nature methods

ARTICLES

https://doi.org/10.1038/s41592-020-0880-2

ZipSeq: barcoding for real-time mapping of single cell transcriptomes

Kenneth H. Hu[1], John P. Eichorst[1,2], Chris S. McGinnis[3], David M. Patterson[3], Eric D. Chow[4], Kelly Kersten[1], Stephen C. Jameson[5], Zev J. Gartner[3], Arjun A. Rao[1,2] and Matthew F. Krummel[1,2]✉

Spatial transcriptomics seeks to integrate single cell transcriptomic data within the three-dimensional space of multicellular biology. Current methods to correlate a cell's position with its transcriptome in living tissues have various limitations. We developed an approach, called 'ZipSeq', that uses patterned illumination and photocaged oligonucleotides to serially print barcodes ('zipcodes') onto live cells in intact tissues, in real time and with an on-the-fly selection of patterns. Using ZipSeq, we mapped gene expression in three settings: in vitro wound healing, live lymph node sections and a live tumor microenvironment. In all cases, we discovered new gene expression patterns associated with histological structures. In the tumor microenvironment, this demonstrated a trajectory of myeloid and T cell differentiation from the periphery inward. A combinatorial variation of ZipSeq efficiently scales in the number of regions defined, providing a pathway for complete mapping of live tissues, subsequent to real-time imaging or perturbation.

Single cell RNA sequencing (scRNA-seq) combined with other multimodal analyses such as surface epitope labeling and repertoire analysis have revealed transcriptional heterogeneity within cell populations. This approach has been especially useful in immunology given the diversity of immune cell types and the microenvironments they experience[1]. Yet scRNA-seq studies lose information on the spatial context where a given single cell transcriptome was localized[2]. Conventional microscopy localizes cells and molecules in space, but is limited in channels. Even with multiplexed imaging techniques (MIBI, CODEX, MERFISH and SeqFish), probes must be selected a priori[3–6]. To better understand how cellular transcriptional heterogeneity is influenced by the local environment and vice versa in a discovery-based, unbiased approach, it becomes necessary to link high-dimensional scRNA-seq data to the spatial dimensions and real-time phenotypical analyses that microscopy affords.

To couple conventional microscopy of live tissues with single cell transcriptomics, we needed a means to demarcate multiple regions of interest (ROI) in real time, in both mouse and human tissues. While excellent for some applications, we decided against grid-based approaches[7–9], which average together gene expression within a region and are defined before imaging on fixed tissue sections. Instead, we sought to develop a method for 'printing' a DNA barcode onto live cells in a spatially defined manner, which can be read out during typical scRNA-seq workflows[10]. We accomplished this by initially coating a photocaged base DNA oligonucleotide onto cells in a tissue/dish and through illumination, control hybridization of subsequent barcoded DNA strands in a light- and thus spatially restricted manner[11,12].

Results

We generated a photo-uncaging system that allowed light-based printing of DNA barcodes onto the surface of cells. A double-stranded piece of DNA was attached to cells, either by a high-affinity antibody (for example, anti-CD45, for pan-immune cells) or via stable lipid insertion using a lipid-modified oligonucleotide (LMO) into the membrane[13]. This double-stranded 'anchor strand' contained a 17 bp overhang sequence (termed 'O1') that is blocked at four sites along its length using 6-nitropiperonyloxylmethyl (NPOM) conjugated to thymidine, and thus is unable to participate in base-pairing[11,14]. Following local illumination with 365 nm light to release the cages, a readout oligonucleotide strand, termed a zipcode or 'ZC' strand can hybridize to O1 (Fig. 1a). This annealed zipcode terminates in a polyA sequence and an Illumina Read 2 Sequence, which then allows for poly-dT based amplification during library construction[10]. We first demonstrated light-based control of hybridization for two zipcodes, by marking two separated populations of primary mouse CD4 and CD8 T cells and demonstrating excellent concordance of populations and the associated zipcodes (Supplementary Fig. 1a–f). Cells with either an even mix of zipcodes or 'incorrectly' assigned could partially be explained by doublets detected computationally[15] (Supplementary Fig. 1f) combined with <100% cell purity following purification from lymphoid tissues. We also noted minimal zipcode mixing when cells were pooled (Supplementary Fig. 1g,h). Additionally, we show that our workflow of ultraviolet illumination and oligonucleotide addition does not adversely affect cell viability of mouse primary lymphocytes (Supplementary Fig. 1i–k). We used a digital micromirror device (DMD) to control the spatial pattern of 365 nm light in a plane conjugate with the image plane of a conventional widefield microscope (Fig. 1b).

To demonstrate spatial printing of a collection of zipcodes, we plated CD8 mouse T cells, labeled with caged anti-CD45 Ab-DNA anchor strands and used complementary O1′ strands that were labeled with one of three fluorophores to visually track spatially controlled annealing. Following three rounds of (1) patterned illumination, (2) zipcode addition and (3) washing, we obtained clear delineation of three regions, showing the linear scaling of resolution

[1]Department of Pathology, University of California San Francisco, San Francisco, CA, USA. [2]ImmunoX Initiative, University of California San Francisco, San Francisco, CA, USA. [3]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA. [4]Department of Biochemistry and Biophysics, Center for Advanced Technology, University of California San Francisco, San Francisco, CA, USA. [5]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA. ✉e-mail: matthew.krummel@ucsf.edu

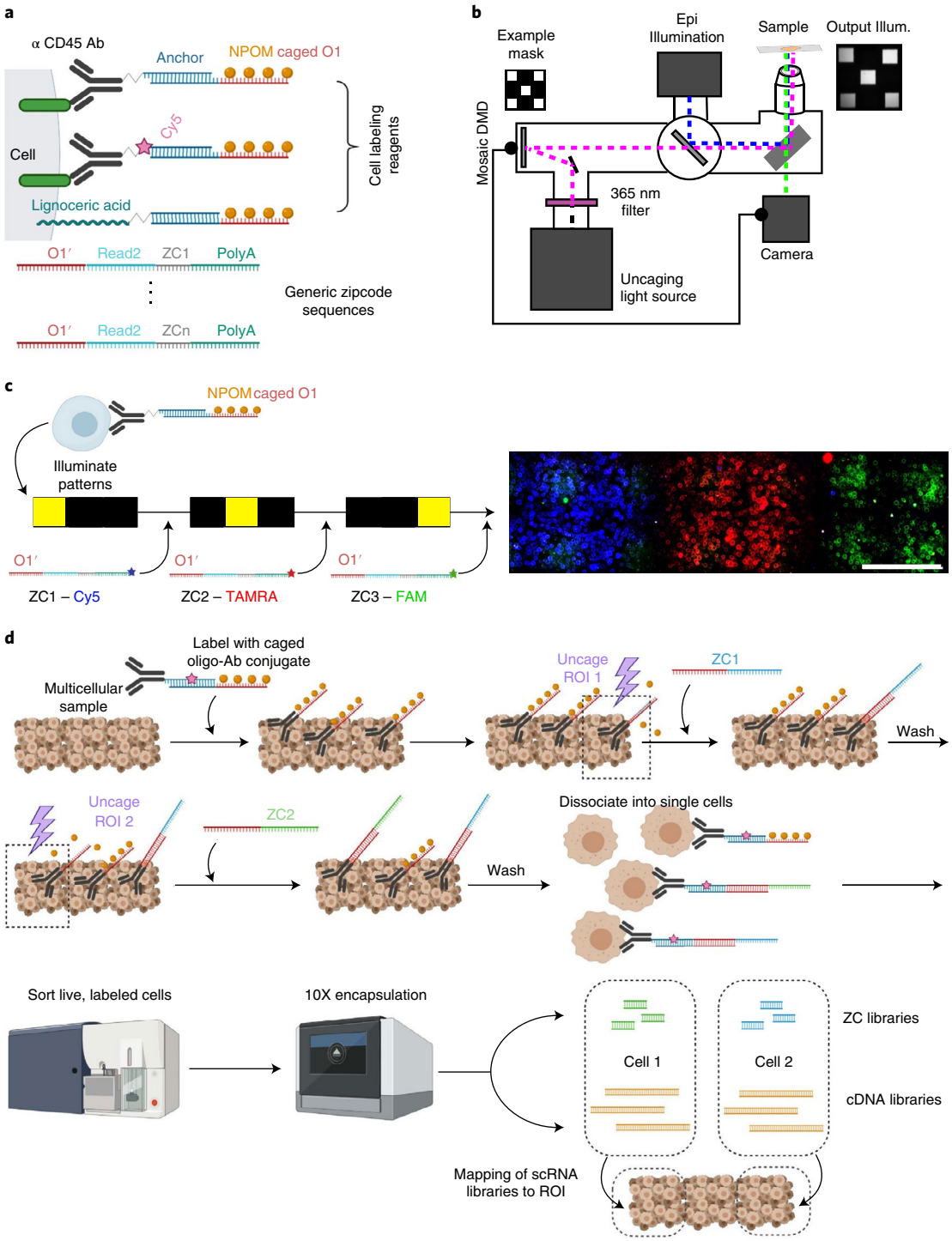


Fig. 1 | Design of ZipSeq oligonucleotides, imaging setup and workflow. a, Schematic of oligonucleotide sequences and labeling moieties used in this paper. Both lipid and antibody are covalently conjugated to an 'anchor' sequence. Meanwhile a caged strand consisting of four photocaging groups on an overhang sequence (O1) linked to the reverse complement of the anchor strand can hybridize with the Ab or lipid DNA conjugate before labeling cells. Readout strand or zipcodes consist of a reverse complement sequence to the caged overhang sequence 1 or 2 (O1′), followed by a partial Illumina Small RNA Read 2 sequence for downstream amplification. In addition, each zipcode strand bears a 8-bp barcode and a 28-bp polyA sequence for capture by poly-dT primers during reverse transcription. **b**, A microscope light path for simultaneous imaging and photo-uncaging of a sample. An example mask generated by the DMD is shown with the resulting illumination pattern visualized on a mirrored slide. **c**, Illustration of proof-of-concept demonstrating ability to spatially control hybridization of fluorescently labeled oligonucleotides. Scale bar, 200μm. **d**, Schematic for workflow for labeling two ROI in a tissue section beginning with labeling of cells in a dish or in tissue with appropriate labeling moiety hybridized to a strand bearing the photocaged ssDNA overhang.

and number of rounds (Fig. 1c). Taking these two proofs of concept together, a schematic of an idealized workflow is depicted in Fig. 1d. A fluorophore (for example, Cy5) can be incorporated into the anchor sequence allowing for enrichment of all Ab-oligonucleotide labeled cells (ZC_{1-n} or ZC unlabeled, which are sorted computationally) for scRNA-seq.

Defining spatially segregated motility and cell division programs in wound healing. We applied ZipSeq to study spatially defined transcriptional programs in a model of wound healing in a monolayer of NIH/3T3 fibroblasts. Twelve hours after 'wounding', we imaged the wound edge (Fig. 2a). We then used the LMO hybridized to the photocaged oligonucleotide (Fig. 1a) to first label all cells and then illuminated a band 0–200 µm from the wound edge ('front') and added zipcode 1. Following hybridization and wash-out of zipcode 1, we illuminated another band 200–400 µm away from the wound edge ('rear') followed by zipcode 2 addition (Fig. 2b). We then dissociated the monolayer into single cells and subjected the cells to a 10× Genomics scRNA-seq pipeline.

During analysis, we removed the cells with low zipcode counts (that is, from neither of the illuminated zones) from the analysis and used the ratio of zipcodes 1 to 2 counts to determine whether a cell derived from the front region versus the rear. We identified three cell clusters in transcriptional space using unsupervised nearest neighbor clustering (Fig. 2c). When we overlaid the ZipSeq region calls, a clear partitioning between front and rear cells was observed (Fig. 2d), with cluster 2 highly enriched for front cells and clusters 0 and 1 relatively enriched for rear cells (Fig. 2e).

