajjath,⁶ Raeline Valbuena,¹ Steven Wang,¹ Bence Daniel,^{2,3} Leif S. Ludwig,^{7,8} Brooke E. Howitt,² Caleb A. Lareau,^{5,*} and Ansuman T. Satpathy^{2,3,9,10,*}

¹Department of Genetics, Stanford University, Stanford, CA 94305, USA

²Department of Pathology, Stanford University, Stanford, CA 94305, USA

³Center for Immunotherapy Design, Stanford University, Stanford, CA 94305, USA

⁴Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

⁵Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

⁶Robin Chemers Neustein Laboratory of Mammalian Cell Biology and Development, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA

⁷Berlin Institute of Health at Charité – Universitätsmedizin Berlin, 10178 Berlin, Germany

⁸Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC) Berlin Institute for Medical Systems Biology (BIMSB), 10115 Berlin, Germany

⁹Parker Institute for Cancer Immunotherapy, San Francisco, CA 94129, USA

¹⁰Lead contact

*Correspondence: liuv@stanford.edu (V.L.), lareauc@mskcc.org (C.A.L.), satpathy@stanford.edu (A.T.S.)

<https://doi.org/10.1016/j.ccell.2026.05.006>

SUMMARY

Innate immune cells constitute the majority of the tumor microenvironment (TME) and mediate anti-tumor immunity and immunotherapy responses. While single-cell T and B cell receptor sequencing have revealed insights into the clonal dynamics of adaptive immunity, the lack of analogous tools has precluded similar analysis of innate immune cells. Here, we describe a method leveraging somatic mitochondrial DNA (mtDNA) mutations to reconstruct clonal lineage relationships between cells in native human tissues. By jointly profiling single-cell chromatin accessibility and mtDNA variants, we resolve clonal dynamics of 218,715 cells from matched tumors, tissues, and blood from patients with lung and ovarian cancers. Clonal tracing reveals that TME-resident myeloid subsets, including macrophages and type 3 dendritic cells (DC3), are clonally related to circulating and tissue-infiltrating monocytes. We further identify distinct DC-biased and macrophage-biased clones, whose circulating monocyte precursors exhibit distinct epigenetic profiles, suggesting intratumoral myeloid differentiation fate may be peripherally programmed before TME infiltration.

INTRODUCTION

The tumor microenvironment (TME) is a dynamic ecosystem that maintains a close connection to the peripheral immune system through the bloodstream. Circulating immune cells continually infiltrate the TME, where their capacity to differentiate and expand shapes anti-tumor immunity and therapy response. Lineage tracing methods such as paired single-cell RNA and TCR sequencing (scRNA/TCR-seq) have advanced understanding of the expansion potential, tissue distribution, and clonal dynamics of immune populations in patients.^{1,2} For example, we and others have applied scRNA/TCR-seq to show that peripherally expanded, tumor-specific CD8⁺ T cell clones replace dysfunctional exhausted T cells in the tumor, which is essential to immune checkpoint blockade response.^{3–5} However, similar lineage-tracing approaches have not been feasible for non-adaptive immune populations due to the absence of somatic recombination events during their development.

Innate immune cells, including myeloid cells, are among the most abundant immune cell types in the TME⁶ and are ontogenetically diverse, consisting of both tissue-resident and circulating bone marrow-derived compartments.⁷ Although phenotypic and quantitative maps have delineated their diversity across tissues and cancer types,^{8–11} and there is growing recognition that ontogeny plays a key role in shaping their function within the TME,^{12,13} the inability to connect cellular phenotype and lineage within native human tissues leaves several open questions. First, do innate immune cells have an intrinsic tissue site preference for infiltration, differentiation, or expansion? Second, do they face clonally selective pressures in the TME, analogous to those experienced by T and B cells? Finally, what specific lineage relationships connect the diverse innate immune constituents within the TME, and what are their ontogenetic origins?

Somatic mtDNA variants have been established as endogenous lineage markers, as the high mutation rate of mtDNA

relative to nuclear DNA drives the accumulation of lineage-distinguishing mutations over time.^{14,15} Recent advancements in single-cell sequencing have enabled the detection of mtDNA mutations^{16–19} allowing comprehensive, simultaneous lineage and epigenetic profiling of thousands of clonotypes across cell types, tissue sites, and cancer types. Therefore, to address these fundamental questions regarding innate immune cell clonality and ontogeny, we performed mitochondrial single-cell assay for transposase accessible chromatin by sequencing (mtscATAC-seq) on 218,715 cells from tumor, non-involved lung tissue (NILT), and/or peripheral blood mononuclear cells (PBMCs) from patients with lung or ovarian cancer to provide a clonally resolved view of the human innate immune response to cancer.

RESULTS

Development of Mitotrek to recover single-cell clonal relationships

To establish a lineage tracing method applicable to donor-matched solid tissue and blood samples and to clonally link immune cells across tissue sites, we applied mtscATAC-seq to profile five patients with early-stage non-small cell lung cancer (NSCLC) who had undergone therapeutic lobectomy, spanning adenocarcinoma (SU-L-002, SU-L-004, SU-L-005; treatment-naïve), squamous cell carcinoma (SU-L-001; post-neoadjuvant chemotherapy), and neuroendocrine (SU-L-003) histologies (Figures 1A and S1A; Table S1; methods). To investigate whether distinct TMEs influence immune clonal architecture, we complemented the NSCLC dataset, characterized by high mutational burden and strong immune infiltration,^{20,21} with mtscATAC-seq data from five ovarian cancer patients, which typically exhibit a more immunosuppressive TME²² (Figure S1B; Table S1). Fresh samples were dissociated and used to generate single-cell ATAC and mtDNA genome-sequencing libraries, enabling simultaneous recovery of chromatin accessibility profiles and mtDNA genotypes from individual cells.

In the lung cancer cohort, we obtained mtscATAC-seq profiles from 83,371 immune, malignant, and stromal cells from tumor and NILT, and 41,587 profiles from peripheral blood. In the ovarian cancer cohort, we profiled 52,154 immune, malignant, and stromal cells from ovarian tumors and 41,603 PBMCs (Table S2; Figure S1C). For one patient (SU-O-002), we additionally generated paired scRNA/TCR-seq data from 6,982 cells spanning peripheral blood and tumor, with TCR-mtDNA correspondence supporting clonal barcoding of immune cells in this dataset¹⁶ (Figure S1D; methods). We recovered 10,584 high-confidence somatic mtDNA mutations present at low pseudobulk frequencies (<1%) as determined by *mgatk*¹⁷ (Figure S1E), a computational pipeline to process mitochondrial reads and generate heteroplasmy estimates from mtscATAC-seq data. Variants with >1% pseudobulk frequencies, accounting for 0.76% of all variants detected, were broadly distributed across cell types and tissue sites, likely representing early-developmental or zygotic mutations²³ that are not informative for resolving recent immune lineage relationships (Figure S1F), and were therefore excluded from downstream analysis.

Analysis of heteroplasmy distributions revealed that mtDNA variants detected at high allele frequencies in solid tissue cells were more frequently shared across cells, including from other

lineages, than in matching PBMC samples, likely due to both technical artifacts (e.g., uptake of ambient mtDNA during single-cell capture²⁴) and biological phenomena (e.g., horizontal mitochondrial transfer events²⁵) that are unlikely to reflect *bona fide* clonal relationships and may be unique to solid tumor profiling^{18,26,27} (Figures S1G–S1L). To mitigate such spurious lineage inferences, we developed Mitotrek, a computational framework that assigns cells to high-confidence clones by prioritizing clonal accuracy over sensitivity (Figure 1B; methods). Briefly, Mitotrek: (1) reformats cell-by-variant heteroplasmy data into a binary “positive-unlabeled” matrix using empirically defined thresholds to account for measurement noise from genetic drift²⁸ and variable sequencing depth, (2) discards variants detected in >20% of cells within a sample, which are likely to generate artifactual linkages, and (3) excludes cells assigned to multiple clones to avoid contamination by non-informative variants of technical or biological origin. Subclonal structure was not explored due to the difficulty in establishing true clonal hierarchy, and potential subclonal relationships were collapsed to single clones by merging highly correlated variants. Benchmarking Mitotrek on full-length scRNA-seq data from single-colony hematopoietic stem and progenitor cells (HSPCs) with known clonal identities¹⁶ yielded assignment accuracies of 91.3% and 85.3% for two independent donors (Figures 1C, 1D, S1M, and S1N), establishing Mitotrek as a high-accuracy framework for resolving clonal relationships in solid human tissues.

Constructing multi-modal atlases of NSCLC and ovarian cancer

We next constructed a single-cell epigenetic atlas of NSCLC by iteratively clustering scATAC-seq profiles from tumor and NILT samples, identifying 53 cell clusters spanning malignant tumor, epithelial, stromal, T cell, NK/ILC, myeloid, B cell, and plasma cell populations (Figures 1E; Figure S2A). All non-malignant cell-type annotations contained cells from multiple donors, consistent with interpatient heterogeneity,²⁹ and cell-type assignments aligned with surface markers used for enrichment (Figures 1F and S2B). Cell-type distributions in tumor and NILT samples matched previous transcriptomic and proteomic datasets^{8,30} (Figure 1G). In patient-matched PBMCs, we recovered all major immune populations with the expected absence of tissue-resident myeloid subsets, and additionally identified circulating HSPCs (Figures S2C–S2F), enabling unbiased downstream lineage analysis.

Applying Mitotrek to the NSCLC dataset, we recovered 5,146 clones comprising 29,010 cells (23.2% of cells passing scATAC-seq QC filters), of which 3,190 clones (62%) contained ≥ 3 cells and 573 (11%) contained ≥ 10 cells. The number of clones recovered per patient (440–2,146) was highly correlated with cell recovery ($R = 0.90$; Figure S2G), and clone assignment rates were consistent across cell types and tissue sites, ranging from 15% to 26% in tumor/NILT and 20%–32% in PBMC (Figures S2H and S2I), with the fraction of multi-assigned cells decreasing with clone size (Figures S2J and S2K). Globally, cells within the same clone were significantly more likely to share a cell type than randomly grouped cells across all tissues (Figure 1H), reflecting lineage fidelity and supporting the validity of Mitotrek-assigned clones. A clone-level cell type association

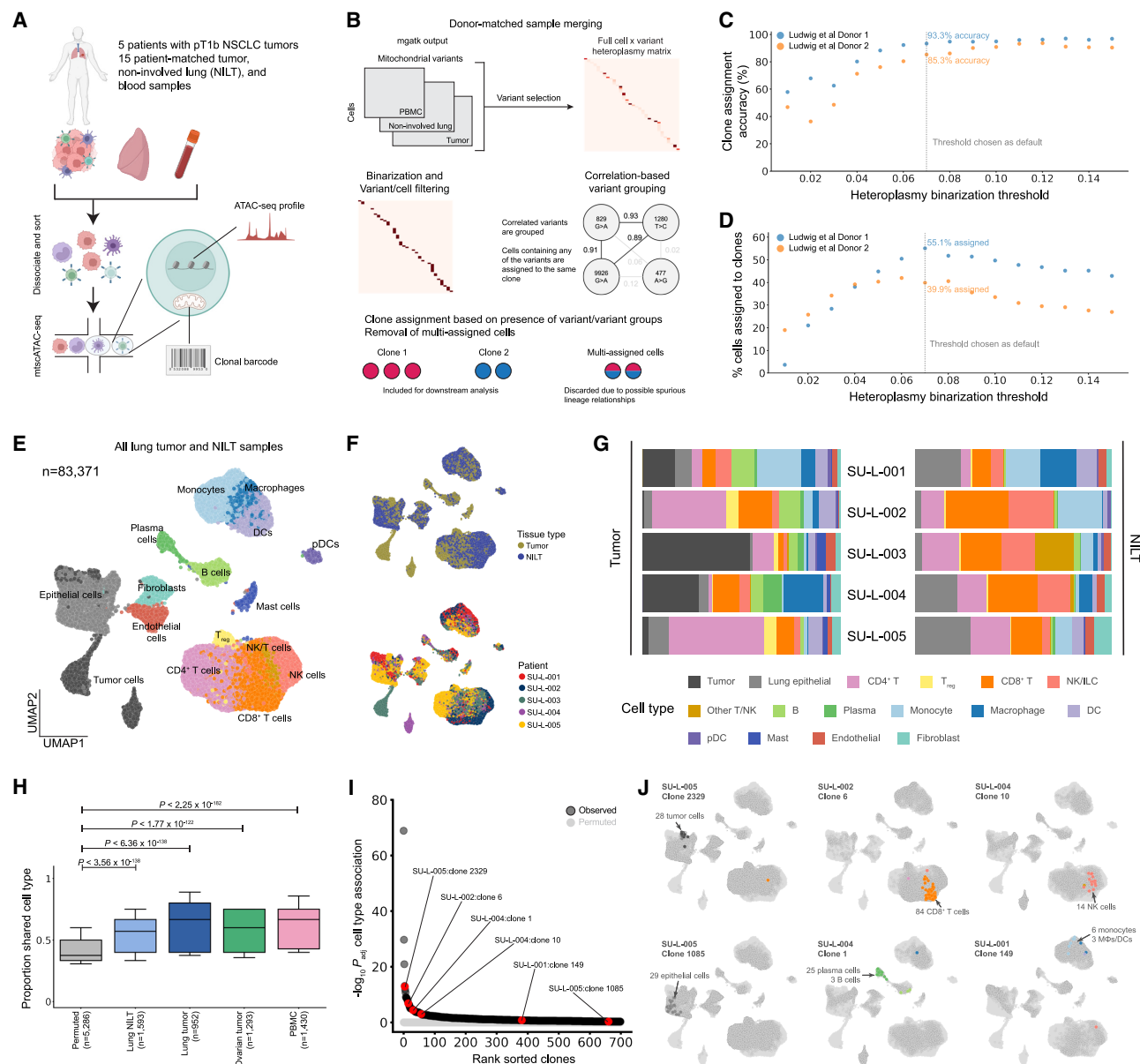


Figure 1. Creating a lineage-embedded atlas of NSCLC from mitochondrial DNA with Mitotrek

(A) Schematic of a simultaneous single-cell epigenetic profiling and clone tracing in patients with early-stage NSCLC. The chromatin accessibility profile and mitochondrial mutations are recovered from each cell.

(B) Schematic of clone calling from mitochondrial variants using Mitotrek.

(C) Benchmarking Mitotrek on gold-standard HSPC colony data.

(D) Same as (C) but for a second HSPC donor.

(E) Uniform manifold approximation and projection (UMAP) of 83,371 cells in lung tumor and non-involved lung tissue (NILT).

(F) UMAP of cells colored by tissue type (top) and patient identity (bottom).

(G) Normalized bar plot showing cell type composition for each patient, partitioned by tumor and non-involved lung.

(H) Distribution of the proportion of cells within each clone (≥ 3 cells) that share the most common cell type for that clone compared to permuted data from the same samples. N = number of clones, Kruskal-Wallis test. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5 \times interquartile range.

(I) Clone associations with cell type. p values represent the Benjamini-Hochberg adjusted Kruskal-Wallis test against overall cell type proportions for clones with at least five cells. Clones shown in (J) are highlighted in red.

(J) Representative clones capturing clonal expansion (tumor, CD8⁺ T, NK, and epithelial) or differentiation (B/plasma and monocyte/macrophage) events in tumor and NILT. See also [Figures S1 and S2](#); [Tables S1 and S2](#).

analysis identified 226 clones (20% of clones with ≥ 5 cells) whose compositions significantly deviated from the background distribution ($\text{Padj} < 0.05$; [Figure 1I](#)), capturing key expansion and differentiation events across immune, stromal, and epithelial compartments ([Figure 1J](#)).

Iterative clustering of ovarian tumor cells identified 40 clusters across the same broad cell type categories ([Figures S3A–S3E](#)). Mitotrek recovered 4,560 clones comprising 20,713 cells, of which 2,389 clones (53%) contained ≥ 3 cells and 287 clones (6%) contained ≥ 10 cells. The number of clones recovered per patient ranged from 61 to 1,487 ([Figure S3F](#)). Consistent with our observations in the NSCLC dataset, clone counts were strongly correlated with cell recovery per patient ($R = 0.93$), and clone assignment rates were uniform across cell types and tissue sites ([Figures S3F–S3J](#)). Together, the lineage-embedded NSCLC and ovarian tumor datasets provide a robust framework for dissecting lineage dynamics across distinct TMEs.

Immune clonal landscapes across distinct tissue sites reveal a diverse clonal repertoire of tissue-infiltrating myeloid cells

As mtscATAC-seq enables a measure of clonal diversity across all cell states, we assessed the relative clone sizes of both immune and tumor cells from NSCLC and ovarian cancer samples. We quantified the distribution of clone fractions by cell type, reasoning that high per-clone fractions reflect cell type-specific clonal expansion events and noting that mitochondrial lineage tracing is particularly adept at marking clonal expansions³¹ ([Figures 2A and S3K](#)). Overall, tumor cells and adaptive immune cells, including T cells, B cells, and plasma cells, comprised the largest clones observed in both lung and ovarian tumors ([Figures 2B and S4A](#)). Among clones composed of $\geq 80\%$ of a single cell type (≥ 3 cells per clone), lung tumors were dominated by tumor cells, CD8^+ T cells, CD4^+ T cells, and B/plasma cells, whereas ovarian tumors had notably lower representation of CD4^+ T cells, suggesting that clonally expanded CD4^+ T cells are more prominent in the lung TME ([Figures S4B–S4F](#)). This inference was corroborated by matched ovarian tumor scRNA/TCR-seq, which showed markedly reduced clone sizes for CD4^+ compared to CD8^+ T cells (mean 7.6 vs. 25.4 cells per TCR clone; Student's t test $p = 0.029$; [Figure S4G](#)).

Myeloid populations, including monocytes, macrophages, and DCs, exhibited smaller clone fractions, compared to adaptive immune cells. Among clones with a substantial number (≥ 5) of myeloid cells detected in lung tumors, the mean myeloid fraction was 58.5%. In contrast, clones with a substantial number of CD4^+ T cells, CD8^+ T cells, and B/plasma cells exhibited higher mean fractions of 78.7%, 78.3%, and 62.1% of the corresponding cell types ([Figure S4H](#)). Similar patterns were observed in ovarian tumors, except for the lower level of clonal expansion of CD4^+ T cells ([Figures S4I–S4K](#)). These data provide direct evidence that sustained clonal expansion of human myeloid cells following bone marrow egress is limited, compared to lymphoid cell types, which clonally expand following antigen recognition. Notably, we observed oligoclonal NK cell expansions in one ovarian tumor (SU-O-005), consistent with adaptive-like NK responses reported in other contexts^{26,32–35} ([Figures S5A–S5D](#)).