Differential expression analysis between front and rear cells (Fig. 2f) identified collections of genes associated with cell motility (for example, *Tagln*, *Acta2* and *Cav1*) enriched in front cells and genes associated with cell division (for example, *H2afv*, *Cenpa* and *Stmn1)* for rear cells as seen in expression overlays onto the uniform manifold approximation and projection (UMAP) representation (Fig. 2g). Gene ontology analysis of biological processes using significantly differentially expressed genes supported the broad segregation of cell division (rear) and motility (front) associated genes in these two regions (Fig. 2h). Using gene signatures for S and G2M phases of the cell cycle[16], we observed that the rear population exhibited significantly higher signature scores for S and G2M phases (Fig. 2i,j). Cluster 1 was especially enriched for cells in S or G2M phase relative to cluster 0 and 2, which were largely in the G0/G1 phase (Fig. 2k). Finally, we validated several gene hits using immunofluorescence. Staining for ACTA2 was broadly enriched near the front (Fig. 2l), tapering after peaking around 100 µm from the wound edge (Fig. 2n), whereas staining for STMN1 demonstrated enrichment (Fig. 2m) approximately 300 µm from the front (Fig. 2n), consistent with the regions we defined earlier (Fig. 2b).

Mapping cortex versus medulla in a live lymph node (LN). We next applied ZipSeq to learn examine gene expression in mouse LNs, which have well-characterized cellular organization. We targeted two regions for ZipSeq: the 'outer' cortex extending from the tissue edge to the T-B margin and an 'inner' region largely comprising the deep T cell zone and the medulla. We first stained live LN sections from an adult C57Bl/6 mouse with fluorescent anti-CD3ε and anti-B220 antibodies to delineate these regions by widefield microscopy (Fig. 3a,b).

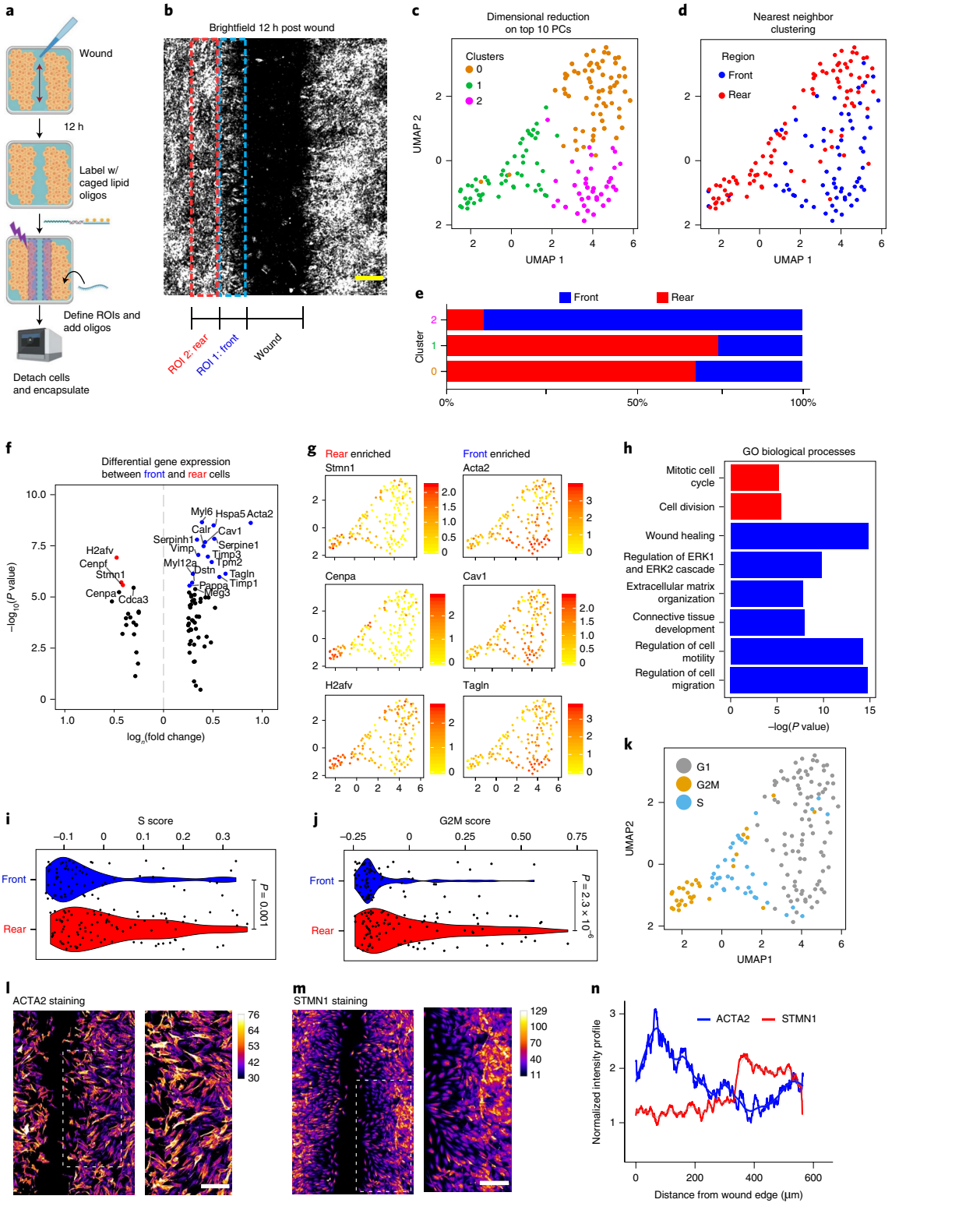
We labeled immune cells within the section with an anti-CD45-based photocaged anchor strand that was also conjugated to Cy5 to allow for purification of immune cells before encapsulation. Using the B220:CD3ε signal as a guide, we first printed zipcode 1 to the outer region followed by printing of zipcode 2 to the inner region. The LN section was then dissociated and live, labeled $CD45^+$ cells were sorted out based on $Cy5^+$ signal and encapsulated. Following merging of zipcode and complementary DNA counts, we separated populations based on dominance of ZC1 and ZC2 (Supplementary Fig. 2). We hypothesize that ambiguous cells represent cells at the border of regions that may have received partial uncaging due to small deviations in stage alignment or bleeding of illumination at edges of the DMD. In parallel, we performed *t*-distributed stochastic neighbor embedding (*t*-SNE) dimensional reduction and identified the main immune cell populations using marker genes. We then overlaid zipcode identity onto the *t*-SNE projection. This revealed clear enrichment of outer cells in the B cell cluster and inner cells for T cell populations (Fig. 3c,d), consistent with expectations and fluorescence imaging.

Using the zipcode spatial information, we performed differential expression analysis within populations based on position. Within the CD4 T cell population, we identified *Ccr7*, *Klf2* and *Klf6* preferentially expressed in cells found in the inner region and calcium binding proteins *S100a6/4* and transcription factor *Rora* preferentially expressed in the outer region (Fig. 3e). Performing a similar

Fig. 2 | ZipSeq mapping of a live cell monolayer following wounding. a, Experimental setup. NIH/3T3 cells were plated 48 h before imaging and allowed to reach confluency. A pipette tip was used to cleanly scrape away a band. The wound was imaged after 12 h and ROI defined. Cells were then labeled with a lipid-oligo conjugate and then uncaged in a series of vertical bands alternating with zipcode addition. Collected cells were then passed into the modified 10× Genomics workflow. **b**, Brightfield of wound 12 h post-wounding with ROI overlaid. Two vertical bands of 200 µm width were drawn with increasing distance from the wound edge (0–200 µm) and (200–400 µm), referred to as front and rear, respectively, for illumination and zipcode hybridization. Scale bar, 200 µm. Imaging representative of two experiments. **c**, UMAP representation of zipcode labeled cells with majority zipcode identity overlaid. For following analysis of scRNA-seq data: $n = 160$ cells, $n_{Front} = 67$ and $n_{Rear} = 93$, and mean $n_{UMI} = 17{,}300$ and mean $n_{Gene} = 3{,}550$ (where n_{UMI} is defined as number of detected unique molecular identifiers per cell and n_{Gene} as the number of detected genes per cell); cutoffs were $n_{Gene} > 1{,}000$, percentage mitochondrial < 0.15. **d**, UMAP representation of labeled cells with cluster overlay. Clusters calculated using Seurat's built-in simulated neural network-based clustering algorithm. **e**, Percentage of cells belonging to either front or rear populations in each cluster as defined in **d**. **f**, Volcano plot from differential expression analysis between front and rear cells. Significance was tested using Seurat's built-in nonparametric Wilcoxon rank-sum test. Colored points represent genes with an adjusted *P* value (Bonferroni corrected) <0.05. **g**, Feature plots overlaid on UMAP representation for three selected genes from differential expression (DE) analysis enriched in either front (*Acta2*, *Cav1* and *Tagln*) or rear cells (*Stmn1*, *Cenpa*, *H2afv*). Color scale indicates log-normalized gene read counts. **h**, Hits from DE analysis were passed through gene ontology analysis. Significantly enriched biological processes shown with −log(*P* value) (Bonferroni corrected). GO, gene ontology. **i**,**j**, Violin plots for S-phase (**i**) and G2M-phase signature score (**j**) for front and rear cells. Gene lists were derived from built-in signatures in Seurat. For S score, $n = 160$ cells, $n_{Front} = 67$ and $n_{Rear} = 93$, and a *P* value of 0.001 using a two-tailed Wilcoxon test (minima, maxima, center and 25th/75th percentiles), −0.145, 0.333, −0.080 and −0.110/−0.021 for the front and −0.134, 0.361, −0.015 and −0.088/0.120 for the rear. For G2M score, $n = 160$ cells, $n_{Front} = 67$ and $n_{Rear} = 93$, and a *P* value of 2.3×10^{-6} using a two-tailed Wilcoxon text (minima, maxima, center and 25th/75th percentiles), −0.243, 0.558, −0.178 and −0.198/−0.122 for the front, and −0.242, 0.708, −0.094 and −0.176/0.124 for the rear. **k**, Assignment to cell cycle phase (S, G2M or G1 phase) based on the signature scores calculated in **i** and **j**. **l**,**m**, Immunofluorescence imaging of fixed NIH/3T3 cells 12 h post-wounding stained for either ACTA2 (**l**) or STMN1 (**m**). Images representative of two wells using the same starting cells, growth, fixation, staining and imaging conditions. The fire LUT from ImageJ has been applied. Zoomed-in insets shown for indicated regions. Scale bar, 100 µm. **n**, Line plot with quantification of mean fluorescence intensity versus distance from edge. Immunofluorescence images from **l** and **m** were first masked for pixels belonging to cells versus background. Then in-cell pixels within vertical bands stepping away from the wound were averaged to create the indicated line-scan profiles with a smoothed fit applied. For **c**–**k** $n = 160$ cells, $n_{Front} = 67$ and $n_{Rear} = 93$, and representative of two independent experiments using distinct starting cell aliquots, imaging, barcoding and sequencing runs.

analysis in B cells identified *Klf2* and *Fosb* in B cells found in the inner versus the outer region (Fig. 3f). Given the appearance of *Klf2* as an inner enriched gene in both of these analyses, we validated its spatial expression pattern using immunofluorescent imaging of LNs (Fig. 3g) from a KLF2-GFP reporter mouse[17]. Using B220 and CD4 staining to identify B and CD4 T cells, respectively, we found that there was indeed more KLF2-GFP expression in both B and CD4 T cells in the interior (Fig. 3h). We similarly validated that CD4 T cells found near and in B cell follicles expressed more S100A6 than those found deeper in the T cell zone (Fig. 3i,j). We noted a more modest difference in S100A6 in B cells (Fig. 3i).

Mapping immune cell differentiation in relation to position within tumors. To map variations in immune cell composition and expression state within a live tumor, we derived cell lines from spontaneous tumors arising in the PyMT-chOVA mouse breast cancer



model, in which mCherry and ovalbumin (OVA) were coexpressed under the MMTV promoter, along with the Polyoma middle T antigen (PyMT)[18]. We orthotopically injected these into the mammary fat pad of female C57Bl/6 mice and 10 d later, we adoptively transferred 2 million CD8 T cells isolated from an OTI UBC-GFP mouse. We allowed the T cells to expand for 4 d and traffic to tumors. Tumors were collected and sectioned into ~150 µm-thick slices (Fig. 4a) for imaging. We observed dense clusters of green fluorescent protein (GFP) OTI T cells in the tumor margin with more dispersed cells in the interior (Fig. 4b).

We defined 'margin' versus 'interior' regions based on the GFP and Cherry signals where margin and interior correspond to antigen-specific CD8 T cell 'dense' and 'sparse' regions. We used anti-CD45-Cy5 labeled anchor strands and labeled immune cells with zipcodes 1 and 2 corresponding to margin and interior, respectively (Fig. 4b). After dissociation of the tumors, we sorted out $Cy5^+$ cells and encapsulated them for scRNA-seq using our modified 10× Genomics scRNA-seq workflow. Analysis of these cells revealed clusters of T lymphocytes and monocytes/macrophages (Fig. 4c) and smaller populations of neutrophils, dendritic cells and natural killer cells. Several of these populations displayed distinct regional distributions within the sample. For example, lymphocytes and neutrophils skewed toward the marginal region while macrophage populations were skewed toward the interior, matching observations made in a subcutaneous colorectal cancer tumor model using whole-volume imaging of cleared tissue[19] (Fig. 4d).

Subsampling the monocyte/macrophage cluster (Fig. 4e) we found that that margin cells were enriched in prototypical monocyte-like genes such as *Ly6c2*. Conversely, interior cells were predominantly enriched for terminal tumor-associated macrophage (TAM) markers such as *C1qc* and *Apoe* (Fig. 4e,f)[20]. Exploring this further, we generated a pseudotime trajectory using Monocle[21] with the $Ly6c^{Hi}$, $Ccr2^{Hi}$ state as the root state (state 1) (Fig. 4h). We observed differentiation of several TAM states (states 2–4) from the root monocyte state as pseudotime advanced with graded changes in gene expression such as loss of *Ly6c2* and *Ccr2* expression and gain of other TAM defining markers (Supplementary Fig. 3). These terminal TAM states could be defined by expression of marker genes (Fig. 4g and Supplementary Fig. 4) consistent with previously described TAM markers[22]. When we overlaid regional localization onto our pseudotime trajectory, we observed the regional localization of cells shift from margin to interior as pseudotime progressed from state 1 to terminal states 2, 3 and 4 (Fig. 4i). The terminal states exhibited their own differences in regional localization with state 2 more marginal versus states 3/4 (Fig. 4i) and differentially expressed genes based on localization (Supplementary Fig. 5).

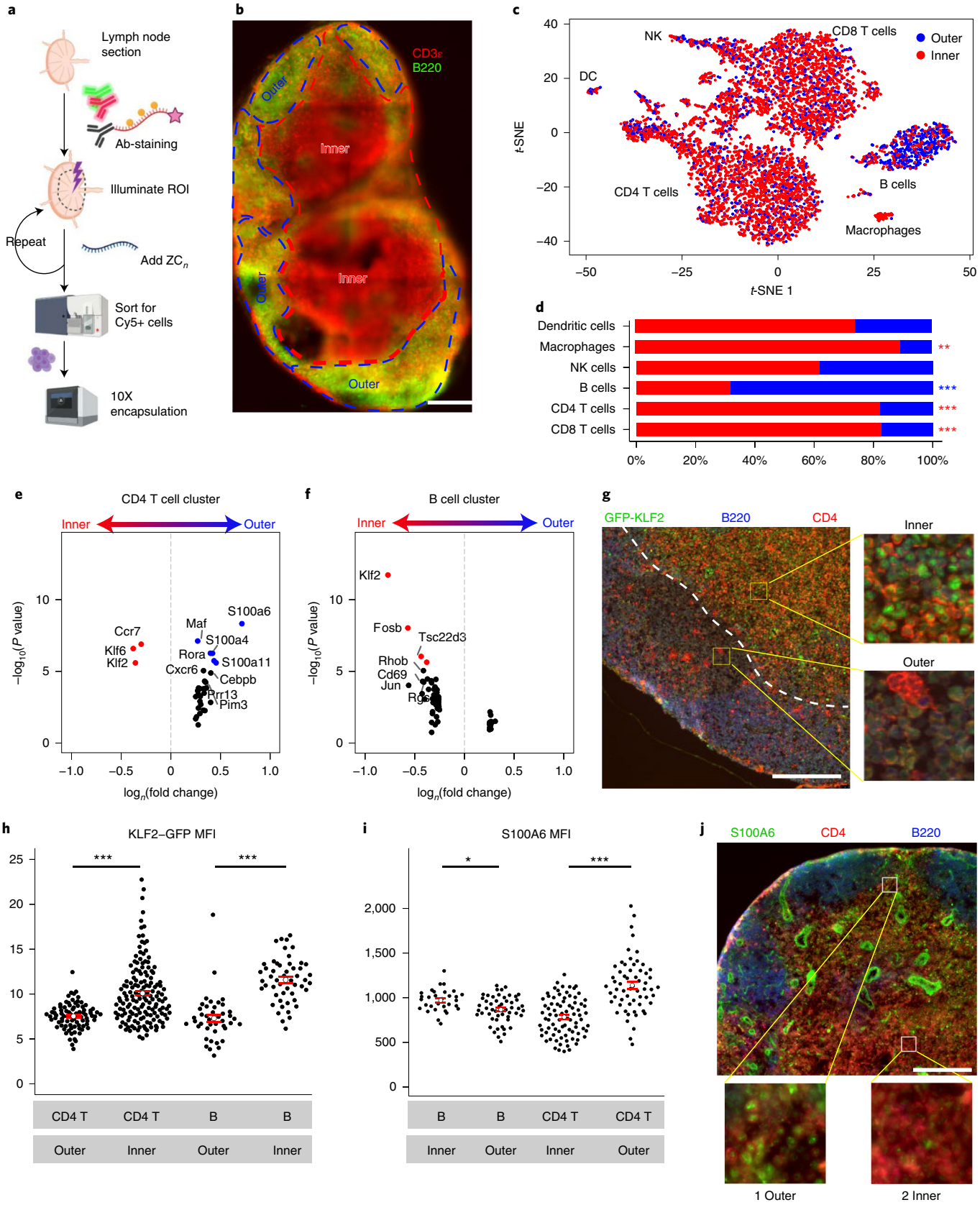
Focusing on the antigen-specific (*Gfp*-expressing) OTI T cells, we also observed segregation of cells in transcriptional space based on regional location (Fig. 4j). We observed a clear enrichment for genes previously associated with exhaustion (versus naïve) in interior localized antigen-specific T cells (Fig. 4k)[23,24]. Given that T cell exhaustion represents a graded process, we also applied a terminal versus stem-like exhaustion signature and observed a clear increase in terminal versus stem-like exhaustion score for interior cells (Fig. 4k)[25]. Similarly, when we performed differential expression analysis for marginal versus interior cells in this GFP^+ subset, the most significant gene hits were enriched in those defining earlier differentiation (for example, *Tcf7* and *Myb*) in margin T cells[26,27] and more committed exhaustion (for example, *Id2* and *Pdcd1*) in the interior T cells (Fig. 4l)[24,25].

Increased resolution in LN reveals spatial patterns of gene expression. To increase the number of labeled regions, we devised two variations of ZipSeq. Instead of a terminating zipcode sequence, we used DNA duplex strand bearing an orthogonal NPOM-caged overhang (O2) sequence to effectively swap the potential binding site from O1 to O2 on illumination in the first round. Multiple regions can be defined at a time, through downstream addition of zipcodes bearing either a complement to overhangs O1 or O2. This approach can theoretically be scaled up through the use of additional orthogonal overhang sequences, resulting in definition of 2^N regions using $N+1$ rounds of illumination and oligonucleotide addition. Figure 5a demonstrates the ability to generate four distinct regions.

Using this design, we looked for gradients of gene expression in the LN. We adoptively transferred B cells (carboxyfluorescein succinimidyl ester (CFSE)-labeled) and CD8 T cells (RFP-labeled) into a C57Bl/6 mouse before gathering and sectioning the LN that revealed T and B zones (Fig. 5b). The sequence of illumination patterns and oligonucleotide additions to generate four concentric regions of labels (ZC1–4) is illustrated in Fig. 5b. Following dissociation

Fig. 3 | ZipSeq mapping of single immune cell transcriptomes within a live lymph node section. a, Schematic of workflow for LN study. A LN was taken from a C57Bl/6 mouse and sectioned into 150 µm slices. Following this, the section was stained for B220 and CD3ε along with the anti-CD45 Ab conjugated anchor strand (with internal Cy5 modification) hybridized to a caged strand. The section was imaged and ROI were illuminated before zipcode 1 or 2 addition. Tissue was then dissociated and labeled live cells ($Cy5^+$) were sorted for encapsulation. **b**, Composite stitched image of LN section used with B220 marked in green and CD3ε in red to delineate inner and outer regions used for zipcoding in subsequent study. Scale bar, 400µm. **c**, *t*-SNE dimensional reduction of sorted live, $Cy5^+$ cells following the10× Genomics scRNA-seq workflow. Assigned regional ID based on ZC1:ZC2 counts overlaid. Immune cell populations were identified using known expression markers on Immgen (n=7,019 cells with n_{Outer}=1,685 and n_{Inner}=5,334 for **c** and **d**) (mean n_{UMI}=4,100 and mean n_{Gene}=1,057: cutoffs used were n_{Gene}>400, percentage mitochondrial <0.15). **d**, Regional distributions of major immune cell populations as identified in **c**. Asterisks denote significance of enrichment with color indicating direction (inner versus outer). ***P<0.0005 **P<0.005 and *P<0.05 calculated by using a hypergeometric distribution to model random selection from the pool of all cells without replacement. Exact P values for DC, Mac., NK, B, CD4 T and CD8 T were 0.42, 0.0014, 0.36, 2.5×10^{-135}, 3.3×10^{-17} and 1.2×10^{-13}, respectively, representing likelihood of obtaining an enrichment greater than or equal to that observed. Asterisk color denotes enrichment toward the outer (blue) or inner (red) relative to the whole population. **e**, Volcano plot showing differential gene expression analysis within the CD4 T cell subpopulation. Colored points represent genes with P<0.05 as calculated by Seurat's built-in nonparametric Wilcoxon rank-sum test (Bonferroni adjusted) (n=1,474 cells). **f**, Same as in **e** for the B cell population (n=626 cells). **g**, Immunofluorescence imaging of fixed LN section taken from a GFP-KLF2 reporter mouse. Section was stained for GFP, CD4 and B220. Dotted line represents demarcation between inner and outer regions used during quantification. Scale bar, 200µm. Zoomed-in insets show representative fields within inner and outer regions. **h**, Mean fluorescence intensity of GFP-KLF2 signal intensity within CD4 T and B cells in immunofluorescence image from **g**, either the inner or outer region (n=84, 152, 51 and 42 cells for B outer, B inner, T outer and T inner, respectively). **i**, Mean fluorescence intensity of S100A6 signal within CD4 T and B cells found in the outer and inner regions of the LN in the immunofluoresence image from **j** (n=81, 60, 31 and 52 cells for T inner, T outer, B inner and B outer, respectively). For **h** and **i**, bee-swarm plots represent intensities of individual cells with bars denoting standard error and center defined as mean. P values for significance testing calculated by two-tailed Wilcoxon's rank-sum test. **j**, Fixed frozen LN section stained for CD4, B220 and S100A6. Zoomed-in insets show representative fields from outer and inner regions, Scale bar, 200µm. Images and quantification in **g**–**j** representative of two independent experiments using sections acquired from the same animal and stained, acquired and analyzed separately. **c**–**f**, Showing scRNA-seq data are representative of two independent experiments performed on LNs from distinct animals, imaging and sequencing runs.

and sequencing, we observed four groupings of cells with dominant zipcode counts for 1–4 while filtering out cells that were ambiguous with no clear zipcode dominance (Supplementary Fig. 6). We observed a strong enrichment for regions 3 and 4 in the B cell cluster versus T cell clusters (Fig. 5c). We noted that the degree of region 4 enrichment in B cells versus T cells (50% versus 4%) (Fig. 5d) was greatly increased compared to our two-region LN study in Fig. 3. In addition, we observed the natural killer cell population enriched



specifically at the interface between T and B cell zones (region 3) as has been previously observed[28] (Fig. 5d).