We next compared the clonal landscape of lung tumors to the paired normal lung tissue environment. Overall, clone sizes were

correlated across tissue types, with the strongest concordance observed between NILT and blood and weaker correlation between blood and tumor ([Figure S6A](#)). We classified clones with ≥ 10 cells based on their enrichment in tumor versus NILT ([Figure S6B](#)). Tumor-enriched clones were dominated by tumor cells as well as CD4^+ T, CD8^+ T, and B/plasma cells, with the adaptive immune clones largely restricted to the tumor site ([Figure 2C](#)). In contrast, NILT-enriched clones included a broader range of cell types with higher representation across tissue sites ([Figure 2D](#)), including clones predominantly composed of invariant NK/T cells and monocytes that were minimally observed in tumor-enriched clones. NILT-enriched clones also showed greater cellular heterogeneity than tumor-enriched clones ([Figure S6C](#)), consistent with reduced clonal expansion in the normal lung tissue environment.

Among the 186 clones that were not preferentially enriched in either tumor or NILT, most were dominated by CD4^+ or CD8^+ T cells ([Figure S6D](#)) and were detected across tumor, NILT, and peripheral blood, consistent with previous reports demonstrating concordance between peripheral and intratumoral T cell clone sizes.^{3,4} In contrast, B cell and plasma cell-dominated clones were restricted to either tumor or NILT, suggesting that B and plasma cell expansion is locally confined. A complementary analysis of PBMC-enriched clones identified 14 clones dominated by CD8^+ T and NK cells, the majority of which were also detected at lower frequencies in tumor and NILT, indicating their capacity for tissue infiltration ([Figure S6E](#)). Together, these results provide a comprehensive view of immune clonal composition across the human TME and periphery.

Inter-cell type clonal relationships reveal broad tissue distribution of bone marrow-derived myeloid cells

Next, we systematically quantified lineage relationships and differentiation patterns between cell types within and across tissue sites. We aggregated all clonotypes for each pair of cell types (including self-self pairs) and measured the fraction of clones shared across distinct cell-pair combinations, reasoning that two cell types were more related if they more frequently shared clonotypes ([Figures 2E and S7A](#); methods). Across all tissue samples, we identified three broad patterns: (1) innate immune cells (monocytes, macrophages, DCs, and NK cells) consistently exhibited high intra-group clone sharing, consistent with recent hematopoietic output without substantial clonal bottlenecks prior to tissue infiltration, (2) adaptive immune cells and tumor cells showed high intra-cell type clone sharing, reflecting histories of clonal selection and expansion, and (3) stromal cells (endothelial cells, fibroblasts, and non-tumor epithelial cells) displayed intermediate levels, consistent with their distant embryonic origins relative to HSPC-derived immune cells. Analogous patterns were observed in PBMCs ([Figure S7B](#)). Notable exceptions included non-regulatory CD4^+ T cells, which exhibited minimal intra-cell type clone sharing despite the expansion of select clones, indicating broader clonal diversity, and endothelial cells, which displayed particularly high intra-cell type sharing in lung and ovarian tumors but not NILT, consistent with the angiogenic processes indispensable for tumor formation.

We next quantified lineage relationships between cell types across patient-matched lung tumors, NILT, and peripheral blood. Among adaptive immune cells, only CD8^+ T cells

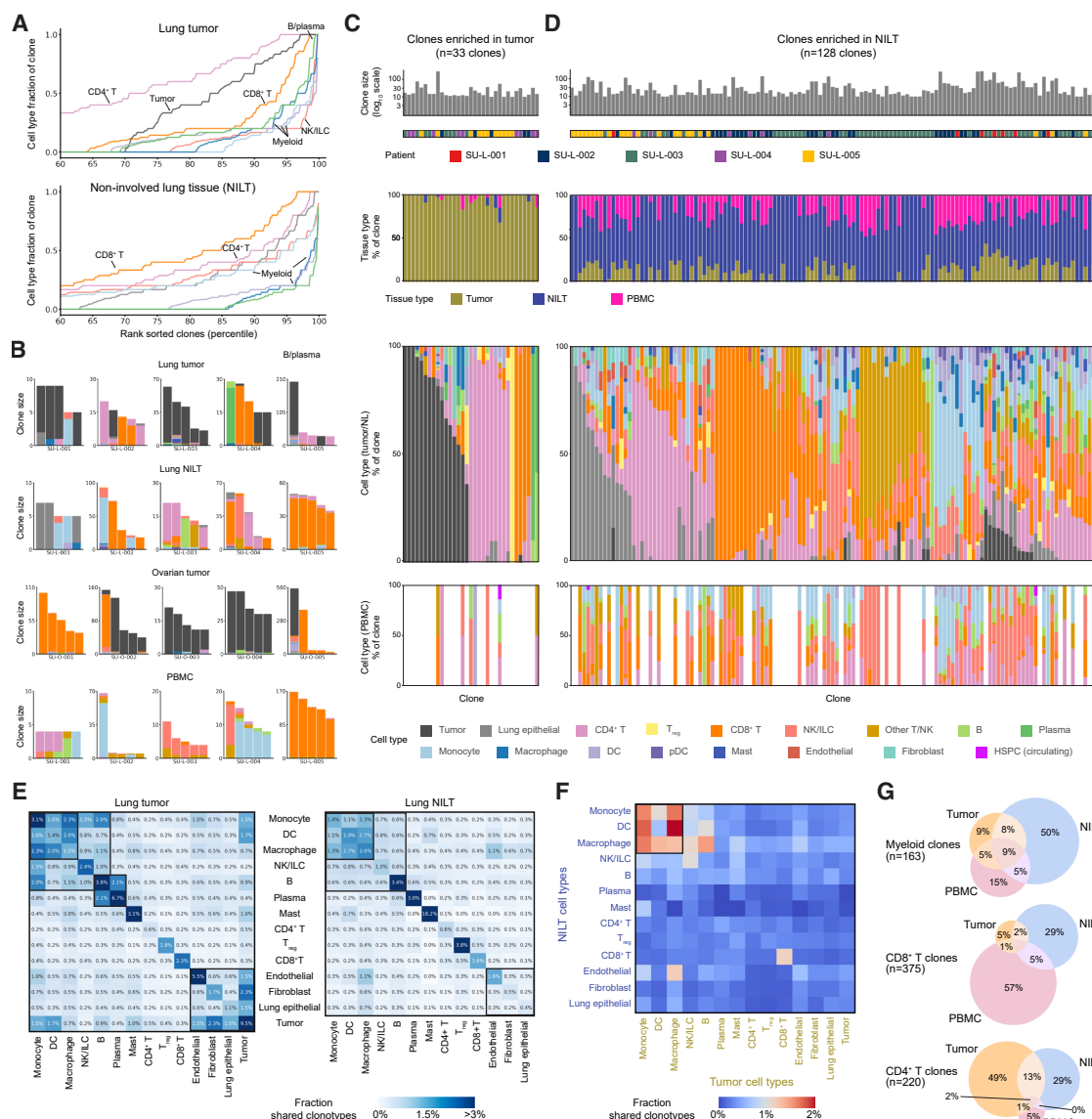


Figure 2. Cross-tissue clonal landscape of tumor-infiltrating immune cells

(A) Cumulative fractions of clones stratified by cell type for cells from lung tumor (top) and NILT (bottom). Clones with ≥ 5 cells are considered for this analysis. (B) Cell types of single cells belonging to the same clone. The top five most abundant clones with the most common cell type $>70\%$ for each patient in the indicated sample are shown. Each bar is colored by cell types of single cells within the clone.

(C) Clones enriched in tumor, as determined by p value < 0.05 from Benjamini-Hochberg adjusted Fisher's exact test against overall tissue site distribution for clones with at least 10 cells in tumor and NILT. Each column represents a unique clone.

(D) Same as in (C) but for the NILT.

(E) Heatmaps showing the fraction of all cell pairs belonging to the same clone and consisting of two cell types within lung tumor (left) and NILT (right). Pairs were restricted to cells from the same donor.

(F) Heatmap showing the fraction of all cell pairs belonging to the same clone and consisting of a lung tumor cell type and a NILT cell type.

(G) Comparison and overlap of clones (≥ 3 cells for the indicated cell type) for myeloid, CD8⁺ T, and CD4⁺ T. Myeloid consists of monocyte, macrophage, and DC. n = number of clones considered for each cell type. See also Figures S3–S7.

consistently exhibited high clone sharing across tissue sites, suggesting that peripheral expansion during the anti-tumor response is unique to this population (Figures 2F and S7C). In contrast, B and plasma cells across tissues were clonally distinct, supporting the notion that B cell-mediated immunity is locally orchestrated, as has been recently noted in the context of infection.³⁶ Among myeloid cells, we observed the greatest

degree of clone sharing across tissue sites, suggesting that these cells readily infiltrate and differentiate in tissue following bone marrow myelopoiesis and egress. Indeed, myeloid-dominated clones spanning multiple tissue sites were the most frequently detected in the NSCLC dataset (27% myeloid clones compared to 8% CD8⁺ T cell clones and 16% CD4⁺ T cell clones; Figure 2G). Together, these findings emphasize the distinct

clonal dynamics of adaptive and innate immune cells in anti-tumor immunity, highlighting the systemic nature of CD8⁺ T cell responses in contrast to the locally restricted nature of B cell responses, and revealing the extensive tissue infiltration and differentiation capacity of bone marrow-derived myeloid cells.

Intratumoral DC3s are epigenetically and clonally linked to circulating monocytes

It has been reported that specific subpopulations of circulating myeloid cells give rise to tumor-resident mononuclear phagocyte (MNP) populations that critically shape the TME.^{13,37,38} In particular, human DC3s have been identified as a CD1c⁺ dendritic cell subset distinct from cDC2s and enriched in NSCLC tumors, and have been suggested to uniquely prime tissue-resident CD8⁺ T cells.^{30,39–42} However, the epigenetic and ontogenetic relationships between human DC3s and other MNPs remain undefined. Given the broad tissue distribution of bone marrow-derived myeloid cells, we sought to understand the differentiation trajectories of circulating myeloid populations and clonally link them to those in the TME. To achieve higher granularity for myeloid cell type annotation, we iteratively re-clustered 21,676 myeloid cells from lung tumor, NILT, and peripheral blood in the NSCLC cohort into 13 distinct clusters, including classical (CD14⁺) and non-classical (CD16⁺) monocytes, alveolar macrophages (AMΦ), interstitial macrophages (IMΦ), two monocyte-derived macrophage populations (MoMΦ1 and MoMΦ2), and DC subsets including cDC1, cDC2, mature DC enriched in regulatory molecules (mregDC),^{9,43,44} and DC3s³⁰ (Figures 3A, 3B, and S8A–S8C; methods). Annotations were validated using gene signatures derived from joint single-cell transcriptomic and proteomic profiling of human lung myeloid cells from 35 NSCLC tumors and 29 patient-matched NILTs³⁰ (Figures S8D and S8E), and were consistent with previous characterizations of human lung tumors and NILTs.

Despite extensive transcriptomic and proteomic characterization of lung myeloid cells, the corresponding epigenetic landscape remains understudied. We used ArchR⁴⁵ to identify differentially accessible regions (DARs) specific to each myeloid cell subtype. Although subtypes exhibited distinct epigenetic profiles, hierarchical clustering grouped them into three broader categories reflecting their myeloid lineage, with IMΦ cells excluded from downstream analyses due to their rarity across donors (96% originated from a single donor; Figure 3C). Transcription factor (TF) motif enrichment analysis of subtype-specific DARs recapitulated established lineage-defining programs, including CEBP motifs in monocytes,^{46–52} AP-1 and metabolic receptor family (RXR, PPAR, and LXRα) motifs in macrophages,^{48–51,53–60} and canonical DC TFs (BCL11A, IRF4/8, RUNX, CBFB, and NF-κB) in DCs^{48–51,61–63} (Figures S8F and S8G).

Unlike cDC2 DARs, DC3-enriched DARs exhibited significantly greater accessibility in monocytes and macrophages (Figures S9A and S9B), and CD14⁺ monocyte DARs were more accessible in DC3s than in other DC subtypes (Figure S9C), suggesting that DC3s retain monocytic epigenetic features. TF motif analysis of DC3 DARs revealed increased accessibility for motifs associated with monocyte and macrophage TF families, including AP-1, MAF, and Mit-TFE (Figures 3D–3F). Using chromVAR, we inferred per-cell TF activity and observed con-

current activation of monocyte-, macrophage-, and DC-associated TF programs in DC3s (Figures 3G–3I and S9D), with CD14⁺ but not CD16⁺ monocytes showing anti-correlated accessibility of monocyte- versus DC-associated motifs, suggestive of transitional states during monocyte-to-DC differentiation (Figures S9E and S9F). Increased RNA expression of AP-1 family members in DC3s from published scRNA-seq data³⁰ further supported these observations (Figure S9G). Collectively, DC3s harbor monocytic epigenetic features indicative of a shared ontogeny distinct from classical DCs.

To reconstruct lineage relationships between DC3s and other immune subsets independent of epigenetic profiles, we analyzed clonal frequency correlations across immune subtypes, reasoning that ontogenetically related populations are also clonally related. As expected, lineage proximity between circulating and tissue-infiltrating monocytes was recapitulated (Figure 3J). Strikingly, DC3s clustered closely with monocytes and MoMΦs (Figures 3K and 3L), whereas cDC2s were less clonally related to monocytes (Figure S9H). For example, a representative clone defined by the 3068G>A mtDNA mutation showed 35 CD14⁺ monocytes in the peripheral blood, 41 CD14⁺ monocytes in the tissue, and 13 DC3s in the tumor and NILT, likely reflecting continuous infiltration and differentiation (Figure 3M). These findings were recapitulated in ovarian tumors, where intratumoral DC3s showed similar monocyte/macrophage TF-motif enrichment, concordant chromVAR activity, and clonal proximity to circulating monocytes (Figures S10A–S10J), suggesting the monocyte-DC3 connection is tumor-agnostic.

Biased monocyte differentiation fates peripherally reprogram the tumor myeloid compartment

While circulating monocytes have the capacity to differentiate into either macrophages or DCs as they infiltrate inflamed tissue,^{53,57,63–65} it is not clear whether differentiation fates are entirely dictated by the tissue environment or there is cell-intrinsic bias in the context of human cancer. To address this question, we next asked whether myeloid clones displayed tissue-specific differentiation preferences by comparing their frequencies across lung tumor, NILT, and peripheral blood (Figures 4A, 4B, S11A, and S11B). Overall, clone frequencies were highly correlated across sites, and the largest clones in tumors and NILT were also detected in blood, indicating infiltration from circulation without tissue site-specific expansion biases. However, within clones, monocytes were proportionally more abundant in NILT, whereas differentiated macrophages and DCs were enriched in tumors (Figures 4C, 4D, and S11C). These observations are consistent with the notion that TME promotes monocyte differentiation, likely driven by elevated inflammatory cues.⁶⁵

To identify the specific myeloid subtypes that monocytes preferentially differentiate into within the TME, we analyzed clone compositions by subtype (Figures 4E, 4F, and S11D). Both CD14⁺ and CD16⁺ monocytes were less represented in tumors compared to NILTs, while SPP1⁺ MoMΦ2 and DC3s were preferentially enriched in tumors. Notably, MoMΦ1 was more abundant in NILTs. If differentiation were purely driven by the tissue environment, the same clone would display different myeloid subtype compositions between tumor and NILT. To test this hypothesis, we clustered clone-level subtype compositions