Definition of four concentric regions allowed us to identify the existence of gradients of gene expression in space. As predicted, within CD4 T cell clusters, expression of *Ccr7* and *Klf2* steadily decrease moving outward toward the B cell zone while *S100a6* expression increases, matching our findings from Fig. 3 and published results[29] (Fig. 5e). In contrast, B cell zone chemokine *Cxcl13* was significantly expressed in region 4 alone (Fig. 5e). This type of analysis allowed us to compare patterns of gene expression across space using genes such as *Klf2* and *S100a6* as references, and calculating cross-correlation scores with all other expressed genes in our dataset, identifying the most 'similar' and 'dissimilar' genes. For example, *Klf2* shares similar profiles with *Nkg7* and *Actn1* (with which it shares a similar differentiation trajectory in vitro)[30] while being most dissimilar to *Rora* and *Nfkb1*. Meanwhile *S100a6* shares its profile with the closely related *S100a4* as well as the costimulatory molecule *Icos* while being most dissimilar to *Syngr2* and *Mrpl14* (Fig. 5f,g).

This approach, while allowing for definition of exponentially increasing number of regions, also requires a similar scaling of distinct orthogonal caged sequences that could become cost prohibitive, so we devised a second variation. Here, each coding segment consists of a zipcode 'block', which is a duplex of a polyA, barcode, a universal hybridization region and an overhang sequence O. This strand is prehybridized to a strand with the universal hybridization region and a caged overhang sequence O′ (Fig. 5h). This schematic necessitates synthesis of only a single caged sequence species and N distinct barcoded polyA strands yielding potentially 2^N regions after N rounds of illumination and addition. Using an in-tube validation experiment, we were able to observe four separate populations of cells using two rounds of illumination and zipcode addition (Supplementary Fig. 7). Merging of cell type identity and zipcode combination showed good agreement with the experimental scheme. To visualize this approach, we used three separate illumination and addition steps with three zipcode blocks, each bearing a distinct fluorophore. This yielded eight (2^3) distinct color combinations, each defining a grid position (Fig. 5i). We repeated this using 20-µm grid squares, demonstrating the capability to define areas on the order of cell neighborhoods (Supplementary Fig. 8).

Discussion

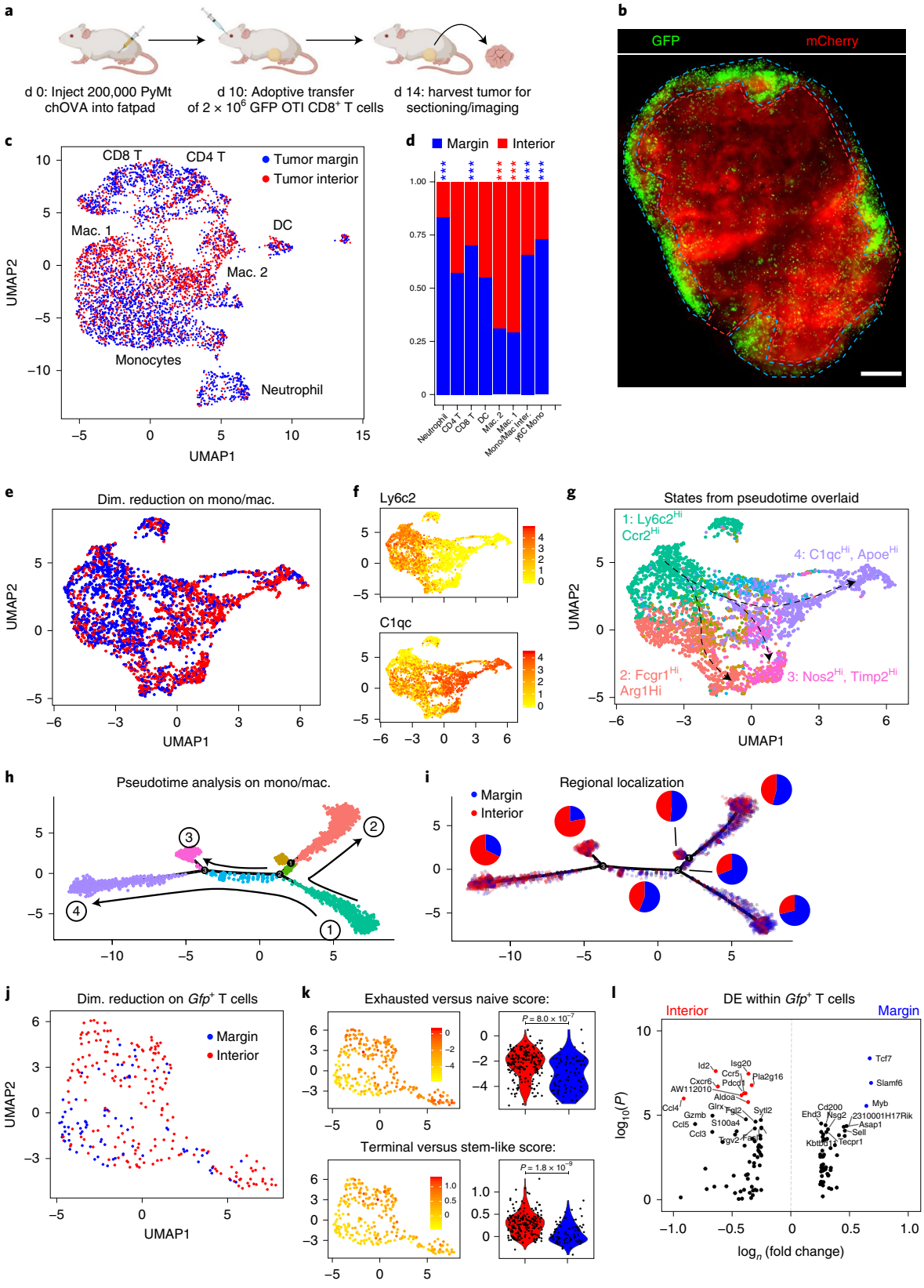
Here we introduce ZipSeq, an approach that allows for on-demand barcoding of cells within defined regions during microscopy. The approach is compatible with live tissue sections and precludes the need for genetically encoded photoactivatable proteins[31] so is applicable to human tissues and allows for definition of multiple regions at once. ZipSeq plugs into the commercially available 10× Genomics workflow[32], and is theoretically compatible with many other scRNA-seq methodologies[33–35], requiring only caged oligonucleotides and a photo-patterning module.

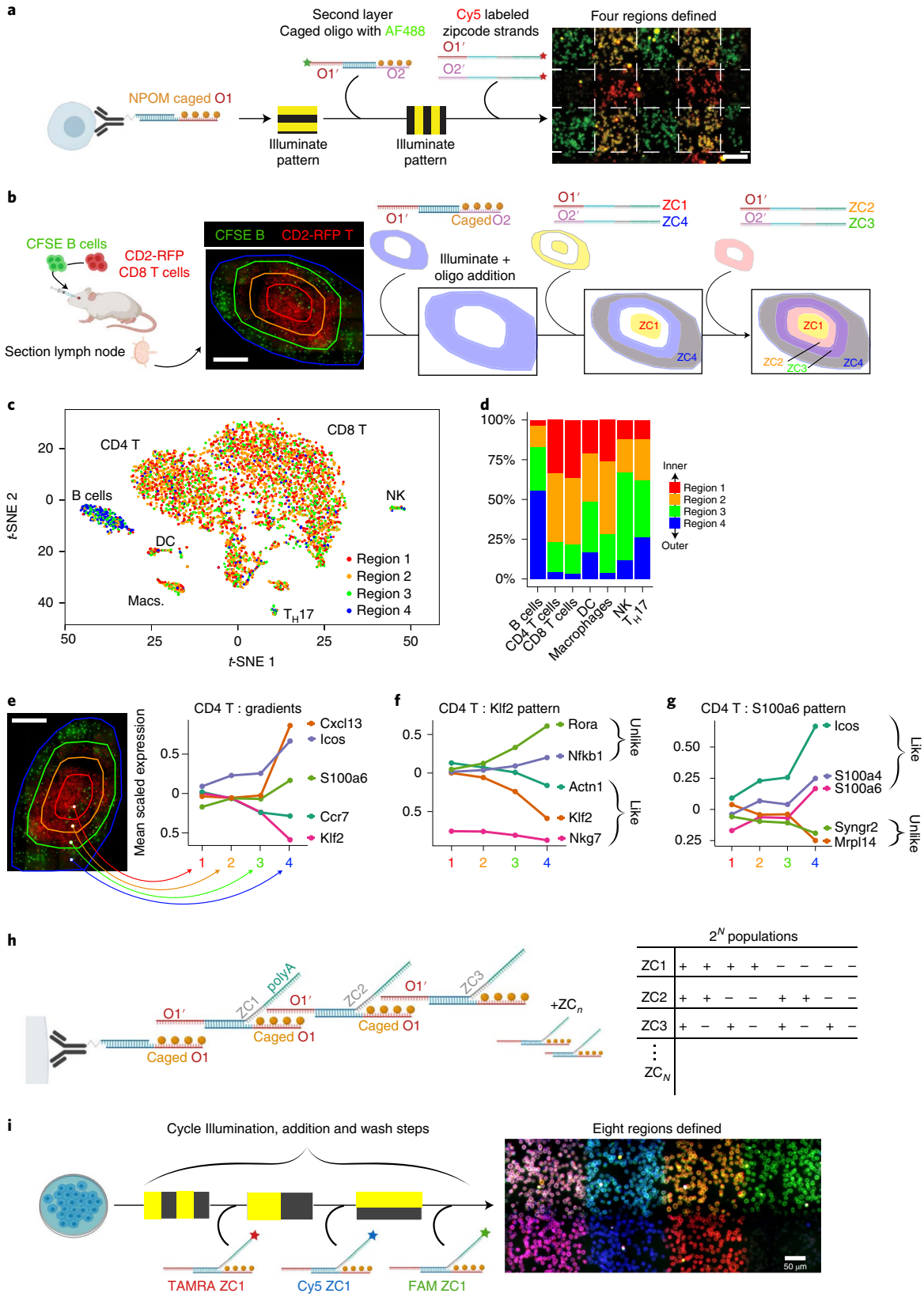
Using ZipSeq, we demonstrate the ability in both an in vitro wound healing model and ex vivo tissue sections (LN and tumor) to assign single cell transcriptomes to regions defined concurrently with fluorescence imaging. In the wound healing model, our approach identified distinct transcriptional programs activated in fibroblasts as a function of distance from the wound edge. We found a migratory cell state enriched at the leading edge (0–200 µm) and a proliferative state enriched at the rear (200–400 µm). This spatial segregation between migration and proliferation has been previously observed in multiple cell types such as epithelial cells