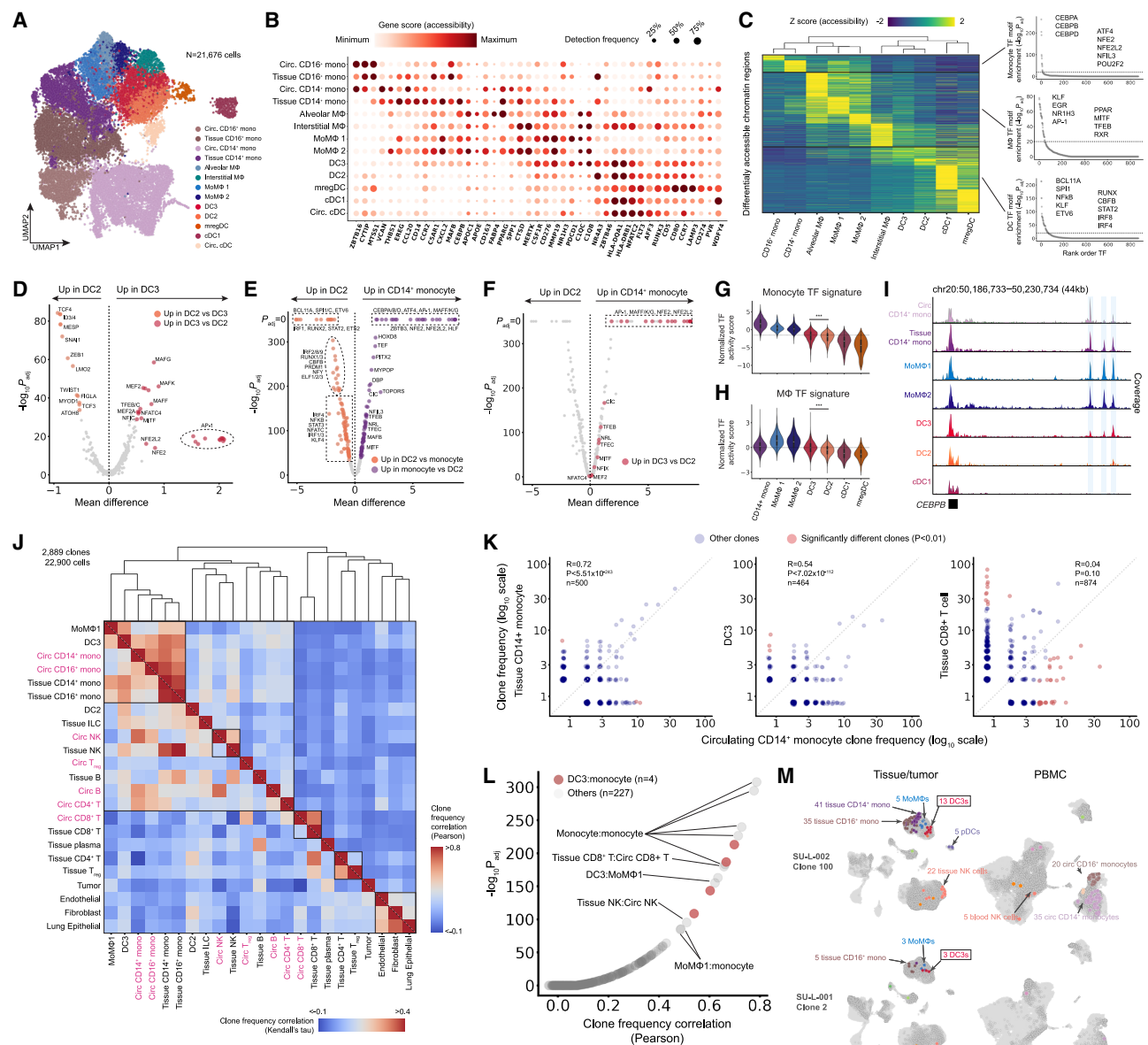


Figure 3. Intratumoral DC3s are epigenetically and clonally related to monocytes

(A) UMAP of myeloid cells from PBMC, lung tumor, and NILT samples of patients with lung adenocarcinoma (LUAD).
 (B) Column-scaled gene accessibility scores and detection frequencies for the indicated genes.
 (C) (Left) Heatmap showing marker peaks for all myeloid cell types. The color indicates column-scaled, mean-adjusted number of reads detected in each peak. (Right) Representative TF motifs that are significantly enriched for monocyte, macrophage, and DC.
 (D) Differentially active TF motifs between DC2 and DC3. p values are calculated using the Benjamini-Hochberg adjusted Kruskal-Wallis test.
 (E) Same as (D) but comparing DC2 and CD14⁺ monocytes.
 (F) Same as (D) but highlighting DC3-up TF motifs with increased accessibility.
 (G and H) Average chromVAR motif deviation scores for monocyte and macrophage TF motifs highlighted in (C). Kruskal-Wallis test.
 (I) Chromatin accessibility tracks for the CEBPB locus for the indicated cell types.
 (J) Cell type-cell type clone frequency correlation across clones (≥ 5 cells across all samples). Color denotes correlation value, computed using Pearson's ρ (upper half), and Kendall's τ (bottom half). Text labels of circulating PBMC cell types are colored pink.
 (K) Scatterplots comparing clone frequencies of circulating CD14⁺ monocytes with those of tissue CD14⁺ monocytes (left), DC3 (middle), and tissue CD8⁺ T cells (right).
 (L) 231 distinct cell type pairs ordered by clone frequency correlation. DC3-monocyte interactions are highlighted in red.
 (M) Representative clones consisting of DC3 and monocytes. For each clone, cell types with at least two cells are highlighted. See also Figures S8–S10.

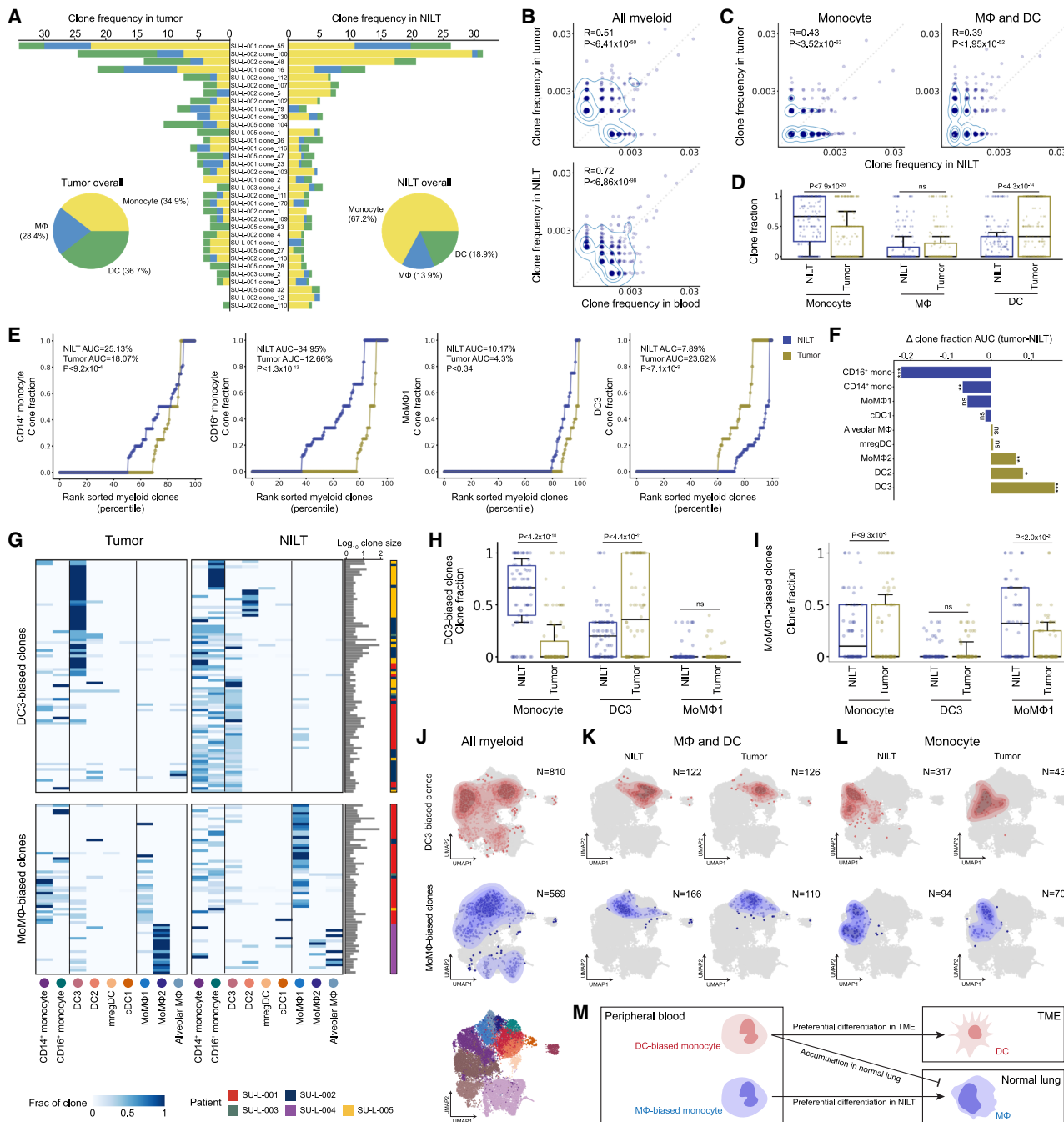


Figure 4. Divergent clonal myeloid differentiation fates in human tissues

(A) Monocyte, macrophage, and DC proportions of largest myeloid clones (≥ 10 cells), split by tumor and NILT.

(B) Scatterplots comparing clone frequencies of circulating myeloid cells with those in tumor (top) and NILT (bottom) myeloid cells. R and p value from Spearman's correlation.

(C) Scatterplots comparing clone frequencies of monocytes (left) and macrophage/DC (right) between NILT and tumor. R and p value from Spearman's correlation.

(D) Distribution of the cell type fraction within each myeloid clone (≥ 5 cells), split by tissue site. p values are from a Wald statistic from binomial generalized linear model (GLM) with a logit link. Boxplots: center line, median; box limits, first and third quartiles; whiskers, $1.5 \times$ interquartile range. (E) Cumulative fraction of clone sizes for the indicated myeloid cell types, split by tissue site. AUC corresponds to the overall clone size for the indicated cell type and tissue site.

(F) Summary of AUC differences between tumor and NILT for all myeloid cell types. A positive value indicates the cell type has larger clone sizes in lung tumors compared to NILTs. Significance is determined by p values from (D) and (E).

(G) Heatmaps showing cell type proportions split by tumor and NILT for DC-biased (top) and MΦ-biased (bottom) clones. Clone sizes correspond to the number of myeloid cells per clone. Only clones detected in both tissue sites or clones with at least three cells are used to identify cell-type biased clones.

(legend continued on next page)

stratified by tissue site (Figure 4G). Interestingly, we observed clone-intrinsic lineage biases, with clones consistently favoring differentiation toward either DCs or macrophages, independent of tissue site. DC-biased clones were enriched for DC3s in tumors and consisted primarily of monocytes in NILTs, with minimal macrophage representation across sites (Figure 4H). In contrast, macrophage-biased clones exhibited distinct behaviors depending on the MoMΦ subtype: MoMΦ1-biased clones were depleted of monocytes and enriched for MoMΦ1 cells in NILTs (Figure 4I), whereas MoMΦ2-biased clones almost exclusively consisted of MoMΦ2 cells in tumors. These findings suggest that monocytes possess intrinsic biases in both differentiation fate and site preference.

We next grouped myeloid cells from DC-biased and macrophage-biased clones to compare their epigenetic profiles. Across tissue sites, cells from the two groups clustered distinctly, even within the same myeloid subtype such as CD14⁺ monocytes in both tissue and periphery (Figures 4J–4M). Remarkably, circulating CD14⁺ monocytes from the two groups were epigenetically distinct even prior to tissue infiltration, with monocytes within each group more similar to one another than to those in the other group (Figures S11E–S11H). TF motif analysis revealed that DC-biased monocytes were enriched for proinflammatory programs (NF-κB, IRF, STAT2, and BLIMP-1), whereas macrophage-biased monocytes were enriched for immunosuppressive programs (MAF family, NRF2, and RUNX1/2)^{66–71} (Figure S11I). Upon tumor infiltration, IRF and BLIMP-1 motifs remained differentially accessible in DC-biased monocytes, with ID3/ID4 additionally becoming selectively accessible, supporting their anti-tumor potential⁷² (Figure S11J).

Corroborating these motif-level inferences, gene body accessibility analysis identified *IRF1*, *IRF3*, and *BLIMP-1* as differentially accessible in DC-biased monocytes and *MAF* (encoding c-MAF, a key driver of macrophage immunosuppressive phenotypes^{73–75}) in macrophage-biased monocytes, with cross-tissue patterns suggesting stage-specific roles before and after tissue infiltration (Figures S11K and S11L). To further support these inferences, we reanalyzed scRNA-seq data from a comprehensive atlas of myeloid cells in human lung tumors^{8,30} and performed pseudotime inference to identify fate-biased peripheral monocytes to both the macrophage and DC3 lineages (Figures S11M–S11P; methods). Differential expression analyses of peripheral CD14⁺ monocytes confirmed fate-based transcriptional programs, including up-regulation of *C1QA*, *C1QB*, and *C1QC* in macrophage-biased monocytes and *FCER1A* and *EREG* in DC3-biased monocytes⁷⁶ (Figure S11Q; full list in Table S3). Taken together, these findings suggest that circulating monocytes exhibit intrinsic epigenetic biases that predispose them to differentiate into either DCs or macrophages within the TME, contributing to their distinct functional roles in shaping the anti-tumor innate immune response.

DISCUSSION

Here, we performed mtscATAC-seq on 218,715 cells across 23 matched tumor, non-involved tissue, and peripheral blood samples from patients with lung and ovarian cancers to generate a clonally resolved map of the human innate immune response to cancer. To ensure the accuracy of clonal tracing in solid tissues, we developed Mitotrek, a tailored analysis framework that prioritizes clonal assignment fidelity over sensitivity or resolution. Our analyses revealed that innate immune cells in tumors are primarily derived from a diverse pool of circulating precursors without evidence of tissue-specific clonal expansion, in contrast to the antigen-driven expansion of adaptive immune cells. Epigenetic and clonal analyses further revealed that DC3s, but not cDC2s, exhibit monocytic regulatory features and are clonally related to monocytes in both circulation and tissue. Although a recent study in mice showed that DC3s arise from bone marrow progenitors shared with monocytes,⁷⁷ the rarity of DC3s in peripheral blood, their abundance in tumor lesions, and the lack of clonal expansion of DC3s in our data collectively suggest that human DC3s in tumors differentiate directly from monocytes upon tissue infiltration.

A key finding of our work is that phenotypically diverse tumor-infiltrating myeloid cells were clonally related to distinct subsets of circulating monocytes, with clones consistently biased toward either DC or macrophage fates in both tumor and non-involved tissue. These biases were accompanied by distinct chromatin accessibility patterns detectable even in peripheral blood monocytes prior to tissue infiltration: macrophage bias was linked to AP-1 family factors, c-MAF, and NRF2, whereas DC bias was linked to IRF, STAT2, and NF-κB motifs,⁷⁸ suggestive of priming by type I interferons, TNF-α, and IL-1 signaling.^{79,80} Consistent with this, a recent study identified a macrophage subset defined by *FOSL2* (an AP-1 factor) activity and derived from circulating monocytes as strongly associated with glioma malignancy,³⁸ supporting our observation that AP-1-enriched monocytes are predisposed to macrophage differentiation within the TME. These findings suggest that the tumor myeloid compartment may be peripherally programmed, offering an alternative to the view that intratumoral monocyte differentiation is primarily driven by local environmental cues, and raise the possibility that tumors epigenetically pre-condition monocytes through systemic cytokine signaling⁸¹ before tissue infiltration.

Limitations of this study

Several limitations of our study could be addressed in future work. First, our findings are based on a modest cohort of patients with lung and ovarian cancers; although we deeply profiled each patient and observed consistent patterns across patients and cancer types, validation in additional patient settings, including other tumor types, will be important to establish generalizability. Second, our approach does not capture the temporal dynamics

(H) Distribution of cell type fraction within DC3-biased clones.

(I) Same as in (H) but for MoMΦ1-biased clones.

(J) Contour plots, representing cell density of DC3-biased clones (top row, red) and MoMΦ-biased clones (bottom row, blue), projected onto the UMAP.

(K) Same as (J) but highlighting tissue-derived monocytes.

(L) Same as (J) but highlighting MΦ/DC.

(M) Schematic illustrating divergent clonal differentiation fate for myeloid cells. See also Figure S11 and Table S3.

of monocyte infiltration and differentiation. Longitudinal sampling or integration with additional progenitor populations would further establish the sequence of events that shape the tumor myeloid compartment. Lastly, our current approach excludes nuclear-derived mutations that could, in principle, contribute to resolving subclonal relationships but are poorly covered by our assay. Future technological advances may enable more reliable inference of subclonal architectures to corroborate the differentiation dynamics that we propose.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ansuman T. Satpathy (satpathy@stanford.edu).

Materials availability

This study did not generate new reagents.

Data and code availability

The ArchR analysis software for epigenetic analysis of scATAC-seq data is available on GitHub (<https://github.com/GreenleafLab/ArchR>). The mgatk software for processing sequencing data for single-cell mitochondrial variant calling is available on GitHub (<https://github.com/caleblareau/mgatk>). The Mitotrek software developed in this work for clone calling using single-cell mitochondrial variant data is available on GitHub (<https://github.com/vincent6liu/mitotrek>). Raw sequencing and processed chromatin accessibility and mtDNA mutation calls are available at the Gene Expression Omnibus accession **GSE302113**. Any additional custom code used for computational data processing and analysis is available from the authors upon request.

ACKNOWLEDGMENTS

We acknowledge the patients who volunteered to participate in these studies. We thank the members of the Satpathy Lab, L. Lanier, and Y. Lavin for stimulating discussions. A.T.S. was supported by a Career Award for Medical Scientists from the Burroughs Wellcome Fund, a Lloyd J. Old STAR Award from the Cancer Research Institute, an Emerging Leader Award from the Mark Foundation for Cancer Research, and the Parker Institute for Cancer Immunotherapy. A.T.S. is a Weill Cancer Hub West Investigator. This work was supported by NIH grants P30CA008748 (C.A.L.), R37CA303960 (B.E.H. and C.A.L.), R01HG012579 (C.A.L.), U01AT012984 (C.A.L.), and U01HG012076 (A.T.S. and L.S.L.). S.M.S. is supported by a Ruth L. Kirschstein Predoctoral Individual National Research Service Award (F31AR083275) and a recipient of the NIH Clinical and Translational Science Award (CTSA) through Rockefeller University. L.S.L. is supported by the German Research Foundation (DFG), including via the Emmy Noether fellowship (LU 2336/2-1) and Heisenberg Programme (LU 2336/10-1). C.A.L. and A.T.S. are supported by the Michelson Medical Research Foundation. This work was supported by the Department of Pathology at Stanford University.

AUTHOR CONTRIBUTIONS

V.L. conceived and designed the study with input from C.A.L. and A.T.S.; V.L., C.A.L., and A.T.S. supervised the work; V.L., K.S., P.K.Y., Z.M., Y.Y., R.R.S., K.H.-G., and W.Z. performed experiments and/or assisted with procurement of biospecimens; V.L. and C.A.L. led bioinformatics analyses with input from A.Y.C. and J.G.; S.M.S., R.V., S.W., B.D., and L.S.L. aided in the interpretation of data; B.E.H. provided pathological analyses of resected tissues; V.L., C.A.L., and A.T.S. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

V.L. has received consulting fees from Genesis Therapeutics and Cartography Biosciences. B.H. has received research support from Cartography Biosciences.

ences, Santa Ana Bio, Link Cell Sciences, and Immunogen, and has received consulting fees from Roche, Leica, Tempus, and Abbvie. C.A.L. is a consultant to Cartography Biosciences. A.T.S. is a founder of Immunai, Cartography Biosciences, Santa Ana Bio, and Arpelos Biosciences, an advisor to Wing Venture Capital and 10x Genomics, and receives research funding from Astellas.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Human samples
- **METHOD DETAILS**
 - Tissue processing
 - Fluorescence-activated cell sorting (FACS)
 - Preparation of mtscATAC-seq libraries
 - Sequencing and upstream processing of mtscATAC-seq data
 - Profiling and validation of mtDNA variants via single-cell TCR
 - scATAC-seq QC, dimensionality reduction and clustering
 - Annotation of mtscATAC-seq dataset
 - Gene signature scoring
 - Peak calling and motif analysis
 - Mitochondrial genotyping and clone calling
 - Benchmarking Mitotrek clone assignment
 - Clone sharing analysis
 - Comparison of clones using binomial generalized linear model with logit link
 - Corroboration via transcriptomic analyses
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.ccell.2026.05.006>.