Fig. 4 | ZipSeq mapping of immune cell transcriptional states in the tumor microenvironment. a, Schematic of experimental setup. Here, 200,000 PyMT-ChOVA tumor cells were injected into the inguinal mammary fatpad of 8-week-old female C57Bl/6 mice. After 14d, 2×10^6 CD8 T cells from a GFP OTI mouse were adoptively transferred. After four more days, each tumor was collected, sectioned, imaged and labeled with anti-CD45 Ab conjugated anchor strand (with internal Cy5 modification) hybridized to a caged strand. **b**, The section was imaged and ROI were illuminated before zipcode 1 or 2 addition as denoted. Imaging of 150-µm-thick live tumor section used for scRNA-seq in following experiments. Red channel denotes mCherry signal from PyMT-ChOVA tumor cells and green channel denotes adoptively transferred GFP OTI T cells. ROI used for zipcode labeling shown overlaid. Scale bar, 400 µm. **c**, UMAP representation of sorted live Cy5$^+$ cells following 10× Genomics scRNA-seq workflow. Cells below n_{UMI} and zipcode count threshold or above mitochondrial percentage threshold were filtered out. Assigned regional ID based on ZC1:ZC2 counts overlaid. Large scale populations annotated based on similarity to known markers on Immgen (n=4,916 cells with n_{Margin}=2,783 and $n_{Interior}$=2,133 cells) (mean n_{UMI}=22,500, mean n_{Gene}=3,939 and mean n_{ZC}=6,564: cutoffs were n_{Gene}>500, percentage mitochondrial <0.15). **d**, Stacked bar charts denoting regional distributions of major immune cell populations from data in **c**. Exact P values for Neut., CD4, CD8, DC, macrophage (mac.) 2, mac. 1, monocyte (mono)/mac inter. and mono. were 7.4×10^{-20}, 0.13, 6.3×10^{-9}, 0.43, 5.8×10^{-81}, 1.8×10^{-30}, 1.2×10^{-7} and 4.3×10^{-22}, respectively, representing the likelihood of obtaining an enrichment greater than or equal to that observed. Asterisk color denotes enrichment toward outer (blue) or inner (red) relative to the whole population. **e**, UMAP dimensional reduction on monocyte/macrophage population subset with regional identity as determined by ZC1:ZC2 ratio (n=3,144 cells with n_{Margin}=1,593 and $n_{Interior}$=1,551 cells for subfigures **e**–**i**) (mean n_{UMI}=25,000, mean n_{Gene}=4,239 and mean n_{ZC}=5,268: cutoffs, n_{Gene}=500, percentage mitochondrial < 0.15). **f**, Feature plot of UMAP representation in **e** with normalized gene expression denoted by color scale for *Ly6c2* as a marker for monocytes and *C1qc* as a TAM marker. **g**, UMAP representation of monocyte/macrophage population with state identity calculated from Monocle pseudotime analysis in **h** overlaid. Arrows represent differentiation trajectory from the monocyte population to the terminal macrophage populations. Each major state is annotated with selected marker genes. **h**, DDR Tree dimensional reduction of monocyte/macrophage population as computed by Monocle with state identities overlaid. Arrows denote increasing pseudotime with the *Ly6C*^{Hi}, *Ccr2*^{Hi} state designated as the root. **i**, DDR Tree dimensional reduction of monocyte/macrophage population plotted with regional localization overlaid. Pie charts represent regional distributions (marginal versus interior) for each state. **j**, UMAP dimensional reduction on cells within the T cell clusters expressing at least one GFP transcript. Regional identity as determined by ZC1:ZC2 ratio overlaid (n=265 cells for subfigures **j**–**l**) (mean n_{UMI}=24,700, mean n_{Gene}=4,083, mean n_{ZC}=12,000: cutoffs were n_{Gene}>500, percentage mitochondrial <0.15). **k**, UMAP representation with gene expression signature scores overlaid. Exhaustion versus naïve gene signature scores were calculated for the *Gfp*$^+$ T cell subpopulation (cells within lymphoid clusters with at least one *Gfp* transcript) and these scores overlaid on the UMAP representation. Violin plot represents this score distribution based on regional identity. The bottom row represents similar quantification of a terminal versus stem-like exhausted signature score. P values calculated using a two-tailed Wilcoxon test. For exhausted versus naïve score, n=265 cells with N_{Margin}=76 and $N_{Interior}$=189; minima, maxima, center and 25th/75th percentiles; −5.5, 0.47, −2.062 and −2.85/−1.50 for the interior, and −5.37, 0.49, −3.20 and −4.25/−2.00 for the margin. For the stem versus terminal score, n=265 cells; minima, maxima, center and 25th/75th percentiles; −0.29, 1.30, 0.27 and 0.08/0.42 for the interior and −0.41, 0.81, 0.02 and −0.09/0.21 for the margin. **l**, Volcano plot showing top differentially expressed genes in the *Gfp*$^+$ T cell subpopulation based on regional identity. Colored points represent genes with P<0.05 as calculated with Seurat's built-in nonparametric Wilcoxon rank-sum test (Bonferroni adjusted). **c**–**l** show scRNA-seq data representative of two independent experiments derived from separate animals, imaging and scRNA-seq runs.

and keratinocytes during wound healing[36,37] with overexpression of *Acta2* and *Serpine1* described at the leading edge in previous wound healing studies[38,39]. It will be informative to repeat this study at higher resolution at different timepoints to observe how these spatial patterns of expression might evolve.

In the context of LNs, this method reports spatially dependent gene expression validated by previous works including *Klf2*, *Ccr7* and *S100a6* expression[17,40,41]. With increasing region number, this method permits identification of genes that map 'similarly' or 'dissimilarly' to a known gene over space. In the context of a tumor





model, this method allows the progression of myeloid and T cell differentiation to be mapped to physical infiltration depth. The myeloid differentiation in particular is consistent with recruited monocytes receiving local cues that skew differentiation trajectories as they arrive in the tumor margins, as described previously in several tumor models[42,43].

We also observed genes associated with T cell exhaustion upregulated in tumor specific CD8 T cells when comparing tumor interior with margins. Notably, *Tcf7*, a main factor in maintenance of a stem-like exhausted phenotype, was enriched in marginal T cells versus interior[25,44]. Comparison with imaging data taken before barcoding suggests that the dispersed, deeper infiltrating

Fig. 5 | Increased mapping resolution reveals spatial patterns of gene expression in cell subpopulations. a, Schematic of oligonucleotide design used for defining four regions by adding a second layer of caged oligonucleotides. A secondary oligonucleotide duplex bearing an orthogonal caged O2 overhang can hybridize to the uncaged O1 sequence. With a combination of zipcode strands with either an overhang region O1′ or O2′, four distinct regions can be defined by four zipcode species. Workflow illustrates the ability to define four such regions through fluorescence tagging of these oligonucleotide strands. Scale bar, 100μm. **b**, Stitched microscopy imaging of inguinal LN section used for following scRNA-seq data. B cells and CD8 T cells denoted in green and red respectively. Regions 1–4 were defined with zipcodes as overlaid onto the image. Scale bar, 400μm. **c**, *t*-SNE dimensional reduction of sorted live, $Cy5^+$ cells from LN section shown in **b** with regional identity overlaid following the scRNA-Seq workflow. Major immune populations are annotated: n=5,489 cells; mean n_{UMI}=3,900 and mean n_{Gene}=1,273; cutoffs were n_{Gene}>400, percentage mitochrondrial <0.15. **d**, Bar chart illustrating distribution of cells in each of the four regions for selected immune cell populations in **c**. **e**, Plots of mean scaled gene expression levels within the CD4 T cell cluster as a function of regional assignment for selected genes. The four regions used are shown overlaid on the LN image from **b**. Scale bar, 400μm. **f**, Plots of average scaled gene expression levels within the CD4 T cell cluster as a function of regional assignment for *Klf2* and two similar and dissimilar genes as calculated by cross-correlation score. Genes with significantly different expression levels and a logFC threshold of 0.4 between at least one pair of regions were considered for analysis. Cross-correlation scores were calculated between the averaged scaled expression levels of these genes and the reference gene. **g**, Similar analysis with *S100a6* as reference (n = 696, 910, 399 and 91) cells for regions 1, 2, 3 and 4, respectively). **h**, Schematic of second design iteration. Briefly, zipcode duplexes contain a barcode and polyA sequence hybridized to a universal photocaged overhang. In this way, zipcode blocks are added on in a combinatorial manner, defining 2^N populations based on presence or absence of a given zipcode block. **i**, Demonstration of combinatorial spatial barcoding of a field of cells in an exponentially scaling manner. Conjugate labeled CD8 T cell were plated and subjected to a 3× sequence of illumination patterns and zipcode block additions bearing distinct fluorophores resulting in eight regions with distinct fluorophore combinations. Scale bar, 50μm. Imaging shown in **a**, **h** and **i** is representative of two independent experiments. scRNA-seq results in **b**–**g** are representative of two independent experiments performed on LNs from two distinct animals imaging and sequencing runs.

antigen-specific T cells we observed (Fig. 5b) are further along the exhaustion pathway compared to the T cells at the edges. The mechanistic link between depth and commitment toward exhaustion bears further investigation. We also noted the enhanced expression of chemokines/receptors in interior versus margin CD8 T cells associated with increased trafficking and infiltration of T cells into the tumor such as *Ccl4/5* and receptor *Ccr5* (Fig. 4l)[45–47]. Our work highlights a number of previously unappreciated spatial gene expression patterns in the tumor microenvironment that require further high-resolution transcriptional analysis such as fluorescence in situ hybridization (FISH) to correlate with.

Approaches built on FISH approaches such as MERFISH and SeqFish allow for subcellular resolution of thousands of transcripts at once, however, will require preselection of probes and can only be performed on fixed sections[5,6]. Additionally, throughput for three-dimensional imaging of larger tissue areas could be a limiting factor. Another class of spatial transcriptomics approach uses prefabricated grids of barcoded poly-dTs or barcoded beads bearing poly-dTs to capture transcripts within the grid position[7–9]. While these approaches offer excellent spatial resolution, it will be difficult to apply them directly to tissue following live imaging. The relatively low read depth offered by several of these approaches (~4×10^4 reads per spot with 100 μm spot-to-spot distance for spatial transcriptomics and 2×10^2 per 10-μm bead for SLIDE-seq), could result in significant loss of information for rarer or lower RNA content cell types as their transcripts become diluted out during the capture step. Because ZipSeq plugs into droplet-based scRNA-seq workflows, it has potential to tap into greater read depth per cell, generating true single cell transcriptomes without the need for deconvolution. Another advantage of ZipSeq is the potential to easily integrate with other multimodal measurements such as concurrent surface epitope labeling using CITE-seq or single cell immune repertoire sequencing[10].

ZipSeq, however, faces limitations, most notably in spatial resolution. We propose that ZipSeq is currently most effective in questions based around microanatomical features observed during imaging that can guide definition of ROI for barcoding. For improved scaling of resolution, we demonstrated the ability to add on layers of secondary caged oligonucleotides to exponentially increase the number of definable regions. With this increased spatial resolution, we can describe the segregation of different cell states in finer detail and detect genes with sharply defined spatial expression in an unbiased fashion.

While many of our studies focused on the immune compartment, alteration of anchoring moiety will potentially expand application to diverse multicellular models. By using a panel of oligonucleotide conjugated antibodies directed against various surface markers, spatial information gained from our approach could theoretically be combined with surface epitope profiling. Combined with live imaging of cells before Zipcoding, time-dependent cell behaviors such as motility can be used to define ROI (for example, transcriptional states of cells in low and high motility zones). In summary, ZipSeq represents a new approach to mapping scRNA-seq data from conventional scRNA-seq workflows using on-demand light-controlled hybridization of DNA barcodes onto cells. We propose that ZipSeq will strengthen our capability to link spatial heterogeneity in multicellular systems to transcriptional heterogeneity of the constituent cell populations.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-020-0880-2.

Received: 2 February 2020; Accepted: 26 May 2020;
Published online: 06 July 2020

References

1. Galli, S. J., Borregaard, N. & Wynn, T. A. Phenotypic and functional plasticity of cells of innate immunity: macrophages, mast cells and neutrophils. *Nat. Immunol.* **12**, 1035–1044 (2011).
2. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
3. Goltsev, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174**, 968–981.e15 (2018).
4. Angelo, M. et al. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* **20**, 436–442 (2014).
5. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, 1360–1363 (2015).
6. Eng, C. H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
7. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78 LP–82 (2016).
8. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).

9. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463 LP–1467 (2019).
10. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
11. Govan, J. M. et al. Optochemical control of RNA interference in mammalian cells. *Nucleic Acids Res.* **41**, 10518–10528 (2013).
12. Ando, H., Furuta, T., Tsien, R. Y. & Okamoto, H. Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat Genet.* **28**, 317–325 (2001).
13. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* **16**, 619–626 (2019).
14. Deiters, A. Light activation as a method of regulating and studying gene expression. *Curr. Opin. Chem. Biol.* **13**, 678–686 (2009).
15. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
16. Kowalczyk, M. S. et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
17. Lee, J. Y. et al. The transcription factor KLF2 restrains $CD4^+$ T follicular helper cell differentiation. *Immunity* **42**, 252–264 (2015).
18. Engelhardt, J. J. et al. Marginating dendritic cells of the tumor microenvironment cross-present tumor antigens and stably engage tumor-specific T cells. *Cancer Cell* **21**, 402–417 (2012).
19. Stoltzfus, C. R. et al. CytoMAP: a spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. *Cell Rep.* **33**, 107523 (2020).
20. Afik, R. et al. Tumor macrophages are pivotal constructors of tumor collagenous matrix. *J. Exp. Med.* **213**, 2315–2331 (2016).
21. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
22. Yang, M., McKay, D., Pollard, J. W. & Lewis, C. E. Diverse functions of macrophages in different tumor microenvironments. *Cancer Res.* **78**, 5492–5503 (2018).
23. Maimela, N. R., Liu, S. & Zhang, Y. Fates of $CD8^+$ T cells in tumor microenvironment. *Comput. Struct. Biotechnol. J.* **17**, 1–13 (2019).
24. Wherry, E. J. et al. Molecular signature of $CD8^+$ T cell exhaustion during chronic viral infection. *Immunity* **27**, 670–684 (2007).
25. Im, S. J. et al. Defining $CD8^+$ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016).
26. Jadhav, R. R. et al. Epigenetic signature of PD-1^+ $TCF1^+$ CD8 T cells that act as resource cells during chronic viral infection and respond to PD-1 blockade. *Proc. Natl Acad. Sci. USA* **116**, 14113–14118 (2019).
27. Chen, Z. et al. TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. *Immunity* **51**, 840–855.e5 (2019).
28. Garrod, K. R., Wei, S. H., Parker, I. & Cahalan, M. D. Natural killer cells actively patrol peripheral lymph nodes forming stable conjugates to eliminate MHC-mismatched targets. *Proc. Natl Acad. Sci. USA* **104**, 12081–12086 (2007).
29. Okada, T. & Cyster, J. G. CC chemokine receptor 7 contributes to Gi-dependent T cell motility in the lymph node. *J. Immunol.* **178**, 2973–2978 (2007).
30. Lee, M. S., Hanspers, K., Barker, C. S., Korn, A. P. & McCune, J. M. Gene expression profiles during human $CD4^+$ T cell differentiation. *Int. Immunol.* **16**, 1109–1124 (2004).
31. Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
32. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
33. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell.* **161**, 1202–1214 (2015).
34. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
35. Meador, J. P. et al. Massively parellel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
36. Safferling, K. et al. Wound healing revised: A novel reepithelialization mechanism revealed by in vitro and in silico models. *J. Cell Biol.* **203**, 691–709 (2013).
37. Park, S. et al. Tissue-scale coordination of cellular behaviour promotes epidermal wound repair in live mice. *Nat. Cell Biol.* **19**, 155–163 (2017).
38. Rockey, D. C., Weymouth, N. & Shi, Z. Smooth muscle α actin (Acta2) and myofibroblast function during hepatic wound healing. *PLoS ONE* **8**, e77166 (2013).
39. Simone, T. M. et al. SERPINE1: a molecular switch in the proliferation-migration dichotomy in wound-"activated" keratinocytes. *Adv. Wound Care* **3**, 281–290 (2014).
40. Weber, J. P. et al. ICOS maintains the T follicular helper cell phenotype by down-regulating Krüppel-like factor 2. *J. Exp. Med.* **212**, 217–233 (2015).
41. Zemmour, D. et al. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat Immunol.* **19**, 291–301 (2018).
42. Movahedi, K. et al. Different tumor microenvironments contain functionally distinct subsets of macrophages derived from Ly6C(high) monocytes. *Cancer Res.* **70**, 5728–5739 (2010).
43. Franklin, R. A. et al. The cellular and molecular origin of tumor-associated macrophages. *Science* **344**, 921–925 (2014).
44. Wang, Y. et al. The transcription factor Tcf1 preserves the effector function of exhausted CD8 T cells during chronic viral infection. *Front. Immunol.* **10**, 1–14 (2019).
45. Parsonage, G. et al. CXCR6 and CCR5 localize T lymphocyte subsets in nasopharyngeal carcinoma. *Am. J. Pathol.* **180**, 1215–1222 (2012).
46. González-Martín, A., Mira, E. & Mañes, S. CCR5 in cancer immunotherapy. *Oncoimmunology* **1**, 106–108 (2012).
47. De Oliveira, C. E. et al. CCR5-dependent homing of T regulatory cells to the tumor microenvironment contributes to skin squamous cell carcinoma development. *Mol. Cancer Ther.* **16**, 2871–2880 (2017).

Methods

Oligonucleotide list.

Amine-modified anchor strand for conjugation to antibody (Fig. 1 and Supplementary Fig. 1)
Amine-C6 linker: ATCGTTTTTTTTTTTTGATGTTGACGG
Amine-modified anchor strand with internal Cy5 fluorophore for conjugation to antibody (Figs. 3 and 4)
Amine-C6 linker: ATCCAG\iCy5\TTTTTTTTTTTTGATGTTGACGG
Caged Strand hybridized to anchor strand before labeling cells with overhang region O1 (NPOM modified thymidines highlighted in red) (Fig. 1 and Supplementary Figs. 1–5)
5′-CGATCTGTGGTTGCTACCCGTCAACATCAAAAAAAAAAAA-3′
Blocking Strand (O1) for neutralizing carried over zipcodes (all figures)
5′-CGATCTGTGGTTGCTAC-3′
Second layer oligonucleotide (contains overhang region O1′) (Fig. 5)
GTAGCAACCACAGATCGCATGAGTCGAATCTCCCAC
Second layer oligonucleotide (contains overhang region O1′) with AF488 (Fig. 5)
AF488-TTGTAGCAACCACAGATCGCATGAGTCGAATCTCCCAC
Second layer NPOM-caged oligonucleotide with overhang region O2 (NPOM modified thymidines highlighted in red) (Fig. 5)
5′-CCTAGATCATGCAGTTCCGTGGGAGATTCGACTCATG-3′
Zipcode 1 with complementary overhang O1′ (Supplementary Figs. 1–5)
GTAGCAACCACAGATCGCACCCGAGAATTCCATGATGCAAAAAAAAAAAAAAAAAAAAAAAAAAA
Zipcode 2 with complementary overhang O1′ (Supplementary Figs. 1–5)
GTAGCAACCACAGATCGCACCCGAGAATTCCAAGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAA
Zipcode 3 with complementary overhang O2′ (Fig. 5)
GAACTGCATGATCTAGGCACCCGAGAATTCCATCAACGAAAAAAAAAAAAAAAAAAAAAAAAAAA
Zipcode 4 with complementary overhang O2′ (Fig. 5)
GAACTGCATGATCTAGGCACCCGAGAATTCCATACGTGAAAAAAAAAAAAAAAAAAAAAAAAAAA
Fluorescent oligonucleotide for flow readout with complementary overhang O1′
GTAGCAACCACAGATCGTATA-Cy5
Adaptor strands for lipid anchored oligonucleotide (Fig. 2)
GATGCTTCACGGGATATTTTTTTTTTTTTGATGTTGACGG
TATCCCGTGAAGCTTGAGTGGAATCCCTTGGCACCCGAGAATTCCA
Lignoceric acid conjugated anchor strand (Fig. 2)
Lignoceric acid: GTAACGATCCAGCTGTCACTTGGAATTCTCGGGTGCCAAGG
Coanchor strand (Fig. 2)
AGTGACAGCTGGATCGTTAC: Palmitic acid
Design version 2 amine-modified anchor strand (Fig. 5 and Supplementary Fig. 7)
Amine-C6 linker: TTTTTCACCCGAGAATTCCAC
Design version 2 universal caged strand (Fig. 5 and Supplementary Fig. 7)
CGATCTGTGGTTGCTACGTGGAATTCTCGGGTG
Design version 2 generic zipcode strand (Fig. 5 and Supplementary Fig. 7)
GTAGCAACCACAGATCGCACCCGAGAATTCCACNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAA
Zipcode 1 strand with complementary overhang O1 with terminal Cy5 (Figs. 1 and 5 and Supplementary Fig. 7)
GTAGCAACCACAGATCGCACCCGAGAATTCCACTGATGCAAAAAAAAAAAAAAAAAAAAAAAA-Cy5
Zipcode 2 strand with complementary overhang O1 with terminal TAMRA (Figs. 1 and 5 and Supplementary Fig. 7)
GTAGCAACCACAGATCGCACCCGAGAATTCCACAGCCATAAAAAAAAAAAAAAAAAAAAAAAA-TAMRA
Zipcode 3 strand with complementary overhang O1 with terminal FAM (Figs. 1 and 5 and Supplementary Fig. 7)
GTAGCAACCACAGATCGCACCCGAGAATTCCACCTCTTCAAAAAAAAAAAAAAAAAAAAAAAA-FAM

All oligonucleotides save for the caged strand were ordered from IDT with high-performance liquid chromatography purification.

NPOM-caged strand was ordered as a custom synthesis from BioSynthesis. Details on properties of NPOM-caged dTs can be found at https://www.biosyn.com/oligonucleotideproduct/light-activated%20npom-caged-dt-modified-oligonucleotid.aspx

Reagents. Nuclease-free bovine serum albumin purchased from VWR (VWRV0332-25G).

Single-stranded salmon sperm DNA purchased from Abcam (ab229278).

The 10× v.2 and v.3 kits were purchased from 10× Genomics. Solid-phase reversible immobilization (SPRI) selection beads came from Beckman Coulter (B23317). Collagenase I and IV were purchased from Worthington Biosciences (LS004194) and (LS004186), respectively. The 2× Kapa HiFi HotStart Master Mix was purchased from Roche (KK2601).

PBS (minus calcium/magnesium), RPMI 1640 and DMEM were ordered from Gibco.

Antibodies. LEAF-purified anti-CD45 antibody (30F-11) (no. 103164), PE-anti-CD3ε (145-2C11) (no. 100307) and FITC-anti-CD45R(B220) (RA3-6B2) (no. 103205) purchased from Biolegend.

For IF, Abs for targets included anti-STMN1 (Abcam 52630), anti-S100A6 (Invitrogen PA5-16590) and anti-ACTA22 (Sigma 1A4).

Conjugation of anchor oligonucleotide with antibody. The Thunder-Link Plus kit (Expedeon)(425-0300) was used to conjugate the amine-modified anchor strand to an anti-mCD45 antibody (clone 30F-110) at a molar ratio of 1:5 Ab:oligo and allowed to conjugate overnight at room temperature before conjugate purification according to instructions.

Lipid conjugated oligonucleotide. Synthesis of the lipid conjugated anchor oligo was performed as in ref. [13]. To hybridize the caged oligonucleotide species, two adaptor sequences were also prehybridized with the anchor and caged strands (sequences shown in the Oligonucleotide list).

Mouse strains. Experiments were performed in 6–8-week-old female C57Bl/6J mice from JAX (no. 000664). For adoptive transfer, CD8 T cells were derived from a CD2-dsRed mouse (MGI no. 5296821) and a OTI (C57BL/6-Tg(TcraTcrb)1100Mjb/J) (no. 003831) crossed with a UBC-GFP (C57BL/6-Tg(UBC-GFP)30Scha/J) (no. 004353).

In-tube validation of zipcoding. The caged strand was prehybridized to the anti-CD45 Ab-anchor conjugate by adding a 1:1 molar ratio of caged to anchor strand and incubating at 37 °C for 15 min and allowing to cool to room temperature. CD4 and CD8 T cells were isolated from a mouse using CD4 and CD8 negative selection kits, respectively (StemCell Technologies nos. 19852 and 19853), and were labeled with the caged strand hybridized to the Ab-oligo conjugate. In the first round, CD4 T cells were illuminated with 365 nm light and the first zipcode strand added to both populations. Following a 4-min incubation and three washes with S2 blocking strand (0.1 μM), the CD8 population was then illuminated and zipcode 2 added and allowed to hybridize. Following a series of washes, the cells were pooled and encapsulated using a 10× Genomics v.2 3′ kit with a target cell number of 4,000.

For the in-tube validation experiment for Design 2 shown in Supplementary Fig. 7, bone-marrow-derived dendritic cells were isolated by flushing marrow from femur and tibia of adult B6 mice. After straining, cells were then plated at 1×10^6 per ml in a low adherent culture dish in DMEM + 10% FCS and pen/strep/glutamine with GM-CSF (7.5 ng ml^{-1}) (Peprotech 315-03). Two days before the experiment IL4 was supplemented at 20 ng ml^{-1}. Cells with lipopolysaccharide were given more lipopolysaccharide (1 mg ml^{-1}) 18 h before use. Bone-marrow-derived macrophages were derived similarly from flushed marrow but with 20 ng ml^{-1} M-CSF (Peprotech 315-02). LB27.4 cells obtained from ATCC (HB-99). Each population of cells was subjected to the specified sequence of illumination and zipcode addition using a similar protocol as described before.

Microscopy. Imaging and photo-uncaging was performed with a customized Zeiss Axiovert 200M inverted microscope fitted with a Mosaic DMD from Andor. Spatially directed photo-uncaging is accomplished through directing light from a mercury arc lamp onto a Mosaic DMD with an 800 × 600 micromirror array in plane with the sample. The sample can be simultaneously imaged using epi-fluorescent excitation. In the imaging software, a user-defined ROI is converted into a mask, which is reflected in the micromirror array. This spatially patterned light is then directed through the microscope and objective onto the sample. Illumination for widefield imaging was provided by a Lambda DG-4. For photo-uncaging, light from a mercury arc lamp was passed through a 365-nm bandpass filter then directed into the DMD. Imaging was performed with a ×20 objective from Zeiss (Plan-NEOFLUAR). For imaging, we used a Photometrics Evolve 512. Metamorph (Molecular Devices) was used to control the microscope and illumination by the Mosaic DMD with a custom Visual Basics program (available on Github, see link provided) to provide a user interface for manual ROI delineation across multiple fields of view.

On-scope validation of spatially controlled DNA hybridization. For imaging shown in Fig. 1 isolated primary mouse CD4 T cells were plated in a LabTek chamber slide coated with anti-CD3ε antibody to adhere cells. Cells were then labeled with Ab-oligo conjugate hybridized to the caged strand (here without internal Cy5 modification). Following two washes and blocking with single-stranded DNA to prevent nonspecific DNA-well interactions, squares with a width of 200 µm were illuminated with 365-nm light from a mercury arc lamp through the Mosaic. A fluorescently labeled zipcode strand (Cy5) was flowed into the well and incubated for 5 min for hybridization. Following three washes with media, the sequence was repeated twice for two other fluorescently labeled zipcode strands (TAMRA and FAM). The illuminated regions were then imaged.

For Fig. 5a, a similar setup was used; however, the first illumination pattern was made up of horizontal 200 µm width followed by hybridization of a duplex strand composed of O1' overhang with a FAM modification hybridized to a O2 photocaged overhang. Following illumination of vertical bands of 200 µm width, oligonucleotides with O1′ or O2′ overhangs with TAMRA modifications were added for hybridization.

For the combinatorial design shown in Fig. 5h and Supplementary Fig. 8, zipcode strands 1–3 were conjugated to fluorophores Cy5, TAMRA and FAM. These zipcode strands were then hybridized to the complementary caged oligonucleotide strand to form three zipcode blocks. Similar to the workflow described previously, a series of illumination, zipcode addition, washing, and blocking generated three patterns of zipcode hybridization overlaid, creating eight distinct color combinations. The grids used in Fig. 5h were designed to generate $200 \times 200\ \mu m^2$ squares. In Supplementary Fig. 8, the grid squares were scaled down to push the limits of DMD resolution to roughly $20 \times 20\ \mu m^2$ squares.

Wound-healing study. NIH/3T3 cells were ordered from ATCC (CRL-1658) and cultured in complete growth medium (DMEM + 10% FCS + 55 mM BME + PSG). Two days before labeling and scRNA-seq, a monolayer of 40,000 cells was plated in a well of a eight-well LabTek chambered coverglass. Cells were allowed to settle and fill in the well until 12 h before imaging and labeling.

A pipette tip was used to introduce a scratch around 0.8 mm in width in the monolayer. Cells were washed once with fresh growth medium to remove debris or floater cells and incubated for a further 12 h. Cells were then transferred to the microscope and the ROI we delineated. Growth medium was washed out and replaced with serum-free phenol red free DMEM with ssDNA to block nonspecific DNA binding in further steps. The anchor-caged strand lipid tag was added and allowed to sit for 10 min followed by the coanchor strand. Following three washes, the desired ROI were illuminated using an 800 ms pulse of 365-nm light.

The first zipcode strand was added to the monolayer and allowed to hybridize for 5 min. Following two washes, a blocking strand was added at a lower concentration for 5 min to prevent residual zipcode strand from binding in the following cycle. After another two washes, the second ROI was illuminated and the steps repeated with the second zipcode strand.

After the last series of washes, the medium was removed and Accutase (Stem Cell no. 07920) was added to detach the cells. After 5 min of incubation at room temperature, cells were gathered and washed with cold PBS + 0.04% bovine serum albumin (BSA) as recommended by the 10× Genomics protocol. Encapsulation was performed for a target cell number of 8,000 using the v.2 chemistry.

LN study. Inguinal LNs were collected from 8-week-old C57Bl/6 female mice and embedded while live in 2% agarose. Using a Leica Vibratome or Precisionary VF 310-0Z Vibratome, the LNs were sectioned into 150-µm-thick slices and affixed to a LabTek chamber slide (Thermo Fisher no. 155409) using Vetbond (3M) applied to the agarose 'rim'. Sections were then incubated with anti-B220 FITC and anti-CD3ε PE along with the anti-CD45 conjugated NPOM-caged anchor strand bearing an internal Cy5 modification for 1 h at 4 °C. Following three washes, the sections were imaged on the scope and ROI delineated and illuminated with pulses of 365-nm light for a 1,000-ms duration. The first zipcode was added at a 1 µM concentration in RPMI and allowed to incubate 10 min. Following three washes, the blocking strand was added at 0.25 µM concentration and incubated for a further 5 min. Following a series of three washes, the process was repeated for the second zipcode. Following a final blocking step, the section was mechanically disrupted and strained over a 100-µm nylon strainer and sorted for live, $Cy5^+$ cells. Sorted cells were washed with cold PBS + 0.04% BSA and encapsulated following 10× Genomics guidelines using a v.2 3′ kit with a target cell number of 8,000.

For the four-region LN study, naïve CD8 T cells were purified from a hCD2-dsRed mouse and 4×10^6 were adoptively transferred into a 6-week-old B6 mouse. Meanwhile, B cells were purified from a B6 mouse using an EasySep B cell negative selection kit (StemCell Technologies no. 19854) and labeled with carboxyfluorescein succinimidyl ester before being transferred at 4×10^6 per mouse. Three days following transfers, the mouse was killed and the inguinal LNs extracted for sectioning and study. Following sorting, cells were encapsulated as described above using a 10× Genomics v.2 3′ kit with a target cell number of 8,000.

Tumor study. For orthotopically injected PyMT-ChOVA models, the PyMT-ChOVA breast cancer cell line was generated from de novo mammary tumors in the PyMT-ChOVA breast cancer mouse model[18]. Briefly, mammary tumors were collected in ice cold PBS and mechanically minced into small fragments. Cells and tissue fragments were cultured in DMEM + 10% FCS and added penicillin, streptomycin and glutamine. After 7–10 d, tissue fragments and debris were washed out with ice cold PBS and attached cells were allowed to grow to confluency. Cells were cultured in growth medium for an additional 3–5 passages to generate the established PyMT-ChOVA mammary tumor cell line. Tumor cell line was kept frozen and thawed directly from stock as needed before injection to avoid unnecessary passages. Then 200,000 cells were injected in Matrigel (Corning no. 356238) into the inguinal fat pad of 8-week-old female C57Bl/6 mice.

C57BL/6-Tg(UBC-GFP)30Scha/J mice were crossed to OT-1 mice to generate a GFP OT-1 mouse strain. LNs were collected from a 6-week-old GFP OT-1 mouse and CD8 T cells were isolated using an EasySep mouse CD8 T cell negative selection kit (StemCell Technologies). Fourteen days after tumor injection, 5×10^6 CD8 T cells from a GFP OTI mouse were adoptively transferred through retro-orbital injection. After 4 d, the mouse was killed and the tumor collected for sectioning on a Leica Vibratome into 150-µm-thick slices. As before in the LN study, the tissue was embedded live into 2% agarose for sectioning. Sections were blocked with ssDNA and BSA in RPMI for 30 min at 4 °C then stained with the anti-CD45 Ab conjugated to NPOM-caged anchor strand with internal Cy5 for 1 h at 4 °C. Following washes, sections were affixed to an Ibidi µ-Slide eight-well using Vetbond. Sections were then imaged with the desired channels, and spatially zipcoded as described before using pulses of 365-nm light for 1,000 ms. Following the final block and wash step, tissue sections were diced finely and incubated with a Collagenase I and IV blend in RPMI and incubated for 30 min at room temperature. The resulting suspension was mechanically agitated by pipetting and then strained on a 100-µm strainer. As before, live Cy5+ cells were sorted out on a FACSAria II, washed in PBS + 0.04% BSA and then encapsulated following 10× Genomics specifications for v.3 3′ chemistry with target cell number of 8,000.

Library construction. Single cell cDNA library construction was performed as directed by 10× Genomics using v.2 or v.3 3′ chemistry depending on experiment. Following the published CITE-seq protocol, an additive primer (partial Read 2 small RNA) was spiked into the cDNA amplification reaction. During the post-cDNA amplification SPRI cleanup step, given that zipcode reads are significantly shorter (<200 basepairs), these reads were separated from cDNA reads by decanting the supernatant. cDNA reads bound to SPRI beads were processed as recommended in the 10× Genomics v.2 or v.3 user guide. Meanwhile the supernatant containing zipcode reads was saved and underwent two successive 3× SPRI cleanup steps. This library was then amplified using primers from CITE-seq protocol[10]. Following fragment analysis on the BioAnalyzer and library quantification by quantitative PCR, the zipcode library was mixed with the associated cDNA library at a 1:10 molar ratio and sequenced on either Illumina HiSeq Rapid Run mode (all studies except Fig. 4) or NovaSeq SP (Fig. 4 studies) using 10× Genomics recommended sequencing parameters based on kit version (v.2 versus v.3).

Processing of raw sequencing reads. Raw read files were processed using CellRanger bcl2fastq to separate cDNA and zipcode libraries. cDNA libraries were processed using standard CellRanger count function. For PyMT tumor studies that made use of alignment to the GFP transgene, the 10× Genomics provided mm10 reference was modified through addition of artificial genes made up of fluorophore coding sequences. Zipcode libraries were counted using a Python script made available by CITE-seq using a whitelist of cells provided by the CellRanger count function determined by a minimum n_{UMI} threshold[10]. The two outputs were then merged in Seurat for further analysis.

Data analysis. Cells with a high mitochondrial read count percentage (above 10% assumed to be dead or dying cells) were filtered out. Cells with low counts for cDNA were filtered out based on the presence of a local peak at the low end of the distribution. Read counts were normalized using log-normalization, scaled and centered and n_{UMI} and mitochondrial percentage regressed out. Principal components analysis was performed and the top ten of these principal components were used to inform the dimensional reduction by Seurat built-in *t*-SNE or UMAP (using the Python implementation of umap-learn package). Meanwhile, nearest neighbor clustering using these principal components was performed using Seurat's built-in FindClusters function at a specified resolution of 0.8. To call zipcode identities, zipcode counts were normalized and the ratios of these normalized counts used to gate cells as belonging to one identity or another with ambiguous cells (between gates) filtered from analysis. For the four-region LN study, *k*-means clustering on normalized zipcode counts was used to generate five clusters corresponding to ZC1–4 dominance along with a centrally localized ambiguous population.

For differential gene expression analysis, we used Seurat's built-in FindMarkers function that implements testing based on the Wilcoxon rank-sum test along with a Bonferroni correction to adjust *P* values. The testing was restricted to genes expressed in at least 10% of cells.

For Monocle analysis, we directly imported the Seurat object of interest into Monocle and used the clusters predefined in Seurat to get a list of differentially expressed genes between clusters[21]. The top 800 genes were then used in the DDRTree dimensional reduction and pseudotime ordering.

For signature analysis, curated lists of genes from literature were passed into Single Cell Signature Explorer to generate scores[48]. Sources for gene lists used for signature scores derived from the following:

- S and G2M phase: from built-in gene lists in Seurat
- Exhaustion versus naïve[24]
- Terminal versus stem-like exhaustion[25]

For analyzing spatial gradient profiles of genes in the four-region LN study, we used the mean-scaled and normalized gene expression for a given gene in each region. To find the most similar and dissimilar gene profiles, we first filtered out genes expressed in <10% of cells for all regions. We then calculated a cross-correlation score compared to a reference gene and ranked these scores from lowest correlation (−1) to highest correlation (1):

$$\mathrm{CC} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\mathrm{var}(x) \times \mathrm{var}(y)}}$$

for genes x and y, where i represents regions 1–4.

For doublet detection, DoubletFinder v.1 was used as described in ref. [15] (https://github.com/chris-mcginnis-ucsf/DoubletFinder). The pK parameter was determined using a parameter sweep as described with the nExp calculated from the number of cells belonging to each cluster determined by Seurat.