Received: July 16, 2025

Revised: December 30, 2025

Accepted: May 8, 2026

REFERENCES

1. Pai, J.A., and Satpathy, A.T. (2021). High-throughput and single-cell T cell receptor sequencing technologies. *Nat. Methods* 18, 881–892.
2. Bradley, P., and Thomas, P.G. (2019). Using T cell receptor repertoires to understand the principles of adaptive immune recognition. *Annu. Rev. Immunol.* 37, 547–570.
3. Yost, K.E., Satpathy, A.T., Wells, D.K., Qi, Y., Wang, C., Kageyama, R., McNamara, K.L., Granja, J.M., Sarin, K.Y., Brown, R.A., et al. (2019). Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat. Med.* 25, 1251–1259.
4. Wu, T.D., Madireddi, S., de Almeida, P.E., Banchereau, R., Chen, Y.-J.J., Chitre, A.S., Chiang, E.Y., Ifthikhar, H., O’Gorman, W.E., Au-Yeung, A., et al. (2020). Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature* 579, 274–278.
5. Pai, J.A., Hellmann, M.D., Sauter, J.L., Mattar, M., Rizvi, H., Woo, H.J., Shah, N., Nguyen, E.M., Uddin, F.Z., Quintanal-Villalonga, A., et al. (2023). Lineage tracing reveals clonal progenitors and long-term persistence of tumor-specific T cells during immune checkpoint blockade. *Cancer Cell* 41, 776–790.e7.
6. Cassetta, L., and Pollard, J.W. (2018). Targeting macrophages: therapeutic approaches in cancer. *Nat. Rev. Drug Discov.* 17, 887–904.
7. van Vlerken-Ysla, L., Tyurina, Y.Y., Kagan, V.E., and Gabrilovich, D.I. (2023). Functional states of myeloid cells in cancer. *Cancer Cell* 41, 490–504.

8. Lavin, Y., Kobayashi, S., Leader, A., Amir, E.-A.D., Elefant, N., Bigenwald, C., Remark, R., Sweeney, R., Becker, C.D., Levine, J.H., et al. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell* 169, 750–765.e17.
9. Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., et al. (2019). Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* 179, 829–845.e20.
10. Steele, N.G., Carpenter, E.S., Kemp, S.B., Sirihorachai, V.R., The, S., Delrosario, L., Lazarus, J., Amir, E.-A.D., Gunchick, V., Espinoza, C., et al. (2020). Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. *Nat. Cancer* 1, 1097–1112.
11. Krishna, C., DiNatale, R.G., Kuo, F., Srivastava, R.M., Vuong, L., Chowell, D., Gupta, S., Vanderbilt, C., Purohit, T.A., Liu, M., et al. (2021). Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy. *Cancer Cell* 39, 662–677.e6.
12. Casanova-Acebes, M., Dalla, E., Leader, A.M., LeBerichel, J., Nikolic, J., Morales, B.M., Brown, M., Chang, C., Troncoso, L., Chen, S.T., et al. (2021). Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature* 595, 578–584.
13. Nalio Ramos, R., Missolo-Koussou, Y., Gerber-Ferder, Y., Bromley, C.P., Bugatti, M., Núñez, N.G., Tosello Boari, J., Richer, W., Menger, L., Denizeau, J., et al. (2022). Tissue-resident FOLR2+ macrophages associate with CD8+ T cell infiltration in human breast cancer. *Cell* 185, 1189–1207.e25.
14. Larsson, N.-G. (2010). Somatic mitochondrial DNA mutations in mammalian aging. *Annu. Rev. Biochem.* 79, 683–706.
15. Stewart, J.B., and Chinnery, P.F. (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.* 16, 530–542.
16. Ludwig, L.S., Lareau, C.A., Ulirsch, J.C., Christian, E., Muus, C., Li, L.H., Pelka, K., Ge, W., Oren, Y., Brack, A., et al. (2019). Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* 176, 1325–1339.e22.
17. Lareau, C.A., Ludwig, L.S., Muus, C., Gohil, S.H., Zhao, T., Chiang, Z., Pelka, K., Verboon, J.M., Luo, W., Christian, E., et al. (2021). Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* 39, 451–461.
18. Weng, C., Yu, F., Yang, D., Poeschla, M., Liggett, L.A., Jones, M.G., Qiu, X., Wahlster, L., Caulier, A., Hussmann, J.A., et al. (2024). Deciphering Cell States and Genealogies of Human Hematopoiesis. *Nature* 627, 389–398.
19. Lareau, C.A., Liu, V., Muus, C., Praktikjono, S.D., Nitsch, L., Kautz, P., Sandor, K., Yin, Y., Gutierrez, J.C., Pelka, K., et al. (2023). Mitochondrial single-cell ATAC-seq for high-throughput multi-omic detection of mitochondrial genotypes and chromatin accessibility. *Nat. Protoc.* 18, 1416–1440.
20. Jardim, D.L., Goodman, A., de Melo Gagliato, D., and Kurzrock, R. (2021). The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell* 39, 154–173.
21. Kahn, B.M., Lucas, A., Alur, R.G., Wengyn, M.D., Schwartz, G.W., Li, J., Sun, K., Maurer, H.C., Olive, K.P., Faryabi, R.B., and Stanger, B.Z. (2021). The vascular landscape of human cancer. *J. Clin. Investig.* 131, e136655. <https://doi.org/10.1172/JCI136655>.
22. Ghoneum, A., Afify, H., Salih, Z., Kelly, M., and Said, N. (2018). Role of tumor microenvironment in ovarian cancer pathobiology. *Oncotarget* 9, 22832–22849.
23. An, J., Nam, C.H., Kim, R., Lee, Y., Won, H., Park, S., Lee, W.H., Park, H., Yoon, C.J., An, Y., et al. (2024). Mitochondrial DNA mosaicism in normal human somatic cells. *Nat. Genet.* 56, 1665–1677.
24. De Rop, F.V., Hulselmans, G., Flerin, C., Soler-Vila, P., Rafels, A., Christiaens, V., González-Blas, C.B., Marchese, D., Caratù, G., Poovathingal, S., et al. (2024). Systematic benchmarking of single-cell ATAC-sequencing protocols. *Nat. Biotechnol.* 42, 916–926.
25. Borchering, N., and Brestoff, J.R. (2023). The power and potential of mitochondria transfer. *Nature* 623, 283–291.
26. Rückert, T., Lareau, C.A., Mashregi, M.-F., Ludwig, L.S., and Romagnani, C. (2022). Clonal expansion and epigenetic inheritance of long-lasting NK cell memory. *Nat. Immunol.* 23, 1551–1563.
27. Miller, T.E., Lareau, C.A., Verga, J.A., DePasquale, E.A.K., Liu, V., Ssozi, D., Sandor, K., Yin, Y., Ludwig, L.S., El Farran, C.A., et al. (2022). Mitochondrial variant enrichment from high-throughput single-cell RNA sequencing resolves clonal populations. *Nat. Biotechnol.* 40, 1030–1034.
28. Jenuth, J.P., Peterson, A.C., Fu, K., and Shoubridge, E.A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat. Genet.* 14, 146–151.
29. Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., Zhang, Y., Zhao, W., Zhou, F., Li, W., et al. (2021). Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* 12, 2540.
30. Leader, A.M., Grout, J.A., Maier, B.B., Nabet, B.Y., Park, M.D., Tabachnikova, A., Chang, C., Walker, L., Lansky, A., Le Berichel, J., et al. (2021). Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. *Cancer Cell* 39, 1594–1609.e12.
31. Wang, X., Wang, K., Zhang, W., Tang, Z., Zhang, H., Cheng, Y., Zhou, D., Zhang, C., Zhong, W.-Z., Ma, Q., et al. (2025). Clonal expansion dictates the efficacy of mitochondrial lineage tracing in single cells. *Genome Biol.* 26, 70.
32. Sun, J.C., Beilke, J.N., and Lanier, L.L. (2009). Adaptive immune features of natural killer cells. *Nature* 457, 557–561.
33. Hammer, Q., Rückert, T., Borst, E.M., Dunst, J., Haubner, A., Durek, P., Heinrich, F., Gasparoni, G., Babic, M., Tomic, A., et al. (2018). Peptide-specific recognition of human cytomegalovirus strains controls adaptive natural killer cells. *Nat. Immunol.* 19, 453–463.
34. Béziat, V., Liu, L.L., Malmberg, J.-A., Ivarsson, M.A., Sohlberg, E., Björklund, A.T., Retière, C., Sverremark-Ekström, E., Traherne, J., Ljungman, P., et al. (2013). NK cell responses to cytomegalovirus infection lead to stable imprints in the human KIR repertoire and involve activating KIRs. *Blood* 121, 2678–2688.
35. Strauss-Albee, D.M., Fukuyama, J., Liang, E.C., Yao, Y., Jarrell, J.A., Drake, A.L., Kinuthia, J., Montgomery, R.R., John-Stewart, G., Holmes, S., and Blish, C.A. (2015). Human NK cell repertoire diversity reflects immune experience and correlates with viral susceptibility. *Sci. Transl. Med.* 7, 297ra115. <https://doi.org/10.1126/scitranslmed.aac5722>.
36. MacLean, A.J., Bonifacio, J.P.P.L., Oram, S.L., Mohsen, M.O., Bachmann, M.F., and Arnon, T.I. (2024). Regulation of pulmonary plasma cell responses during secondary infection with influenza virus. *J. Exp. Med.* 221, e20232014. <https://doi.org/10.1084/jem.20232014>.
37. Park, M.D., Reyes-Torres, I., LeBerichel, J., Hamon, P., LaMarche, N.M., Hegde, S., Belabed, M., Troncoso, L., Grout, J.A., Magen, A., et al. (2023). TREM2 macrophages drive NK cell paucity and dysfunction in lung cancer. *Nat. Immunol.* 24, 792–801.
38. Yu, K.K.H., Abou-Mrad, Z., Törkenczy, K., Schulze, I., Gantchev, J., Baquer, G., Hopland, K., Bander, E.D., Tosi, U., Brennan, C., et al. (2026). Characterization of a pathogenic subpopulation of human glioma associated macrophages linked to glioma progression. *Cancer Cell* 44, 129–145.e12.
39. 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. <https://doi.org/10.1126/science.aah4573>.
40. Bourdely, P., Anselmi, G., Vaivode, K., Ramos, R.N., Missolo-Koussou, Y., Hidalgo, S., Tossello, J., Núñez, N., Richer, W., Vincent-Salomon, A., et al. (2020). Transcriptional and Functional Analysis of CD13c+ Human Dendritic Cells Identifies a CD163+ Subset Priming CD8+CD103+ T Cells. *Immunity* 53, 335–352.e8.

41. Brown, C.C., Gudjonson, H., Pritykin, Y., Deep, D., Lavallée, V.-P., Mendoza, A., Fromme, R., Mazutis, L., Ariyan, C., Leslie, C., et al. (2019). Transcriptional basis of mouse and human dendritic cell heterogeneity. *Cell* 179, 846–863.e24.
42. Dutertre, C.-A., Becht, E., Irac, S.E., Khalilnezhad, A., Narang, V., Khalilnezhad, S., Ng, P.Y., van den Hoogen, L.L., Leong, J.Y., Lee, B., et al. (2019). Single-cell analysis of human mononuclear phagocytes reveals subset-defining markers and identifies circulating inflammatory dendritic cells. *Immunity* 51, 573–589.e8.
43. Maier, B., Leader, A.M., Chen, S.T., Tung, N., Chang, C., LeBerichel, J., Chudnovskii, A., Maskey, S., Walker, L., Finnigan, J.P., et al. (2020). A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature* 580, 257–262.
44. Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H.D., Krishnan, I., Maroni, G., Meyerovitz, C.V., Kerwin, C.M., et al. (2019). Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity* 50, 1317–1334.e10.
45. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* 53, 403–411.
46. Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc. Natl. Acad. Sci. USA* 94, 569–574.
47. Kim, S., Chen, J., Ou, F., Liu, T.-T., Jo, S., Gillanders, W.E., Murphy, T.L., and Murphy, K.M. (2024). Transcription factor C/EBP α is required for the development of Ly6C hi monocytes but not Ly6C lo monocytes. *Proc. Natl. Acad. Sci. USA* 121, e2315659121. <https://doi.org/10.1073/pnas.2315659121>.
48. Alder, J.K., Georgantas, R.W., Hildreth, R.L., Kaplan, I.M., Morisot, S., Yu, X., McDevitt, M., and Civin, C.I. (2008). Kruppel-like factor 4 is essential for inflammatory monocyte differentiation in vivo. *J. Immunol.* 180, 5645–5652.
49. Feinberg, M.W., Wara, A.K., Cao, Z., Lebedeva, M.A., Rosenbauer, F., Iwasaki, H., Hirai, H., Katz, J.P., Haspel, R.L., Gray, S., et al. (2007). The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *EMBO J.* 26, 4138–4148.
50. Tussiwand, R., Everts, B., Grajales-Reyes, G.E., Kretzer, N.M., Iwata, A., Bagaitkar, J., Wu, X., Wong, R., Anderson, D.A., Murphy, T.L., et al. (2015). Klf4 expression in conventional dendritic cells is required for T helper 2 cell responses. *Immunity* 42, 916–928.
51. Carter, J.H., and Tourtellotte, W.G. (2007). Early growth response transcriptional regulators are dispensable for macrophage differentiation. *J. Immunol.* 178, 3038–3047.
52. Kirschenbaum, D., Xie, K., Ingelfinger, F., Katzenelenbogen, Y., Abadie, K., Look, T., Sheban, F., Phan, T.S., Li, B., Zwicky, P., et al. (2024). Time-resolved single-cell transcriptomics defines immune trajectories in glioblastoma. *Cell* 187, 149–165.e23.
53. Devalaraja, S., To, T.K.J., Folkert, I.W., Natesan, R., Alam, M.Z., Li, M., Tada, Y., Budagyan, K., Dang, M.T., Zhai, L., et al. (2020). Tumor-derived retinoic acid regulates intratumoral monocyte differentiation to promote immune suppression. *Cell* 180, 1098–1114.e16.
54. Sharma, S.M., Bronisz, A., Hu, R., Patel, K., Mansky, K.C., Sif, S., and Ostrowski, M.C. (2007). MITF and PU.1 recruit p38 MAPK and NFATc1 to target genes during osteoclast differentiation. *J. Biol. Chem.* 282, 15921–15929.
55. Liao, X., Sharma, N., Kapadia, F., Zhou, G., Lu, Y., Hong, H., Paruchuri, K., Mahabeshwar, G.H., Dalmás, E., Venteclef, N., et al. (2011). Kruppel-like factor 4 regulates macrophage polarization. *J. Clin. Investig.* 121, 2736–2749.
56. A-González, N., and Castrillo, A. (2011). Liver X receptors as regulators of macrophage inflammatory and metabolic pathways. *Biochim. Biophys. Acta* 1812, 982–994.
57. Goudot, C., Coillard, A., Villani, A.-C., Gueguen, P., Cros, A., Sarkizova, S., Tang-Huau, T.-L., Bohec, M., Baulande, S., Hacohen, N., et al. (2017). Aryl hydrocarbon receptor controls monocyte differentiation into dendritic cells versus macrophages. *Immunity* 47, 582–596.e6.
58. Trizzino, M., Zucco, A., Deliard, S., Wang, F., Barbieri, E., Veglia, F., Gabrilovich, D., and Gardini, A. (2021). EGR1 is a gatekeeper of inflammatory enhancers in human macrophages. *Sci. Adv.* 7, eaaz8836. <https://doi.org/10.1126/sciadv.aaz8836>.
59. Moore, K.J., Rosen, E.D., Fitzgerald, M.L., Randow, F., Andersson, L.P., Altschuler, D., Milstone, D.S., Mortensen, R.M., Spiegelman, B.M., and Freeman, M.W. (2001). The role of PPAR-gamma in macrophage differentiation and cholesterol uptake. *Nat. Med.* 7, 41–47.
60. Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159, 1312–1326.
61. Murphy, T.L., Grajales-Reyes, G.E., Wu, X., Tussiwand, R., Briseño, C.G., Iwata, A., Kretzer, N.M., Durai, V., and Murphy, K.M. (2016). Transcriptional control of dendritic cell development. *Annu. Rev. Immunol.* 34, 93–119.
62. Satpathy, A.T., Briseño, C.G., Cai, X., Michael, D.G., Chou, C., Hsiung, S., Bhattacharya, D., Speck, N.A., and Egawa, T. (2014). Runx1 and Cbfb regulate the development of Flt3+ dendritic cell progenitors and restrict myeloproliferative disorder. *Blood* 123, 2968–2977.
63. Villar, J., Cros, A., De Juan, A., Alaoui, L., Bonte, P.-E., Lau, C.M., Tiniakou, I., Reizis, B., and Segura, E. (2023). ETV3 and ETV6 enable monocyte differentiation into dendritic cells by repressing macrophage fate commitment. *Nat. Immunol.* 24, 84–95.
64. Williams, M., Mildner, A., and Yona, S. (2018). Developmental and functional heterogeneity of monocytes. *Immunity* 49, 595–613.
65. Jakubzick, C.V., Randolph, G.J., and Henson, P.M. (2017). Monocyte differentiation and antigen-presenting functions. *Nat. Rev. Immunol.* 17, 349–362.
66. Kusiak, A., and Brady, G. (2022). Bifurcation of signalling in human innate immune pathways to NF- κ B and IRF family activation. *Biochem. Pharmacol.* 205, 115246.
67. Kim, S.J. (2015). Immunological function of Blimp-1 in dendritic cells and relevance to autoimmune diseases. *Immunol. Res.* 63, 113–120.
68. Kobayashi, E.H., Suzuki, T., Funayama, R., Nagashima, T., Hayashi, M., Sekine, H., Tanaka, N., Moriguchi, T., Motohashi, H., Nakayama, K., and Yamamoto, M. (2016). Nrf2 suppresses macrophage inflammatory response by blocking proinflammatory cytokine transcription. *Nat. Commun.* 7, 11624.
69. Bellissimo, D.C., Chen, C.-H., Zhu, Q., Bagga, S., Lee, C.-T., He, B., Wertheim, G.B., Jordan, M., Tan, K., Worthen, G.S., et al. (2020). Runx1 negatively regulates inflammatory cytokine production by neutrophils in response to Toll-like receptor signaling. *Blood Adv.* 4, 1145–1158.
70. Tomimatsu, M., Matsumoto, K., Ashizuka, M., Kumagai, S., Tanaka, S., Nakae, T., Yokota, K., Kominami, S., Kajiwara, R., Okuzaki, D., et al. (2022). Myeloid cell-specific ablation of Runx2 gene exacerbates post-infarct cardiac remodeling. *Sci. Rep.* 12, 16656.
71. Zezulín, A.U., Yen, D., Ye, D., Howell, E.D., Bresciani, E., Diemer, J., Ren, J.-G., Ahmad, M.H., Castilla, L.H., Touw, I.P., et al. (2023). RUNX1 is required in granulocyte-monocyte progenitors to attenuate inflammatory cytokine production by neutrophils. *Genes Dev.* 37, 605–620.
72. Deng, Z., Loyher, P.-L., Lazarov, T., Li, L., Shen, Z., Bhinder, B., Yang, H., Zhong, Y., Alberdi, A., Massague, J., et al. (2024). The nuclear factor ID3 endows macrophages with a potent anti-tumour activity. *Nature* 626, 864–873.
73. Cao, S., Liu, J., Song, L., and Ma, X. (2005). The protooncogene c-Maf is an essential transcription factor for IL-10 gene expression in macrophages. *J. Immunol.* 174, 3484–3492.

74. Liu, M., Tong, Z., Ding, C., Luo, F., Wu, S., Wu, C., Albeituni, S., He, L., Hu, X., Tieri, D., et al. (2020). Transcription factor c-Maf is a checkpoint that programs macrophages in lung cancer. *J. Clin. Investig.* **130**, 2081–2096.
75. Cao, S., Liu, J., Chesi, M., Bergsagel, P.L., Ho, I.-C., Donnelly, R.P., and Ma, X. (2002). Differential regulation of IL-12 and IL-10 gene expression in macrophages by the basic leucine zipper transcription factor c-Maf fibrosarcoma. *J. Immunol.* **169**, 5715–5725.
76. Odell, I.D., Steach, H., Gauld, S.B., Reinke-Breen, L., Karman, J., Carr, T.L., Wetter, J.B., Phillips, L., Hinchcliff, M., and Flavell, R.A. (2022). Epiregulin is a dendritic cell-derived EGFR ligand that maintains skin and lung fibrosis. *Sci. Immunol.* **7**, eabq6691.
77. Liu, Z., Wang, H., Li, Z., Dress, R.J., Zhu, Y., Zhang, S., De Feo, D., Kong, W.T., Cai, P., Shin, A., et al. (2023). Dendritic cell type 3 arises from Ly6C+ monocyte-dendritic cell progenitors. *Immunity* **56**, 1761–1777.e6.
78. Shih, V.F.-S., Davis-Turak, J., Macal, M., Huang, J.Q., Ponomarenko, J., Kearns, J.D., Yu, T., Fagerlund, R., Asagiri, M., Zuniga, E.I., and Hoffmann, A. (2012). Control of RelB during dendritic cell activation integrates canonical and noncanonical NF- κ B pathways. *Nat. Immunol.* **13**, 1162–1170.
79. Park, S.H., Kang, K., Giannopoulou, E., Qiao, Y., Kang, K., Kim, G., Park-Min, K.-H., and Ivashkiv, L.B. (2017). Type I interferons and the cytokine TNF cooperatively reprogram the macrophage epigenome to promote inflammatory activation. *Nat. Immunol.* **18**, 1104–1116.
80. Sheu, K.M., and Hoffmann, A. (2022). Functional hallmarks of healthy macrophage responses: Their regulatory basis and disease relevance. *Annu. Rev. Immunol.* **40**, 295–321.
81. Hiam-Galvez, K.J., Allen, B.M., and Spitzer, M.H. (2021). Systemic immunity in cancer. *Nat. Rev. Cancer* **21**, 345–359.
82. Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978.
83. Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom.* **19**, 477.
84. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-human CD45 V500 (clone HI30)	BD Biosciences	Cat# 560779; RRID: AB_1937324; Lot 7172744
Anti-human CD3 FITC (clone OKT3)	Invitrogen	Cat# 11-0037-41; RRID: AB_2533240; Lot 2007722
Biological samples		
Fresh human lung tumor and NILT	Stanford Tissue Procurement Shared Resource	N/A
Fresh human ovarian tumor	Stanford Tissue Procurement Shared Resource	N/A
Human peripheral blood mononuclear cells (PBMC)	This paper	N/A
Critical commercial assays		
Chromium Next GEM Single Cell ATAC Reagent Kits v1.1	10× Genomics	Cat# PN-1000209
Chromium Single Cell 5' kit	10× Genomics	Cat# PN-1000695
Nuclei Isolation for Single Cell ATAC Sequencing	10× Genomics	CG000169
Deposited data		
Raw and processed mtscATAC-seq and mtDNA mutation calls	This paper	GEO: GSE302113
Software and algorithms		
Mitotrek	This paper	https://github.com/vincent6liu/mitotrek
mgatk	Lareau et al. ¹⁷	https://github.com/caleblareau/mgatk
ArchR	Granja et al. ⁴⁵	https://github.com/GreenleafLab/ArchR
CellRanger-ATAC Pipeline v.2.0.0	10× Genomics	https://support.10xgenomics.com
chromVAR	Schep et al. ⁸²	https://github.com/GreenleafLab/chromVAR
Slingshot	Street et al. ⁸³	https://github.com/kstreet13/slinsshot
Seurat	Hao et al. ⁸⁴	https://satijalab.org/seurat/
Azimuth	Hao et al. ⁸⁴	https://azimuth.hubmapconsortium.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human samples

Fresh ovarian tumors, lung tumors, NILT samples, and peripheral blood were collected at the time of surgery by Stanford Tissue Procurement Shared Resource facility with the appropriate written informed consent and institutional IRB approval. Summary statistics and patient history are available in [Table S1](#). For the early-stage NSCLC dataset, exclusion criteria included previous systemic treatment or radiotherapy. For the ovarian cancer dataset, exclusion criteria included tumors of non-ovary or unknown origin. We note that one patient (SU-L-003) had a neuroendocrine tumor and was receiving olaparib (PARP inhibitor) at the time of surgery for concurrent tubo-ovarian carcinosarcoma, which was confirmed as an independent primary tumor by pathology studies ([Figure S1A](#)).

METHOD DETAILS

Tissue processing

For each patient, we collected tumor resections, peripheral blood samples, and non-adjacent non-neoplastic tissue (confirmed by pathological analysis) from the same resected lobe. Because the immediately adjacent tissue often exhibits pathological alterations, we excluded these tissues from genomic profiling. All tumor and non-adjacent healthy tissues were procured following surgical

resection. Samples were dissociated and viably cryopreserved for downstream library preparation and sequencing. In brief, solid tumor specimens on ice were minced to pieces of <1 mm³ and transferred to 5 mL of digestion medium containing DNase I (100 µg/mL) and collagenase P (2 mg/mL) in Advanced DMEM/F-12. Minced tissue was transferred into C-tubes for use in the gentleMACS Octo Dissociator system at 37°C at 20 rpm for 20 min. After digestion, the cell suspension was filtered through a 70-µm filter, which was washed with an additional 10 mL of DMEM/F-12. The sample was then centrifuged at 400 g at 4°C for 5 min. Any residual undigested tissue was further digested for an additional 20-min incubation with additional digestion medium. After centrifugation, the supernatant was discarded, and the pellet was resuspended in 500 µL of ACK red blood cell lysis buffer and incubated for 1 min on ice, followed by the addition of ice-cold PBS. The cell count and viability were determined by trypan blue staining by using a Countess II FL automated cell counter, before proceeding to cell-sorting.

Fluorescence-activated cell sorting (FACS)

Cells were classified into T cells (CD45⁺CD3⁺), other leukocytes (CD45⁺CD3⁻), and malignant or stromal cells (CD45⁻CD3⁻). The antibodies used included anti-human CD45 conjugated to V500 (clone HI30, 560779, lot 7172744, BD Biosciences) and anti-human CD3 conjugated to fluorescein isothiocyanate (FITC) (clone OKT3, 11-0037-41, lot 2007722, Invitrogen), both diluted at 1:200. Live/dead staining was performed using propidium iodide (P3566, Invitrogen) at a final concentration of 2.5 µg/mL. Cell sorting was conducted on a BD FACSria III cell sorter (BD Biosciences).

Preparation of mtscATAC-seq libraries

For the generation of mtscATAC-seq libraries, we adapted the 10× Genomics scATAC-seq platform NextGEM v1.1 kits. In brief, mtscATAC-seq was performed with modifications to the “Nuclei Isolation for Single Cell ATAC Sequencing” (CG000169 Rev D) user guide, where we fixed and permeabilized cells to retain mitochondria and mtDNA within their host cell by removing Tween 20 as part of the lysis buffer.¹⁹ For the library preparation we followed the “Chromium Next GEM Single Cell ATAC Reagent Kits v1.1” (CG000209 Rev F), user guide with only minor modifications as described and highlighted below and otherwise refer the reader to the original and highly detailed workflow by 10× Genomics.¹⁹

Sequencing and upstream processing of mtscATAC-seq data

All libraries were sequenced on an Illumina Novaseq 6000 device using a 10 × 16 × 151 × 151 read configuration to accommodate the 10× ATAC cell barcode in the i5 piece of the read. Libraries were sequenced to a target of 30,000–35,000 reads/cell as previously recommended.¹⁹ Raw.bcl files were converted into per-sample.fastq files using Illumina bcl2fastq.

Initial processing of mtscATAC-seq data was performed using the CellRanger-ATAC Pipeline v.2.0.0 by mapping scATAC-seq reads with cellranger-atac count to the GRCh38 reference genome, hardmasked for regions that would otherwise interfere with mapping to the mitochondrial genome (as previously detailed¹⁷). The outputs included fragments files for downstream epigenomics analyses and bam files for mtDNA genotyping.

Mitochondrial genotyping was performed on these fragment files with mgatk in tenx mode using the barcodes identified as cells by CellRanger (i.e., cells passing ATAC filter). Only cells with at least 10× coverage of the mitochondrial genome were included in the analysis, achieving a median coverage of 30×–70× across experiments. Across all quality-controlled cells, we observed a mean per-cell mitochondrial genome coverage of 37.8, with a mean coverage of 39.2 among cells assigned to clones used for downstream clonal analyses.

Profiling and validation of mtDNA variants via single-cell TCR

For one ovarian donor (SU-O-002), excess CD45⁺ tumor and peripheral blood mononuclear cells were available for additional profiling to corroborate mtDNA-derived clonotypes. To achieve this, single cells were partitioned into GEMs (gel bead-in-emulsions) on a Chromium Controller (10× Genomics; Single Cell 5' kit, Cat. PN-1000695) following the manufacturer's protocol with TCR profiling. Reverse transcription was performed in-GEM, followed by emulsion breakage, cDNA cleanup, and amplification. Separate gene-expression and TCR libraries were constructed per 10× recommendations. Library quality and size distributions were verified by TapeStation. Pooled libraries were sequenced on an Illumina Novaseq 6000 instrument for 151 × 151 R1 and R2 read lengths. Raw.fastq files were processed using CellRanger v8 default parameters for both gene expression and VDJ quantification to the hg38 reference.

For downstream analyses, we used mgatk to generate a per-cell, per-mutation heteroplasmy matrix.¹⁹ Analyzing 5' scRNA-seq data for mtDNA mutations incurs several limitations as only ~5% of the mitochondrial genome is transcribed and well-covered by this kit¹⁷ and the high error rate of mitochondrial transcription leads to many more spurious mutations.¹⁶ Hence, we restricted our analyses to a subset of high-quality cells with an overall coverage of 20× at well-covered sites detected from the 5' kit and containing valid TCRs and high-quality gene expression data (i.e., at least 1,000 UMIs and 500 distinct genes detected). For these cells, we considered TCRs with at least 3 cells and performed a supervised Mann-Whitney statistic to determine mutations enriched within a specific TCR, which is the same workflow that we previously used for Smart-seq2 data for T cells.¹⁶ Figure S1D shows the 10 mutations with the highest statistic where all cells with a mutation detected and a valid TCR were plotted. As many T cells may not be individually barcoded by mtDNA and only ~5% of the mitochondrial genome can be profiled with the 5' kit, we specifically do not report overlap statistics but use these data to illustrate the concordance of mtDNA variation with orthogonal clonal markers.

scATAC-seq QC, dimensionality reduction and clustering

CellRanger output fragment files were loaded and converted to Arrow files using the `createArrowFiles` function in ArchR. Quality control metrics were computed for each cell, and only cells with TSS enrichments greater than 4 were kept for all samples. Cells were also filtered based on the number of unique fragments sequenced using a cut-off of 1000. Doublet scores for all cells were computed using the ArchR functions `addDoubletScores` with $k = 10$, `knnMethod = "LSI"`, and `LSIMethod = 1`.

Sample groupings were defined based on cell types, tissue sites, and disease status (e.g., all cells, lung tumor and NILT samples from patients with NSCLC in Figure 1; myeloid cells, all samples from patients with NSCLC in Figure 3). An ArchR project was then created for each of the sample groupings and doublets were filtered with `filterDoublets` with a filter ratio of 1. For each ArchR project, dimensionality reduction was performed with `addIterativeLSI` using default parameters to embed ATAC-data in latent semantic indexing (LSI) space. Next, clustering was performed with `addClusters` using default parameters.

Annotation of mtscATAC-seq dataset

An iterative clustering approach was used to annotate cells, where after each round of clustering, select clusters with relatively high epigenetic similarity (e.g., T and NK cells) are merged and reclustered to achieve desired granularity and higher clustering accuracy. For the NSCLC lung tumor/NILT data sample grouping and ovarian tumor sample grouping, clusters were annotated based on gene score of known marker genes, including *EPCAM*, *KRT18* (epithelial and tumor cells), *VWF*, *PECAM1* (endothelial cells), *COL1A2*, *FBLN1* (fibroblasts), *CD3D*, *CD4*, *FOXP3*, *CD8A*, *KLRD1*, *NCR1* (T and NK cells), *MS4A1*, *PAX5* (B cells), *TNFRSF17*, *VOPP1* (plasma cells), *CD14*, *LYZ* (monocytes), *APOC1*, *CD163* (macrophages), *HLA-DQA1*, *ZBTB46* (DCs), *CLEC4C* (PDCs), *TPSAB1*, *KIT* (mast cells).

9 initial clusters in the lung tumor/NILT data annotated as epithelial or tumor cells were grouped and reclustered, resulting in 22 clusters. 9 clusters were annotated as tumor for satisfying the following criteria 1) highly patient specific and 2) enriched over 4-fold in tumor compared to NILT. The other 13 clusters were annotated as lung epithelial subtypes based on high gene score of following signatures *AGER*, *PDPN*, *CLIC5* (alveolar type 1), *SFTPB*, *SFTPC*, *SFTPD*, *MUC1*, *ETV5* (alveolar type 2), *FOXJ1*, *TUBB1*, *TP73*, *CCDC78* (ciliated), *KRT5*, *KRT17*, *MIR205HG* (basal), *MUC5B*, *SCGB1A1*, *BPIFB1*, *PIGR*, *SCGB3A1* (secretory).

6 initial clusters in the lung tumor/NILT data annotated as T/NK cells were further divided into 19 clusters. 5 initial clusters in the ovarian tumor data annotated as T/NK cells were further divided into 13 clusters. These T/NK clusters were annotated using known marker genes, including *CD3D* (broad T), *CD8A* ($CD8^+$ T), *CD4* ($CD4^+$ T), *GNLY*, *NCR1* (NK/ILC), *FOXP3* (T_{reg}). To support the marker gene-based annotation, cells were projected to human PBMC reference data using the Azimuth application.⁸⁴ The Azimuth-predicted cell types were dominated by $CD4^+$ T, $CD8^+$ T, T_{reg} , NK, and ILC, which are all associated with distinct clusters and consistent with the mark-gene based annotations. One cluster in the lung data is marked by the gene score pattern of *CD3D*⁺, *CD8A*⁺, *CD4*⁺, *CD56*⁺ and high gene score of NK signature (*GNLY*, *PRF1*, *GZMB*, *KLRB1*, *CCL3*, *KLRF1*, *NCR1*), which was defined as invariant NKT cells.

For the myeloid reclustering annotation, given that DCs comprised a minor fraction of myeloid cells in peripheral blood relative to lung tumors and NILTs (4.4% vs. 27.4%, respectively), we were unable to fully resolve DC subtypes in blood and thus grouped circulating DCs as a single population.

Gene signature scoring

Gene scores for individual genes are computed as implemented in ArchR. When computing the composite gene score for a gene signature, gene scores of individual genes in the signature were Z score normalized across all cells, and each cell is then scored by taking the mean z-scaled gene scores in the gene signature.

Peak calling and motif analysis

For the NSCLC myeloid epigenetic analysis, peak calling was performed as implemented in ArchR. Both level1 (monocyte, macrophage, DC) and level 2 (e.g., $CD14^+$ monocyte, DC3, MoMΦ1, etc) cell type annotation was used as grouping in `addReproduciblePeakSet()` to identify regulatory elements associated with both broad myeloid cell types as well as subtypes. Differential peak analysis was performed using `getMarkerFeatures`. TF motifs enriched in peak sets are identified using `peakAnnoEnrichment`. Per-cell TF motif activities were calculated using `chromVAR`.⁸²

Mitochondrial genotyping and clone calling

By default, `mgatk` calls high-confidence heteroplasmic variants by selecting variants with strand correlation > 0.65 and variance-mean ratio > 0.01. In this study, most donors have multiple samples from different tissue sites. Given that the default `mgatk` filters are already conservative, a union approach to selecting high-confidence heteroplasmic variants is employed to increase sensitivity, where if a variant passes the default `mgatk` filters in any sample, then it is considered a valid, informative variant in all other samples from the same donor (e.g., if a variant is determined to be high quality in the lung tumor sample, it is unlikely to be spurious if it is also observed in the peripheral blood sample from the same donor). We noted one polymorphic and highly homologous region (CCCTCCC in GRCh38 chrM:307–314) that led to spurious connections under the relaxed filtering criteria and explicitly disregarded variants from this region. This mitochondrial variant processing procedure for clone calling outputs a cell by variant heteroplasmy matrix combining all samples for each donor and is implemented in the `mitotrek.processing` module.