Immunofluorescence: for immunofluorescence study on KLF2-GFP reporter mouse LNs, sections were prepared, stained and imaged from PE-primed mouse on day 14 as in ref. [17]. For immunofluorescences of S100a6, LNs from 8-week-old C57Bl/6 mice were embedded in optimal cutting temperature compound (Sakura no. 4583) and cryosectioned into 10-µm-thick slices. Slices were fixed with 4% PFA and permeabilized by TritonX-100. Following blocking with 5% goat serum, primary antibodies were added and incubated at 4 °C overnight. Following washes, slices were then incubated with secondary Abs for 1 h at room temperature, washed and then incubated for 5 min at room temperature with DAPI. Following another wash, slices were mounted with Vectashield (Vector Labs H-1900) and a coverslip, then imaged.

For immunofluorescence studies on wound healing monolayers, NIH/3T3 fibroblasts were prepared in LabTek chambered slides as before and 12 h after wounding were fixed with 4% PFA, permeabilized with TritonX-100 and blocked with 5% goat serum.

Quantification of LN immunofluorescence: analysis was performed within Fiji. Depending on the cell type of interest (CD4 T or B cells), CD4 or B220 signal was used to identify cells within the tissue. Identified cells were selected randomly based on this signal and mean fluorescence intensity (MFI) of target channel was calculated within this mask. Wilcoxon rank-sum test (two-tailed) used for comparison of MFIs between populations.

For wound healing, binary masks for cell area were generated through thresholding of images and applied to target channel. Columns of 10-pixel width were taken progressively from the wound edge and pixels in the mask were averaged to generate a column fluorescence intensity average.

Statistical testing. Wilcoxon rank-sum test (two-tailed) used for comparison of MFIs between populations for immunofluorescence data. To see whether regional enrichment for a given immune cell population from scRNA-seq data was significant, we used a hypergeometric distribution to determine the probability of drawing a regional composition with equal to or greater enrichment compared to observed at random from the total cell pool. This probability was directly reported as the P statistic.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

For all scRNA-seq studies described here, transcript counts as determined by CellRanger count function as well as raw zipcode fastqs/counts as well as the modified genome for transgene alignment during CellRanger count can be found in the Gene Expression Omnibus under accession number GSE145502.
Lists of gene hits from differential expression analysis can be found in the Extended Data.
Raw image files from which figures are derived from can be found on Dryad at https://doi.org/10.7272/Q6H993DV.

Code availability

Visual Basics code for custom Metamorph User Program for delineation of multi-FOV spanning ROI for Mosaic illumination can be found on Github: https://github.com/BIDCatUCSF/VB-Plugin-for-Patterned-Illumination.
Python script used to generate zipcode counts from fastq available from CITE-seq[10] (https://hoohm.github.io/CITE-seq-Count/).

References

48. Pont, F., Tosolini, M. & Fournié, J. J. Single-Cell Signature Explorer for comprehensive visualization of single cell signatures across scRNA-seq datasets. *Nucleic Acids Res.* **47**, e133 (2019).

Acknowledgements

We thank the Biological Imaging Development Center at the University of California San Francisco (UCSF) for help with microscopy data collection and instrumentation. We also thank the Parnassus Flow Cytometry Core for flow cytometry instrumentation, supported by grant no. P30DK063720 and the Institute for Human Genomics at UCSF for sequencing and bioinformatics support. In addition, we thank the Computational Biology and Informatics core at the UCSF Helen Diller Family Comprehensive Cancer Center for computing resources. This work was supported from NIH/NCI grant nos. P30DK063720 (M.F.K.), 1R01CA197363 (M.F.K.) and R01GM135462 (Z.J.G.), and by the Parker Institute for Cancer Immunotherapy Opportunity grant (PICI). K.H.H. was supported by a NIH T32 training grant (no. 5T32CA177555-02) and is a PICI Scholar. M.F.K. is a PICI member researcher.

Author contributions

K.H.H. conceived and performed the experiments, analyzed data and wrote the manuscript. M.F.K. conceived the experiments, provided administrative and financial support, and wrote the manuscript. K.K. generated the PyMT-chOVA cell line. C.S.M., D.M.P., E.D.C. and Z.J.G. generated the lipid anchored oligonucleotide. J.P.E. developed custom interface for controlling illumination patterns during imaging. S.C.J. generated the KLF2-GFP reporter mouse and associated IF data. A.A.R. assisted with analysis of scRNA-seq data.

Competing interests

K.H.H. and M.F.K. are listed on a patent application regarding the ZipSeq approach.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-020-0880-2.

Correspondence and requests for materials should be addressed to M.F.K.

Peer review information Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Matthew Krummel

Last updated by author(s): 2020/02/05

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Raw .bcl files from sequencer demultiplexed and converted to fastq using Cellranger (v3.0.2) mkfastq. Cellranger (v3.0.2) count used to align to genome to get gene counts. Python script from M. Stoeckius et al. 2017 used to count zipcode sequences For Microscopy data acquisition Metamorph (v7.6.5.0) was used
Data analysis	All scRNA-Seq data analysis performed within Seurat v2.3.4 (Published by Satija Lab) and Monocle 2 (Published by Trapnell Lab). Microscopy quantification performed in Fiji (v2.0.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Transcript counts as determined by CellRanger count function as well as raw Zipcode fastq's/counts as well as the modified genome for transgene alignment during CellRanger count can be found on GEO. (GSE145502) Available 6-5-2020

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes for scRNA-Seq were ultimately determined by cell recovery and 10X encapsulation efficiency. We aimed for target cell numbers to allow for enough cells (on the order of hundreds) in cell populations of interest while avoiding overloading of the chip. Cell numbers used during scRNA-Seq analysis considered when calculating significance in differential gene expression and distribution. For quantification of cell MFI's in Figure 3h,i we considered the variance and possible effect sizes we might observe and used a power calculation to estimate samples sizes needed.
Data exclusions	Data for scRNA-Seq was excluded based on commonly used metrics for dead cells or noise due to free mRNA. (high mitochondrial read % or low # detected Genes).
Replication	Verification of scRNA-Seq results from Wound heal and Lymph node (Fig. 2 and 3) done with immunofluorescence where possible. 4 region LN study in Figure 5 recapitulates gene targets found in the 2 region LN study in Figure 3 representing another replication. scRNA-Seq runs for Fig. 2,3,4,5,S1,S7 were representative of two independent runs derived from distinct starting cell pops/animals and separate imaging, barcoding and sequencing runs. All imaging figures were representative of at least two independent staining, acquisition and analysis experimental runs.
Randomization	No assignment of participants/animals to experimental groups. Sampling of cells for IF quantification was done randomly from the pool of identified cell masks.
Blinding	Blinding not relevant, analysis was not based on subjective measures or selection

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	LEAF Purified anti-mCD45 (Biolegend, clone: 30F-11, 103164) (Lot B240479. Estimated dilution used: 2.5 ug/mL) PE anti-CD3e (Biolegend, clone 145-2C11, 100307) (B253779 1ug/mL) FITC anti-CD45R (eBioscience, clone RA3-6B2, lot E00308-1632) (1 ug/mL) anti-STMN1 (Abcam, EP1573Y, 52630) (Lot GR302254-8) anti-S100a6 (ThermoFisher, polyclonal, PA5-16590) (Lot TH2611115) (1:250) anti-ACTA2 (Sigma, clone 1A4, A2547-100UL) (107M4893V) (1:250) anti-CD45R (Biolegend, RA3-6B2) (B230687) (5 ug/mL) anti-CD4 (Biolegend, GK1.5) (B271741) (5 ug/mL) Goat anti-Rabbit IgG AF488 (Polyclonal)(Lot 2051237) 2 ug/mL Goat anti-rat IgG Af647(Clone Poly4054) (Lot B269695) 1 ug/mL Goat anti-mouse IgG AF568 (polyclonal, lot 1613919) 2ug/mL Goat anti-rabbit IgG AF568 (polyclonal, lot 1180090) 2 ug/mL
Validation	All Biolegend antibodies have multiple citations, and published flow cytometric validation data on website anti-STMN1 Ab tested for ICC/IF with 21 cited publications

anti-S100a6 tested on FFPE samples with 1 cited publication
anti-Acta2 tested for IF and IHC with 1140 cited publications

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	NIH3T3's and LB27.4 acquired from ATCC, PyMTChOVA cell line derived in-house using method described
Authentication	NIH3T3's and LB27.4's are authenticated by ATCC using STR. Cells for studies were directly ordered from ATCC at the time of use and used at low passage numbers <5 to avoid genetic drift or contamination issues. PyMT ChOVA cells were verified by sequencing for the transgene cassettes.
Mycoplasma contamination	Untested
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines used

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	-Isolated CD4/8 T lymphocytes, BMDC's and BMDM's derived from 6-8 week old C57Bl/6 mice male and female obtained from JAX -For Lymph node studies, lymph nodes taken from 8 week old female C57Bl/6 mice obtained from JAX. Tumor studies also used 8 week old female C57Bl/6 mice. -For adoptive transfers: B cells derived from 6-8 week old female C57Bl/6 mice CD8 T cells from either CD2-DsRed (Tg(CD2-dsRed)#Kio)(MGI:5296821) (6-10 week old females) or OTI (JAX 003831) crossed with B6-GFP (JAX 004353) (6-10 week old females)
Wild animals	No Wild animals were used in this study
Field-collected samples	No Field-collected samples were used in this study
Ethics oversight	Approved by UCSF IACUC

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For supplementary figure 1A, source of cells was from blasting primary murine lymphocytes isolated from spleen/lymph nodes of B6 mouse. CD4 and CD8 T cells were isolated from a mouse using CD4 and CD8 negative selection kits respectively (StemCell Technologies #19852 #19853) and blasted using anti-CD3/anti-CD28 beads. Following 2 days of stim, cells were separated from beads, and supplemented with IL-2. Cells were harvested and stained with anti-CD4 and anti-CD8 antibodies prior to flow cytometry. Cells from the same starting populations were used for the scRNA-Seq shown in the rest of Supp. Figure 1. For supplemental figure g,h, the caged strand with overhang O1 was pre-hybridized to the anti-CD45 Ab-anchor (Cy5) conjugate by adding a 1:1 molar ratio of caged to anchor strand and incubating at 37C for 15 minutes and allowing to cool to RT. In the first round, CD4 T cells were illuminated with 365 nm light and the first Zipcode strand conjugated to TAMRA fluorophore was added to one population while the other received Zipcode strand conjugated to FAM fluorophore. Following a 5 min. incubation and 2 washes, the cells from both populations were pooled in medium and allowed to incubate for 0', 20', 40', 60', 80' at RT to mimic Zipcoding assay conditions. Cells were then spun down, resuspended and assayed using the flow cytometer. Displayed normalized fluorescence intensity indicates TAMRA and FAM signal normalized to Cy5 to represent 'occupancy' of open anchor strands conjugated to Antibody. For supplemental figure i-k, the caged strand with overhang O1 was pre-hybridized to the anti-CD45 Ab-anchor (Cy5) conjugate by adding a 1:1 molar ratio of caged to anchor strand and incubating at 37C for 15 minutes and allowing to cool to RT. In the first round, CD4 T cells were illuminated with 365 nm light and the first Zipcode strand conjugated to TAMRA fluorophore was added and allowed to hybridize as above. Cells were then put through a mock series of illuminations, washes and incubations to mimic Zipcode assay conditions, then stained with Zombie NIR as well as DAPI to identify dying/dead cells, where Zombie NIR - /DAPI –

	cells are defined as viable.
Instrument	LSRII from BD for analysis
Software	FlowJo v10 for analysis and FACSDIVA for collection
Cell population abundance	For Supp. Fig 1, CD4+ and CD8+ T cell populations were approx 95% pure as defined by gates.
Gating strategy	For Supp Fig 1, cells were gated based on FSC and SSC to identify cells vs debris. Then FSC-A vs. FSC-H along the diagonal for singlet cells. Gates for CD8 and CD4 positivity drawn using cells stained with isotype controls conjugated to matching fluorophores for subfigure a. For Subfigure G and H, cells were gated based on FSC and SSC to identify cells vs. debris. Then FSC-A and FSC-H used to identify singlets. For subfigures i-k, the gating strategy is shown in I.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.