Clones are then called for each donor using `mitotrek.core.assign_cell_to_clones`. Variants present in >20% cells from a donor (likely technical or homoplasmic) or less than 3 cells are removed. The heteroplasmy matrix is then binarized with a cutoff of 0.07

based on the rationale that exact heteroplasmy levels are not reliable given the stochasticity from mitochondrial genome distribution during cell division and variance in per-cell mitochondrial genome coverage. The heteroplasmy cutoff value is chosen based on benchmarking experiments using published data (Figures 1C and 1D). Although our filtering strategy effectively removes variants that are most likely to represent technical artifacts or broadly distributed homoplasmic mutations, it also reflects an inherent trade-off between accuracy and sensitivity.

To convert mitochondrial variants to clones, an undirected weighted graph is constructed in which vertices are variants and edges are defined by the Pearson correlation coefficient between two variants across cells, computed from the binarized heteroplasmy matrix. After removing edges with weights less than 0.5, each connected component in the graph was treated as a distinct clonotype. Most resulting connected components contained only one variant, whereas highly correlated variants (i.e., frequently co-occurring in cells) were grouped into one connected component. For clones defined by a single variant, any cell positive for the associated variant in the binarized heteroplasmy matrix is assigned to the clone. For clones defined by multiple variants, a cell is required to be positive for all associated variants. Finally, cells assigned to multiple clones are discarded.

Benchmarking Mitotrek clone assignment

Full-length scRNA-seq via the Smart-seq2 technology of single-cell derived colonies from two donors published by Ludwig et al.¹⁶ was downloaded as fastq data from the Gene Expression Omnibus (accession GSE115214). Mitochondrial genotyping was performed using mgatk and cells with at least 100× mitochondrial genome coverage were retained for downstream analysis. Clones were called using mitotrek.core.assign_cell_to_clones with a binarization cutoff threshold set from 0.01 to 0.15. For each ground truth clone label (established in the experimental protocol by physical separation of the clones), the predicted clone label was determined by taking the mode of called clones among cells in the ground truth clone to compute accuracy, which is implemented in mitotrek.core.clone_calling_accuracy.

Clone sharing analysis

We consider two mutually exclusive sets of cells A and B (e.g., two cell types). There are N clones c_1, c_2, \dots, c_N containing at least one cell in $A \cup B$. Note that each clone c_i belongs to a set P_j which contains all clones detected in donor j . We summarize the clone counts in the two groups as a $N \times 2$ matrix X where $X_{i,1}$ and $X_{i,2}$ correspond to number of clone c_i cells in A and B , respectively. The fraction of clones shared between the two groups is computed as.

$$\frac{\sum_{c_i} X_{i,1} \times X_{i,2}}{\sum_{P_j} D_{j,1} \times D_{j,2}} \text{ where } D_{j,1} = \sum_{c_i \in P_j} X_{i,1} \text{ and } D_{j,2} = \sum_{c_i \in P_j} X_{i,2} \text{ for all } c_i \in P_j$$

The numerator computes the number of observed cell pairs between two groups that share a clone. The denominator is a normalization factor that represents all possible cell pairs between two groups, accounting for only cell pairs within the same donor, since cross-donor cell pairs are not valid.

The fraction of shared clonotypes provides a more sensitive and robust measure than correlation coefficients, which can be disproportionately influenced by outliers, capturing relative clone sharing while appropriately accounting for the number of cells detected in each cell type.

Comparison of clones using binomial generalized linear model with logit link

To account for differences in clone size when comparing clonal lineage biases, we analyzed clonal proportions using a binomial generalized linear model (GLM) with a logit link. For each clone, the number of cells of a given lineage was modeled as a binomial outcome, with the total number of cells assigned to that clone used as the binomial denominator. This framework weights clones by their size, such that larger clones contribute proportionally more information than smaller clones, while testing for differences in underlying lineage bias between conditions or groups. Unless otherwise noted, all statistical comparisons of clonal proportions reported in the manuscript are based on this binomial GLM-based inference.

Corroboration via transcriptomic analyses

Using a subset of 45,524 cells profiled from lung tumors and non-involved tissues from Leader et al.,³⁰ we defined the major differentiation trajectories for CD14⁺ monocytes to DC3s or monocyte-derived macrophages using Slingshot⁸³ (Figures S11M–S11O). From this embedding, we could characterize 1,708 DC3-biased and 1,854 macrophage-biased cells that were annotated as CD14⁺ monocytes from the periphery (defined as a 2-fold biased pseudotime for either lineage; Figure S11P). Differential expression analyses between these putative fate-biased monocytes were performed using the FindMarkers function in Seurat with default parameters. These transcriptomic analyses add further support to the clonal relationships that we described using mtscATAC-seq, where monocytes adopt a fate-biased epigenomic program, including in the periphery, and the cell state changes can be verified via transcriptional changes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of single-cell sequencing data was performed in python (v3.9.4) and R (v4.3.5). Statistical analysis of flow cytometry data was performed in GraphPad Prism (v9.0).

Cancer Cell, Volume 44

Supplemental information

Clonal lineage tracing of innate immune cells in human cancer

Vincent Liu, Katalin Sandor, Patrick K. Yan, Zhuang Miao, Yajie Yin, Robert R. Stickels, Andy Y. Chen, Kamir Hiam-Galvez, Jacob Gutierrez, Wenxi Zhang, Sairaj M. Sajjath, Raeline Valbuena, Steven Wang, Bence Daniel, Leif S. Ludwig, Brooke E. Howitt, Caleb A. Lareau, and Ansuman T. Satpathy

Supplemental information for

Clonal lineage tracing of innate immune cells in human cancer

Vincent Liu¹⁻³, Katalin Sandor²⁻³, Patrick K. Yan²⁻³, Zhuang Miao¹⁻³, Yajie Yin²⁻³, Robert R. Stickels²⁻³, Andy Y. Chen²⁻⁴, Kamir Hiam-Galvez²⁻³, Jacob Gutierrez⁵, Wenxi Zhang²⁻³, Sairaj M. Sajjath⁶, Raeline Valbuena¹, Steven Wang¹, Bence Daniel²⁻³, Leif S. Ludwig^{7,8}, Brooke E. Howitt², Caleb A. Lareau⁵, Ansuman T. Satpathy^{2-3,9-10}

¹Department of Genetics, Stanford University, Stanford, CA 94305, USA

²Department of Pathology, Stanford University, Stanford, CA 94305, USA

³Center for Immunotherapy Design, Stanford University, Stanford, CA 94305, USA

⁴Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

⁵Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

⁶Robin Chemers Neustein Laboratory of Mammalian Cell Biology and Development, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA

⁷Berlin Institute of Health at Charité – Universitätsmedizin Berlin, 10178 Berlin, Germany

⁸Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC) Berlin Institute for Medical Systems Biology (BIMSB), 10115 Berlin, Germany

⁹Parker Institute for Cancer Immunotherapy, San Francisco, CA 94129, USA

¹⁰Lead Contact

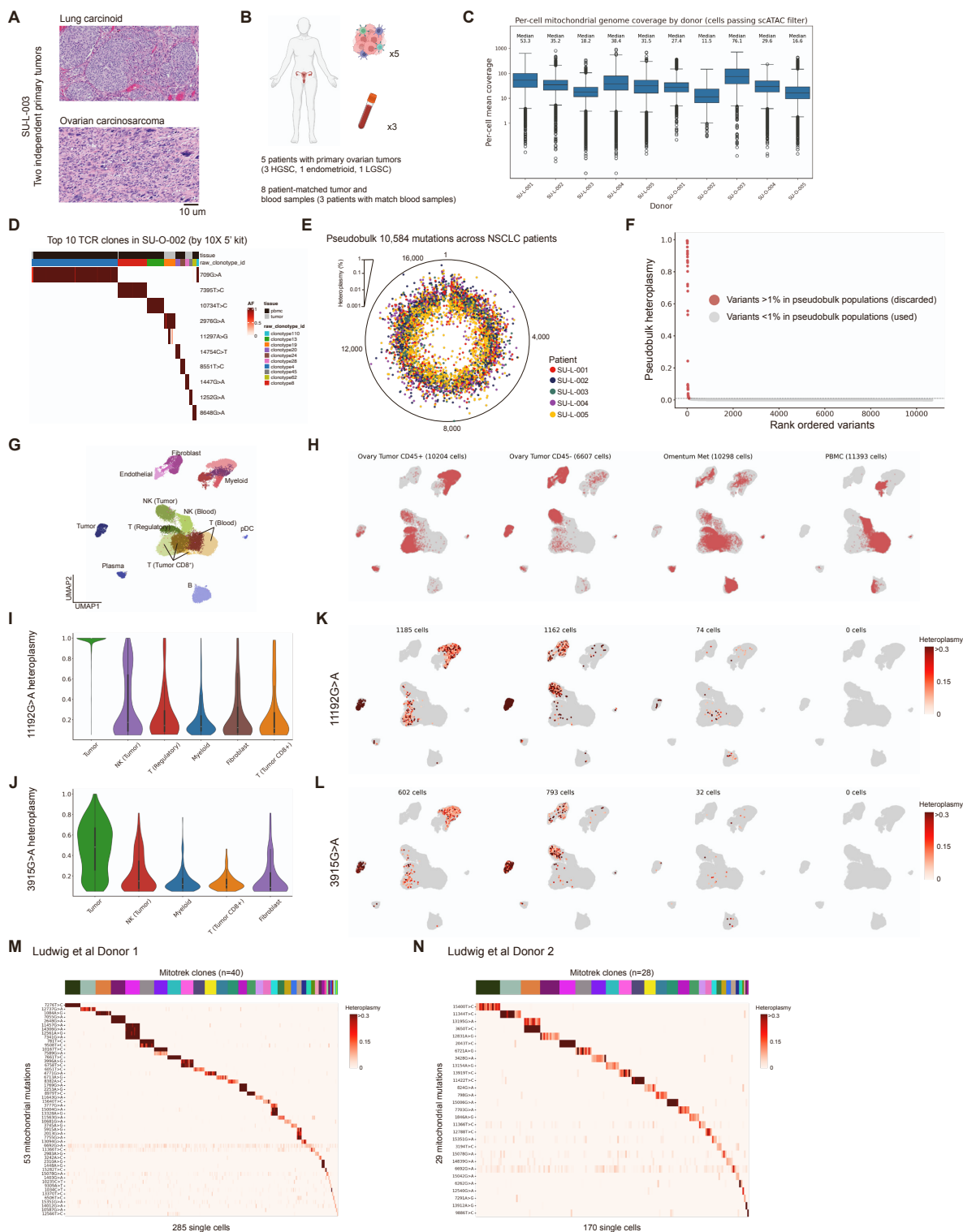


Figure S1: Benchmarking Mitotrek using ground-truth data, Related to Figure 1

(A) Digital pathology scans of patient SU-L-003. Top: scan of low-grade lung carcinoid tumor. Bottom: scan of high-grade ovarian carcinoma.

(B) Schematic summarizing patient and sample information for the ovarian tumor data.

(C) Mitochondrial genome coverage across patients. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range.

(D) Validation of mtDNA clonotypes in T cells by TCR sequencing. Shown in the heatmap are individual cells among the 10 most abundant clonotypes with the heteroplasmy of 10 mutations enriched in a specific TCR (all heteroplasmy is plotted), showing high clonal specificity.

(E) Distribution of mgatk-nominated variants along the mitochondrial genome, averaged across cells and colored by patient.

(F) Pseudobulk heteroplasmy for all variants detected by mgatk across all samples. Variants with >1% pseudobulk heteroplasmy are excluded from downstream analysis. 1% is chosen as the conservative threshold after observing the overall heteroplasmy distribution.

(G) UMAP embeddings of tumor-infiltrating immune cells from matched primary (ovarian) and metastatic (omentum) tumors, and PBMCs from HGSC patient SU-O-005.

(H) Distribution on the UMAP of cells from indicated samples. Cells from the primary ovarian tumor were sorted by CD45 to separate tumor-infiltrating immune cells.

(I) Heteroplasmy levels of 11192G>A, a tumor-specific mitochondrial variants in tumor-infiltrating immune cells processed together with tumor cells.

(J) Same as (I) but for 3915G>A.

(K and L) Mutations projected onto UMAP embeddings across samples. Tumor-specific variants are indiscriminately detected at lower heteroplasmy levels in all cells from the same sample, suggesting mitochondrial transfer and/or technical artifacts (ambient mtDNA).

(M and N) Heatmap showing the heteroplasmy levels of variants (rows) that are identified as clone markers to group cells (columns) in each Mitotrek clone. Position of each variant and the base pair change are shown.

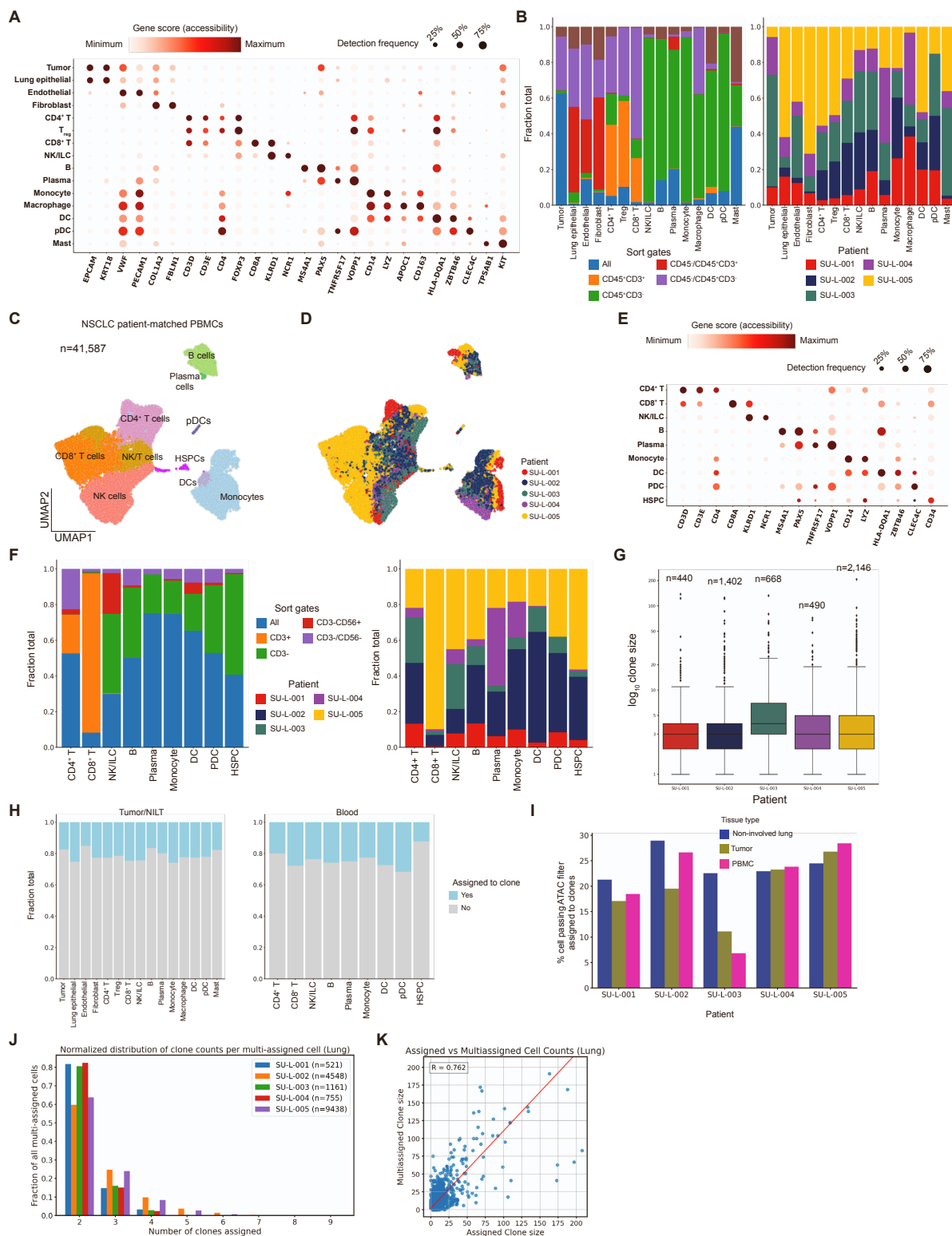


Figure S2: NSCLC data summary and cell type annotation, Related to Figure 1

(A) For cell types annotated in tumor and NILT samples, column-scaled gene accessibility scores and detection frequencies for the indicated genes.

(B) For tumor and NILT samples, bar plots indicating (left) relative proportions of markers used for sorting that were detected in each cell type (certain samples were not sorted, and select sorted samples were merged during single-cell capture, due to sample-specific considerations to optimize single-cell yield) and (right) relative proportions of cells from each patient detected in each cell type.

(C) UMAP of 41,587 PBMCs from patients with lung tumors.

(D) UMAP of cells colored by patient identity.

(E) For cell types annotated in PBMC samples, column-scaled gene accessibility scores and detection frequencies for the indicated genes.

(F) For PBMC samples, bar plots indicating (left) relative proportions of markers used for sorting that were detected in each cell type (certain samples were not sorted, and select samples were additionally sorted using CD56 to enrich innate lymphocytes and myeloid cells) and (right) relative proportions of cells from each patient detected in each cell type.

(G) Clone size distributions for patients with NSCLC. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range.

(H) Bar plots summarizing relative proportions of cells assigned to clones across cell types in tumor/NILT and blood. No significant cell-type bias was observed.

(I) Fraction of cells passing ATAC filters that are successfully assigned to clones.

(J) Characterization of multi-assigned clones in lung tumors. Bar plot showing the distribution of the number of clones assigned for multi-assigned cells.

(K) Clone size comparison between excluded multi-assigned cells and uniquely assigned cells used for downstream analysis. Statistic and P value from Spearman correlation.

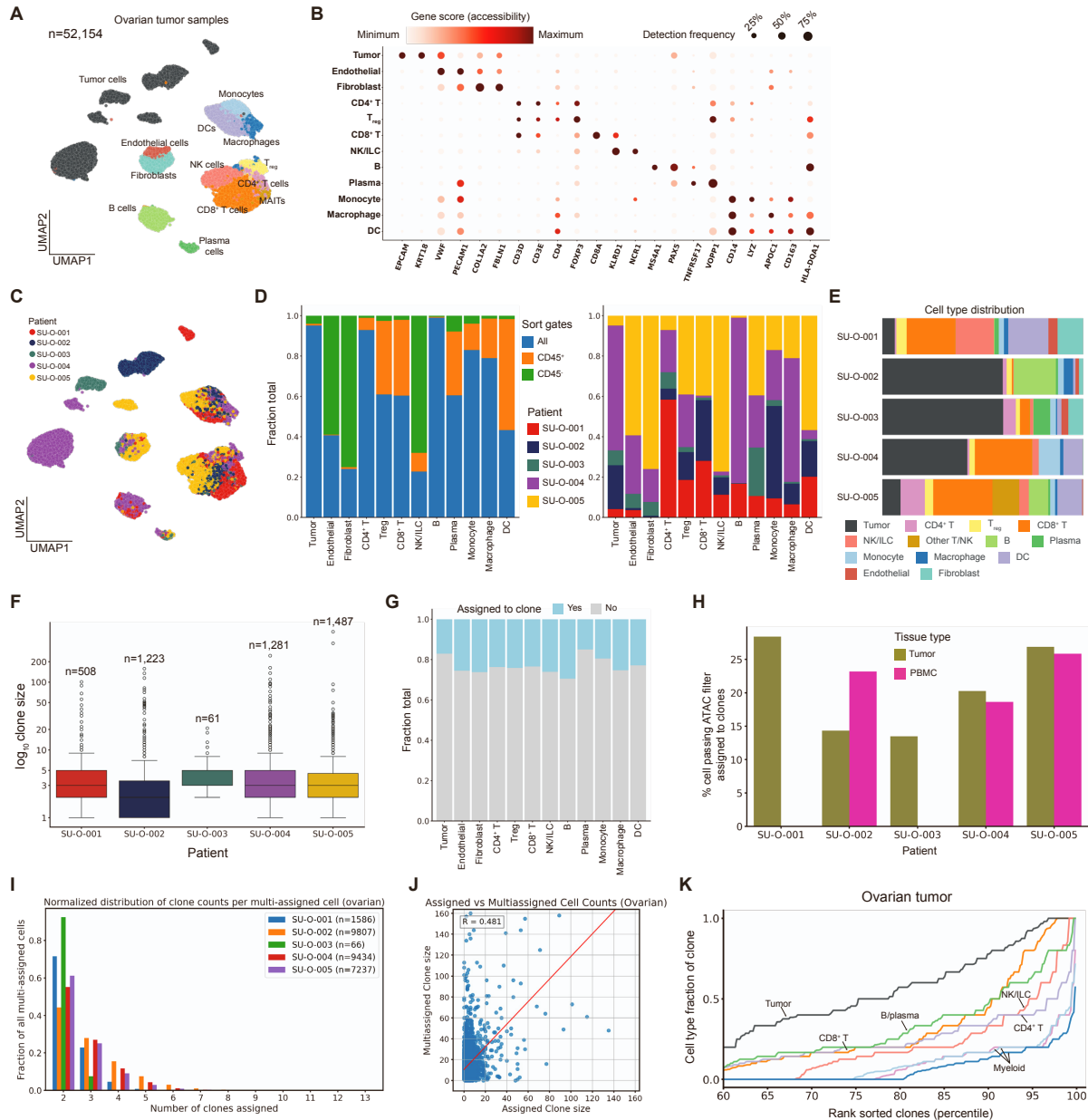


Figure S3: Ovarian cancer data summary and cell type annotation, Related to Figure 2

(A) UMAP of 52,154 cells in ovarian tumors. Cell types denoted by color are inferred after iterative sub-clustering of each of the myeloid, lymphoid, and stromal compartments.

(B) For cell types annotated in ovarian tumor samples, column-scaled gene accessibility scores and detection frequencies for the indicated genes.

(C) UMAP of cells colored by patient identity.

(D) For ovarian tumor samples, bar plots indicating (left) relative proportions of markers used for sorting that were detected in each cell type (certain samples were not sorted, and select sorted samples were merged during single-cell capture, due to sample-specific considerations to optimize single-cell yield) and (right) relative proportions of cells from each patient detected in each cell type.

(E) Normalized bar plot showing cell type composition for each patient.

(F) Clone size distributions for patients with ovarian cancer. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range.

(G) Bar plots summarizing relative proportions of cells assigned to clones across cell types in ovarian and blood. No significant cell-type bias was observed.

(H) Fraction of cells passing ATAC filters that are successfully assigned to clones. Peripheral blood samples were obtained from three patients.

(I) Characterization of multi-assigned clones in ovarian tumors. Bar plot showing the distribution of the number of clones assigned for multi-assigned cells.

(J) Clone size comparison between excluded multi-assigned cells and uniquely assigned cells used for downstream analysis. Correlation test: Spearman.

(K) Cumulative fractions of clones stratified by cell type for cells from ovarian tumor samples. Clones with ≥ 5 cells are considered for this analysis.

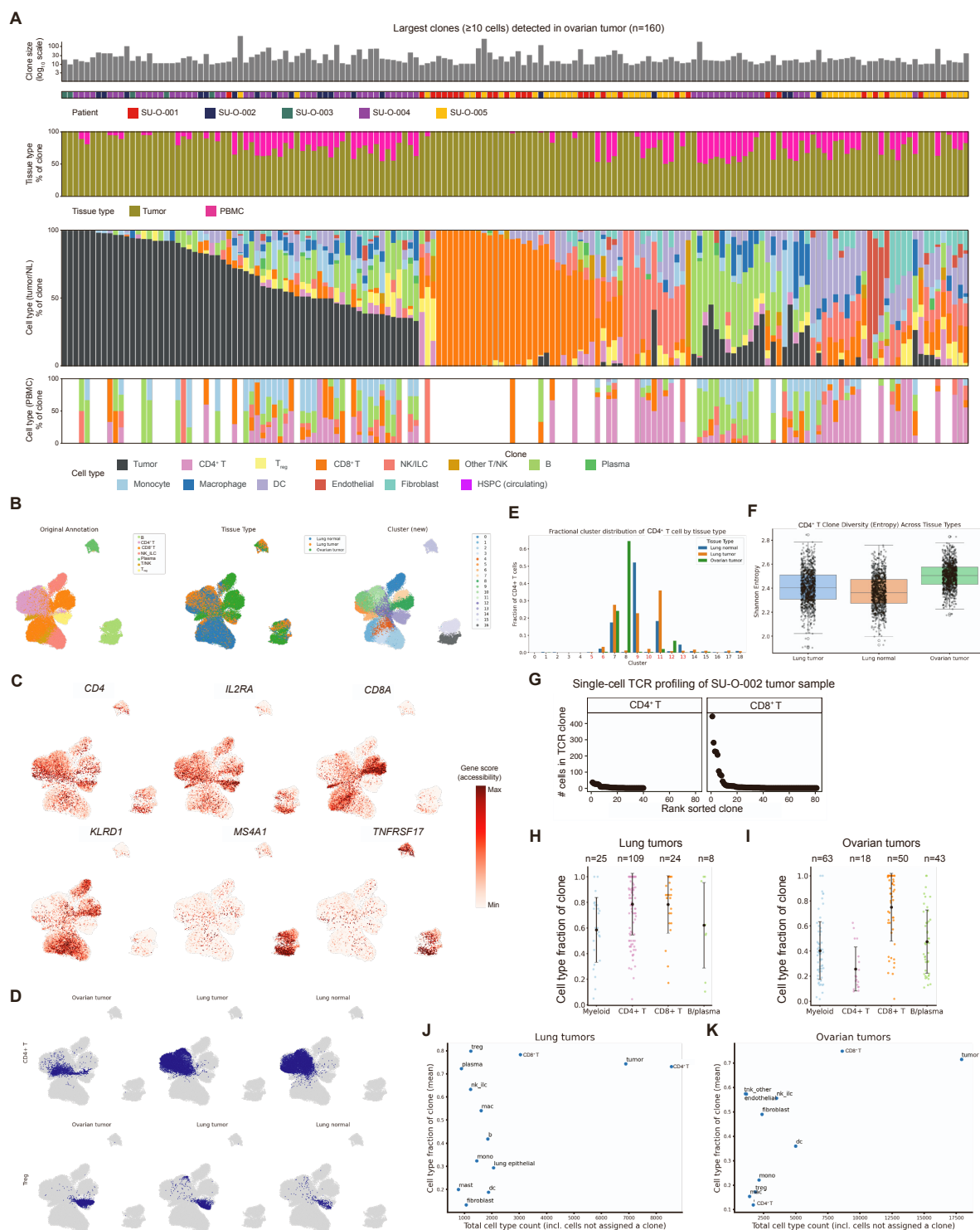


Figure S4: Additional clonal landscape analyses, Related to Figure 2

(A) All clones with at least 10 cells detected in ovarian tumor samples. Each column represents a unique clone.

(B) UMAP of lymphocytes in all solid tissue samples, including both lung and ovarian tumors.

(C) Accessibility of lymphocyte markers distinguish them on UMAP.

(D) Visualization of CD4⁺ T and Treg stratified by tissue type on UMAP, showing overlap by tissue type and supporting the validity of the original cell type annotation.

(E) Cluster distribution of originally labeled CD4⁺ T cells.

(F) CD4⁺ T-cell clonal diversity across tissue types. Shannon entropy was computed from clone frequencies after restricting the analysis to the seven CD4⁺ T clusters. For each tissue type, 500 cells (including both assigned and unassigned clones) were repeatedly sampled with replacement (1,000 iterations) to normalize for cell number and quantify clonal diversity. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range. No statistical test is performed.

(G) Distribution of scTCR-seq derived clone sizes for CD4⁺ and CD8⁺ T cells from SU-O-002. Clones were defined using CellRanger vdj and plotted for clonotypes with at least 2 cells observed.

(H) Annotation of lung tumor clones. Clones with ≥5 cells of the indicated cell type are considered, and the distribution of the indicated cell type's clone fraction is plotted. For myeloid cells, monocytes, macrophages, and DCs are grouped. For CD4⁺ T cells, T_{reg}s and other CD4⁺ T cells are grouped. n represents the number of clones considered for each indicated cell type.

(I) Same as (H) but for ovarian clones.

(J) Relationship between cell type abundance and mean clone fraction in lung tumors. Mean clone fraction and total number of cells are plotted for each cell type, showing no correlation.

(K) Same as (J) but for ovarian tumors.

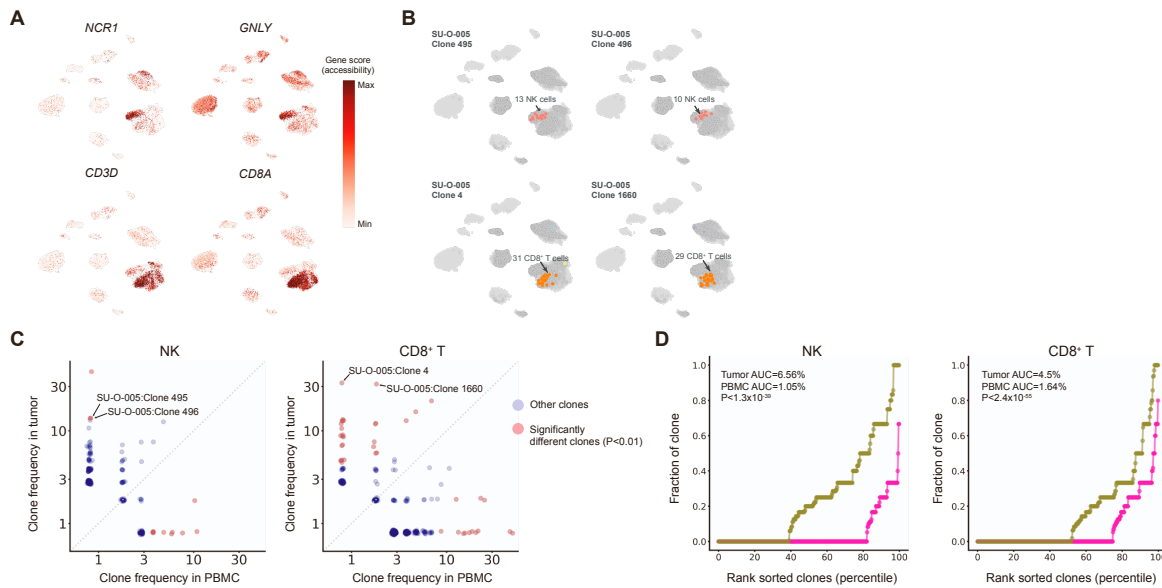


Figure S5: A case report of tumor-specific clonal expansion, Related to Figure 2

(A) Accessibility of NK and CD8⁺ T markers distinguish them on UMAP.

(B) Representative clones capturing clonal expansion events of NK cells (top row) and CD8⁺ T cells (bottom row) in SU-O-005. For each clone, cells from the clone's donor are highlighted with shaded circles, and cells assigned to that clone are colored by their cell type.

(C) Scatterplots comparing clone frequencies of circulating cells with those infiltrating the tumor. The largest clones are locally expanded and minimally detected in the periphery. Significance is determined by Benjamini-Hochberg adjusted Fisher's exact test.

(D) Cumulative fraction of clone sizes for the indicated lymphoid cell types, split by tissue site. AUC corresponds to the overall clone size for the indicated cell type and tissue site. P values are reported from a Wald statistic from binomial GLM.

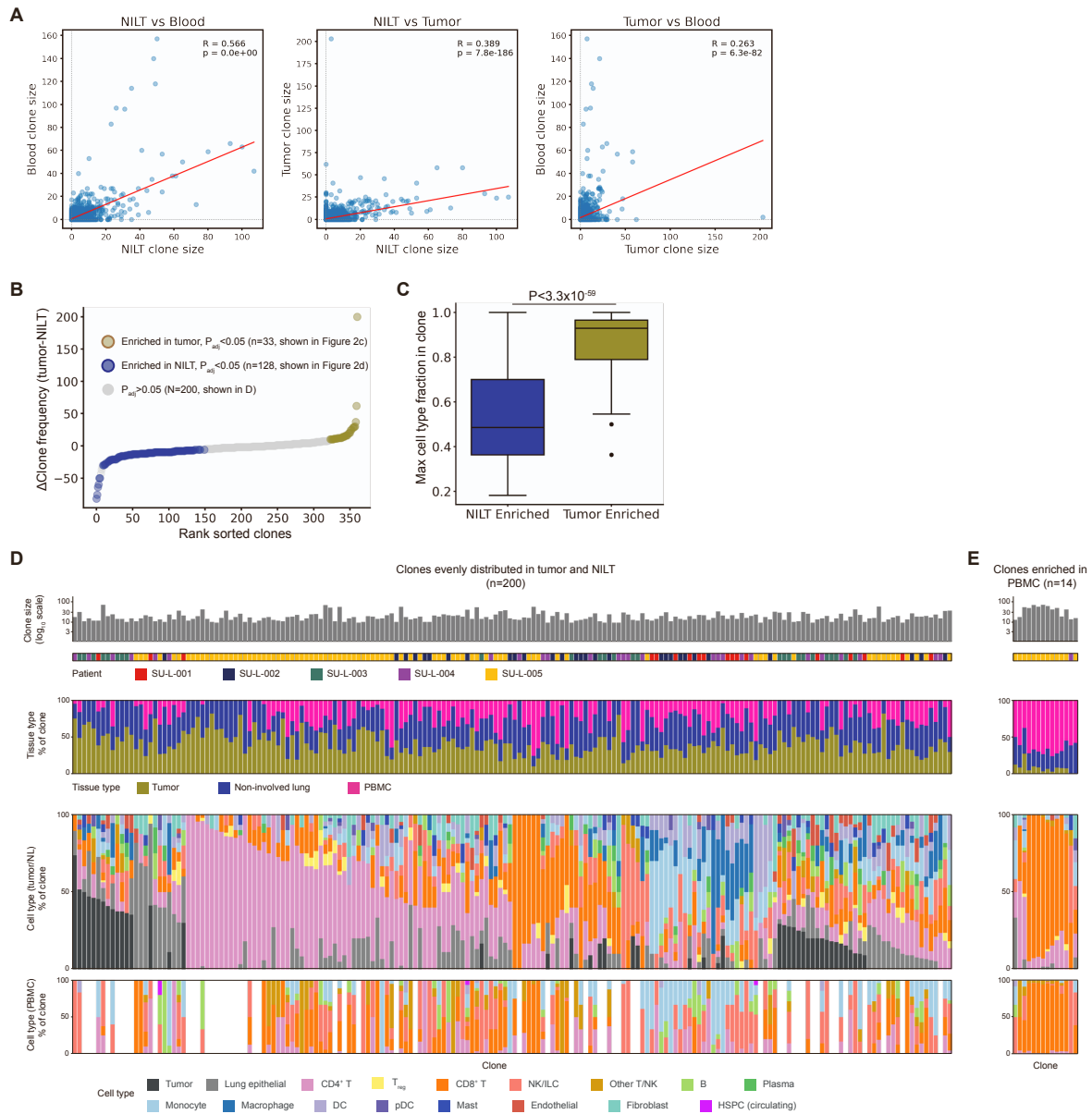


Figure S6: Comparative clonal analysis across tissue sites in NSCLC, Related to Figure 2

(A) Summary of clone Spearman correlations between anatomical sites. Clone sizes are significantly correlated between tissue sites.

(B) Enrichment of clones in NILT or tumor. Significance is determined by Benjamini-Hochberg adjusted Fisher's exact test against overall tissue site distribution for clones with at least 10 cells in tumor and NILT.

208 (C) Comparison of dominant cell type fraction distribution between clones enriched in
 209 NILT and in tumor. Boxplots: center line, median; box limits, first and third quartiles;
 210 whiskers, 1.5× interquartile range. P value from binomial GLM.
 211
 212 (D) Clones enriched in neither tumor nor NILT, as determined by p value > 0.05 from
 213 Benjamini-Hochberg adjusted Fisher's exact test against overall tissue site distribution for
 214 clones with at least 10 cells in tumor and NILT. Each column represents a unique clone.
 215
 216 (E) Clones enriched in PBMC from the same analysis shown in (D).

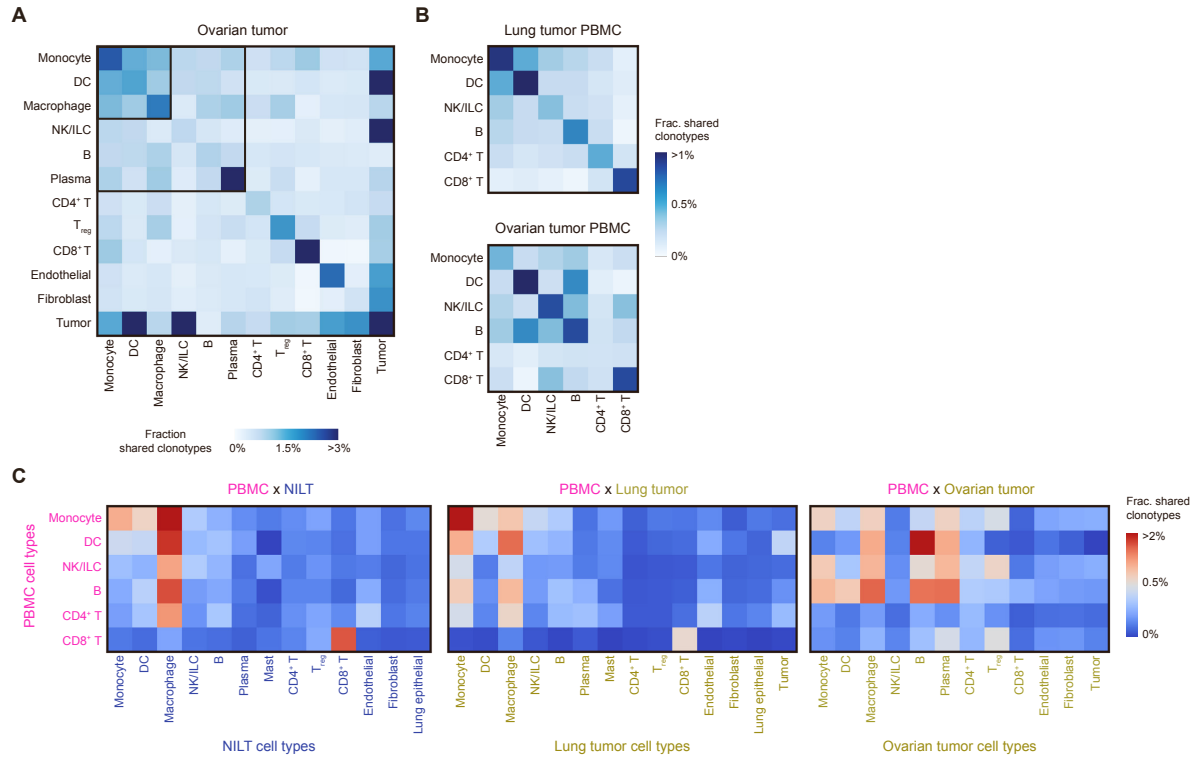


Figure S7: Additional cross-tissue clonal analysis, Related to Figure 2

(A) Heatmaps showing the fraction of all cell pairs belonging to the same clone and consisting of two cell types within ovarian tumor. Pairs were restricted to cells from the same donor.

(B) Same as in (A) but for PBMCs.

(C) Heatmap showing the fraction of all cell pairs belonging to the same clone and consisting of a PBMC cell type and a solid tissue cell type.

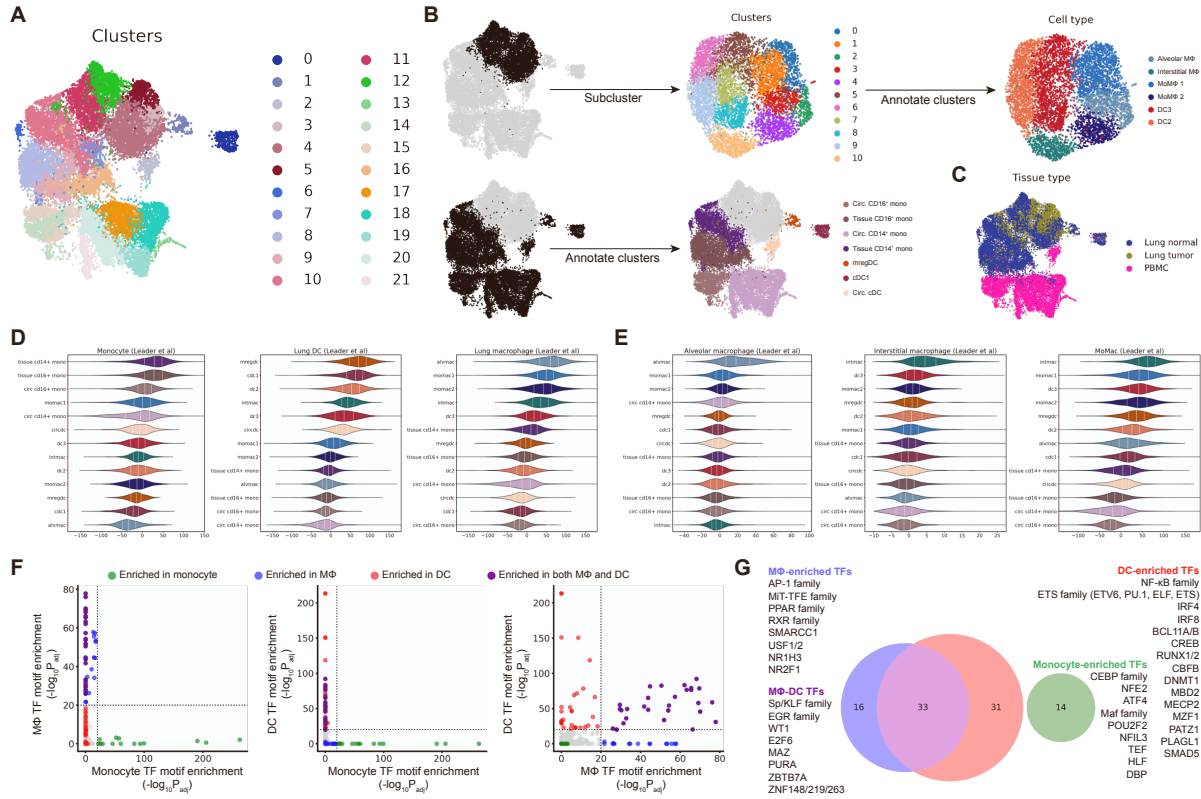


Figure S8: NSCLC myeloid annotation and epigenetic analysis, Related to Figure 3

(A) UMAP of myeloid cells from PBMC, lung tumor, and NILT samples of patients with NSCLC, colored by their original cluster assignments.

(B) Myeloid annotation scheme. Monocyte clusters separated clearly into CD14⁺ and CD16⁺ subsets, which were annotated without further subclustering. Macrophages and DCs were subclustered to achieve higher granularity and annotated based on subclustering results.

(C) UMAP of cells colored by tissue sites.

(D) Gene scores of published signatures derived from single-cell RNA-seq data.

(E) Same as (D) but for proteomic data.

(F) TF motifs enriched in the marker peaks of three major MNP cell types. P-values are calculated from the Benjamini-Hochberg adjusted Wilcoxon signed-rank test.

(G) Summary of TFs whose motifs are enriched in broad MNP cell types. A $-\log_{10}P_{adj} > 20$ cutoff was used.

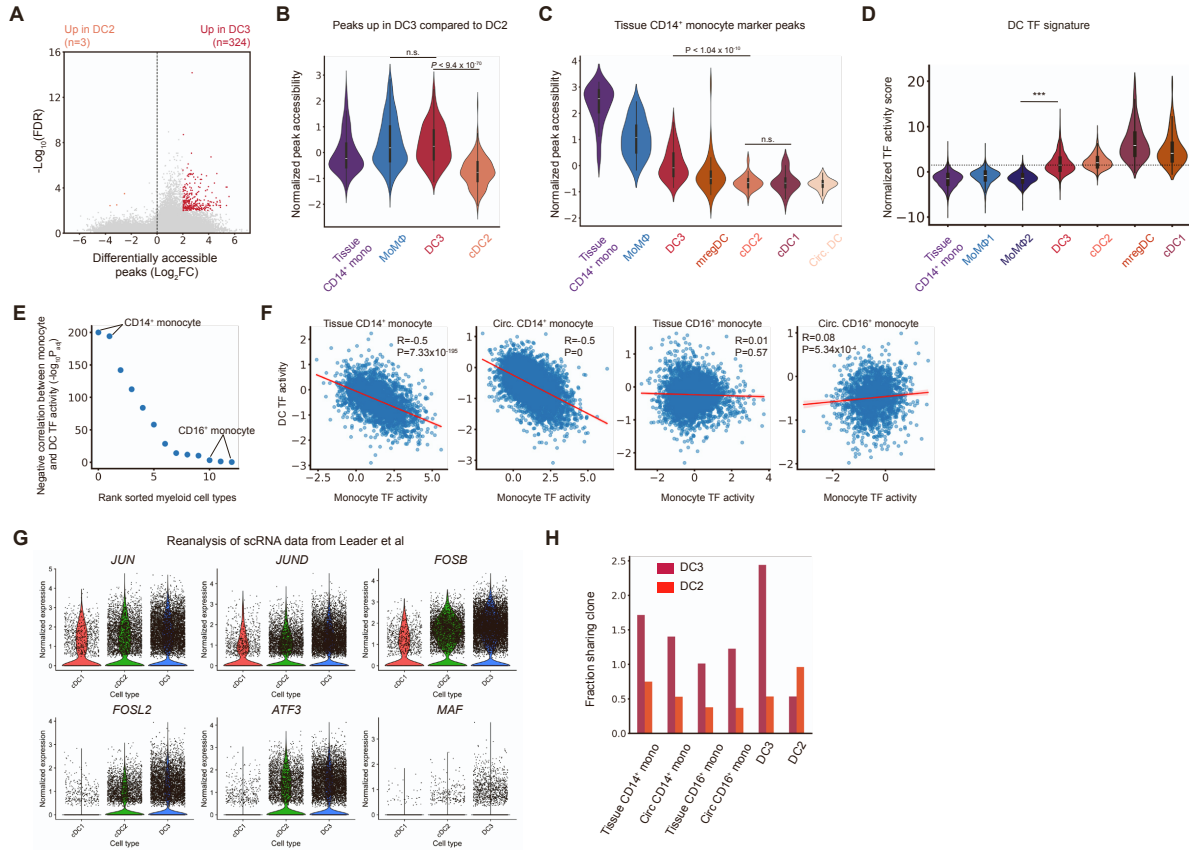


Figure S9: DC3 epigenetic and clonal analysis, Related to Figure 3

(A) Differentially accessible genomic regions in DC3 vs DC2.

(B) Normalized sum accessibility of genomic regions significantly more accessible in DC3 compared to DC2 in indicated cell types. P-values are from a Mann-Whitney U test.

(C) Normalized sum accessibility of genomic regions significantly more accessible in CD14⁺ monocytes compared to other myeloid subtypes. P-values are from a Mann-Whitney U test.

(D) Average chromVAR motif deviation scores for DC TFs highlighted in **Figure 3C**. Kruskal-Wallis test.

(E) Statistical significance of monocyte-DC TF motif accessibility correlation in all myeloid subtypes. CD14⁺ monocytes in tissue and circulation display the most significant negative correlation between monocyte and DC TF motif accessibilities.

(F) Monocyte and DC TF activities are negatively correlated only in CD14⁺ monocytes but not in CD16⁺ monocytes. Correlation test: Spearman.

(G) RNA expression of monocyte-associated TFs in DC subtypes. scRNA-seq data and annotation are procured from a prior report.

274

275 (H) Monocytes share more clonotypes with DC3 than with DC2.



(B) Column-scaled gene accessibility scores and detection frequencies for the indicated genes.

(C) Differentially active TF motifs between DC2 and DC3. P-values are calculated using the Benjamini-Hochberg adjusted Kruskal-Wallis test.

(D) Same as (C) but for DC2 and CD14⁺ monocytes

(E) Same as (D) but colored by TFs up in DC3 vs DC2 shown in (C).

(F) Average chromVAR motif deviation scores for the indicated monocyte-associated TFs.

(G) Same as (F) but for macrophage-associated TFs.

(H) Same as (F) but for DC-associated TFs.

297 (I) Cell type-cell type clone frequency correlation across clones (≥ 5 cells across all
298 samples). Color denotes correlation value, computed using Pearson's ρ . Text labels of
299 circulating PBMC cell types are colored pink.

300
301 (J) Scatterplots comparing clone frequencies of circulating CD14⁺ monocyte and tissue
302 CD14⁺ monocyte (left), DC3 (middle), and tumor CD8⁺ T cell (right). Significantly different
303 clones with $P < 0.05$ adjusted Fisher's exact test are highlighted red. Correlation test:
304 Spearman.

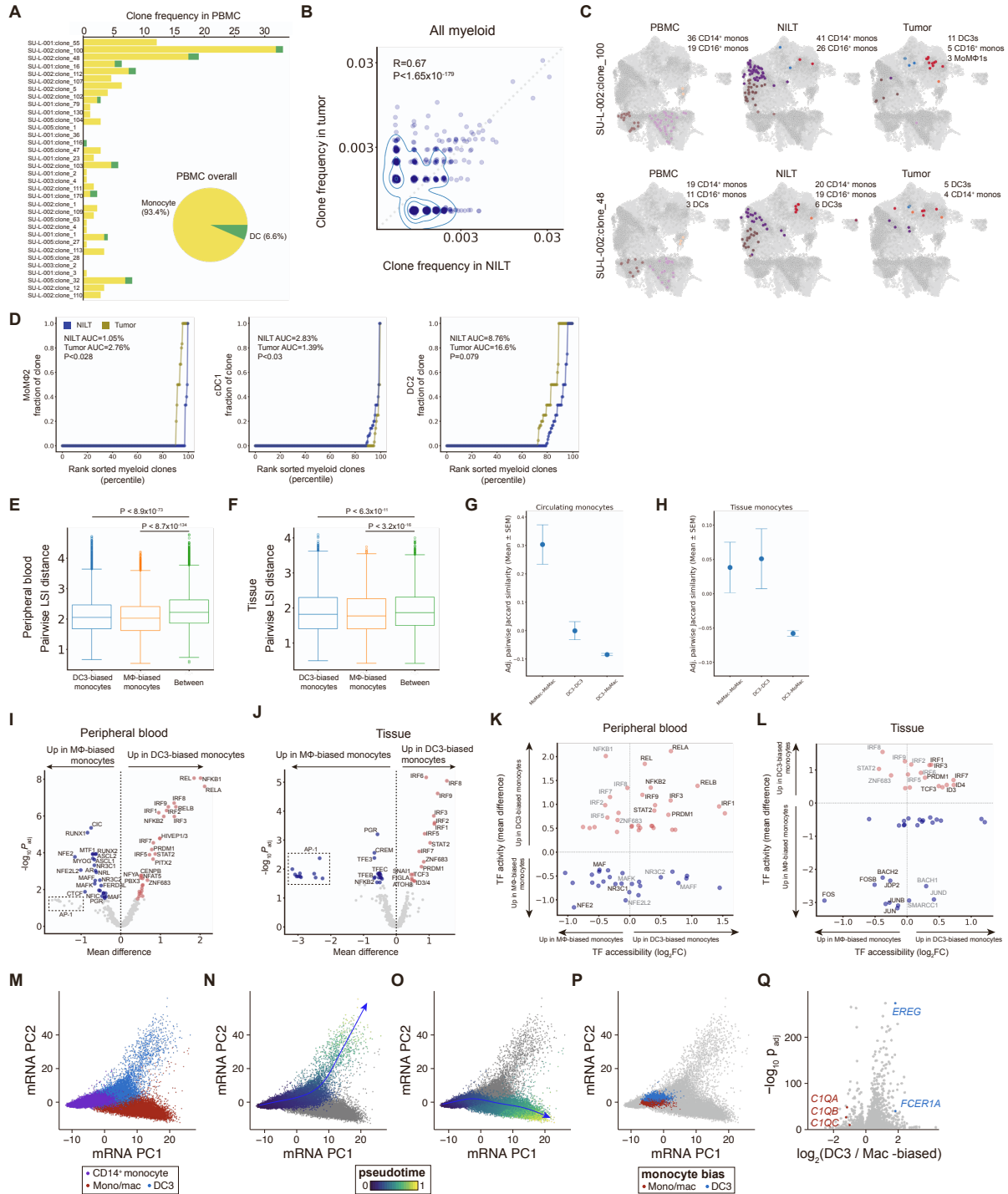


Figure S11: Epigenetic comparison of DC- and macrophage-biased myeloid clones, Related to Figure 4

(A) Monocyte and DC proportions of largest myeloid clones (≥ 10 cells) in PBMC samples of patients with NSCLC.

(B) Scatterplots comparing clone frequencies of myeloid cells in tumors with those in NILT. Contours visualize density. Correlation Test: Spearman.

(C) Representative clones capturing myeloid cell type distribution across tissue sites. For each clone, cells from the clone's donor are highlighted with shaded circles, and cells assigned to that clone are colored by their cell type.

(D) Cumulative fraction of clone sizes for the indicated myeloid cell types, split by tissue site. AUC corresponds to the overall clone size for the indicated cell type and tissue site. P-value from binomial GLM.

(E) Epigenetic similarity as measured by distance in the LSI space within and between DC3-biased and macrophage-biased monocytes in circulation. Statistic Kruskal-Wallis test.

(F) Same as (E) but for tissue-resident cells.

(G) Normalized genome-wide similarity among monocyte fate-biased subpopulations of monocytes in circulation. Genome-wide chromatin accessibility similarity was computed using the Jaccard index after binarizing 500-bp cell-by-bin matrices and normalizing raw values to correct for sequencing-depth-dependent inflation, following Fang et al.. Shown are mean \pm SEM adjusted pairwise similarities for macrophage-biased, DC-biased, and cross-group monocyte pairs in circulation.

(H) Same as (G) but for tissue-derived monocytes.

(I) Differentially active TF motifs between DC3-biased and macrophage-biased monocytes in peripheral blood. P-values are calculated using the Benjamini-Hochberg adjusted Kruskal-Wallis test.

(J) Same as (I) but for tissue-resident monocytes.

(K) Joint comparison of TF gene body accessibility and inferred genome-wide TF activity to nominate TFs driving differential fate outcomes of monocytes in circulation

(L) Same as (K) but for tissue-resident monocytes.

(M) Reduced dimensionality of 45,524 cells from prior lung profiles colored by cell cluster.

(N) Pseudotime trajectory of DC3 differentiation from scRNA-seq profiles.

(O) Same as (N) but for monocyte-derived macrophages.

(P) Fate-biased monocytes used for differential expression.

357 (Q) Differential expression analyses of monocytes biased toward macrophages or DC3s.
358 Selected genes supporting cell state are highlighted.