



ARTICLE OPEN

Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly

Xianwen Ren¹, Guojie Zhong¹, Qiming Zhang¹, Lei Zhang¹, Yujie Sun¹ and Zemin Zhang¹

Single-cell RNA sequencing (scRNA-seq) has revolutionized transcriptomic studies by providing unprecedented cellular and molecular throughputs, but spatial information of individual cells is lost during tissue dissociation. While imaging-based technologies such as in situ sequencing show great promise, technical difficulties currently limit their wide usage. Here we hypothesize that cellular spatial organization is inherently encoded by cell identity and can be reconstructed, at least in part, by ligand-receptor interactions, and we present CSOmap, a computational tool to infer cellular interaction de novo from scRNA-seq. We show that CSOmap can successfully recapitulate the spatial organization of multiple organs of human and mouse including tumor microenvironments for multiple cancers in pseudo-space, and reveal molecular determinants of cellular interactions. Further, CSOmap readily simulates perturbation of genes or cell types to gain novel biological insights, especially into how immune cells interact in the tumor microenvironment. CSOmap can be a widely applicable tool to interrogate cellular organizations based on scRNA-seq data for various tissues in diverse systems.

Cell Research (2020) 30:763–778; <https://doi.org/10.1038/s41422-020-0353-2>

INTRODUCTION

High-throughput single-cell RNA sequencing (scRNA-seq) has emerged as a revolutionary approach to dissect cellular compositions and characterize molecular properties of complex tissues,¹ and has been applied to a wide range of fields resulting in profound discoveries.² However, spatial information of individual cells is lost during the process of tissue dissociation. While it is paramount to investigate the molecular composition of individual cells in the spatial contexture, current methods such as RNA hybridization,³ in situ sequencing,⁴ immunohistochemistry,⁵ and purifying predefined subpopulations for subsequent transcriptomic profiling⁶ are limited by the throughput and complex experimental procedures that are only accessible by a handful of laboratories. The combination of scRNA-seq with in situ RNA patterns or tissue shapes provides computational solutions for high-throughput mapping of the spatial locations of many individual cells,^{7–10} but such methods rely on the availability of spatial references, limiting their wide applications. It is of pressing need to develop a novel method to reconstruct cell spatial organizations de novo from scRNA-seq data to further release the great power of such technology.

The spatial organization of individual cells has recently been shown to be self-assembled via ligand-receptor interactions,^{11,12} implying that cellular spatial organization is inherently encoded by their identity. We argue that the spatial relationship of cells may be reconstructed de novo, at least in part, by integrating scRNA-seq data with ligand-receptor interaction information. Here we formulate this hypothesis as a mathematical model, referred to as

CSOmap (Cellular Spatial Organization mapper), and evaluate its performance computationally and experimentally on a diverse scRNA-seq datasets for various human and mouse tissues. All results support that CSOmap not only can reconstruct cell spatial organizations de novo from scRNA-seq data alone, but also can quantify the statistical significance of cell-cell interactions and reveal the potentially critical ligand-receptor pairs mediating such interactions. In particular, CSOmap allows in silico perturbations to evaluate the potential effects of gene overexpression or knock-down and cell adoptive transfer or depletion on the changes of cell spatial organizations. We applied CSOmap to tumor-infiltrating immune cells and gained new insights into the role of regulatory T cells in tumor immunity.

RESULTS

Overview of CSOmap

With the hypothesis that cell spatial organization is inherently encoded by cell identity, we formulate the computation process from scRNA-seq data to cell spatial organization based on three assumptions: (1) the potential of cellular interactions can be approximated by a function of the abundance of interacting ligands and receptors, and their affinity; (2) cells with high interacting potentials tend to locate in close proximity; (3) cells compete for their interacting partners due to physiological and spatial constraints. We formulate these hypotheses as a mathematical optimization model (named as CSOmap) that predicts coordinates of each cell in a three-dimensional pseudo-space

¹Beijing Advanced Innovation Centre for Genomics, Peking-Tsinghua Centre for Life Sciences, Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, Peking University, 100871 Beijing, China

Correspondence: Xianwen Ren (renxwise@pku.edu.cn) or Zemin Zhang (zemin@pku.edu.cn)

These authors contributed equally: Xianwen Ren, Guojie Zhong

Received: 14 February 2020 Accepted: 20 May 2020

Published online: 15 June 2020

based on input scRNA-seq data and known ligand-receptor interactions^{13,14} (Fig. 1). The algorithmic process is composed of two main steps. The first is to estimate the cellular interacting potentials by integrating thousands of ligand-receptor pairs, resulting in a cell-by-cell affinity matrix (Fig. 1a). The second is to embed the inherently high-dimensional affinity matrix into three-dimensional space (Fig. 1b). The limited availability of space determines that it is not feasible to position cells with the same interacting potentials equally close to their partners. Thus, we applied Student's t-distribution to resolve the cell competition problem, enlightened by the widely used visualization technique t-SNE.¹⁵ After embedding cells into three-dimensional space, spatial structures/patterns of cells can be analyzed by density-based clustering¹⁶ (Fig. 1c), connections and corresponding statistical significance among predefined cell types can be summarized (Fig. 1d), and the dominant ligand-receptor pairs underlying a specified pair of cell types can be calculated (Fig. 1e). When a critical gene or cell population was determined, CSOmap can be further applied to examine the effects of *in silico* perturbations including gene overexpression or knockdown and cellular depletion or adoptive transfer (Fig. 1f).

Evaluating the validity of CSOmap based on public scRNA-seq datasets

We first evaluated CSOmap on various publicly available scRNA-seq datasets. In a scRNA-seq dataset of pancreas,¹⁷ the spatial separation of the endocrine and exocrine compartments provides a natural reference for assessing the performance of CSOmap. On both the human and mouse pancreatic scRNA-seq data, CSOmap successfully recapitulated such spatial separation (Fig. 2), with endocrine cells forming one structure and exocrine cells composing the other compartment. The visual separation in human was further supported by random permutation-based statistical testing, with endocrine rather than exocrine cells showing significant interactions with endothelial cells (Fig. 2). CSOmap was then applied to the scRNA-seq data of human placenta and decidua.¹⁸ CSOmap successfully reconstructed the early maternal-fetal interface, i.e., fetal placenta cells, rather than maternal blood cells, showing significant interactions with maternal decidua cells (Supplementary information, Fig. S1). Quantitative evaluation based on the scRNA-seq data of mouse liver lobules⁸ showed that CSOmap reached high consistence ($R = 0.85$, $P < 0.01$, Spearman correlation, Supplementary information, Fig. S1b) with the

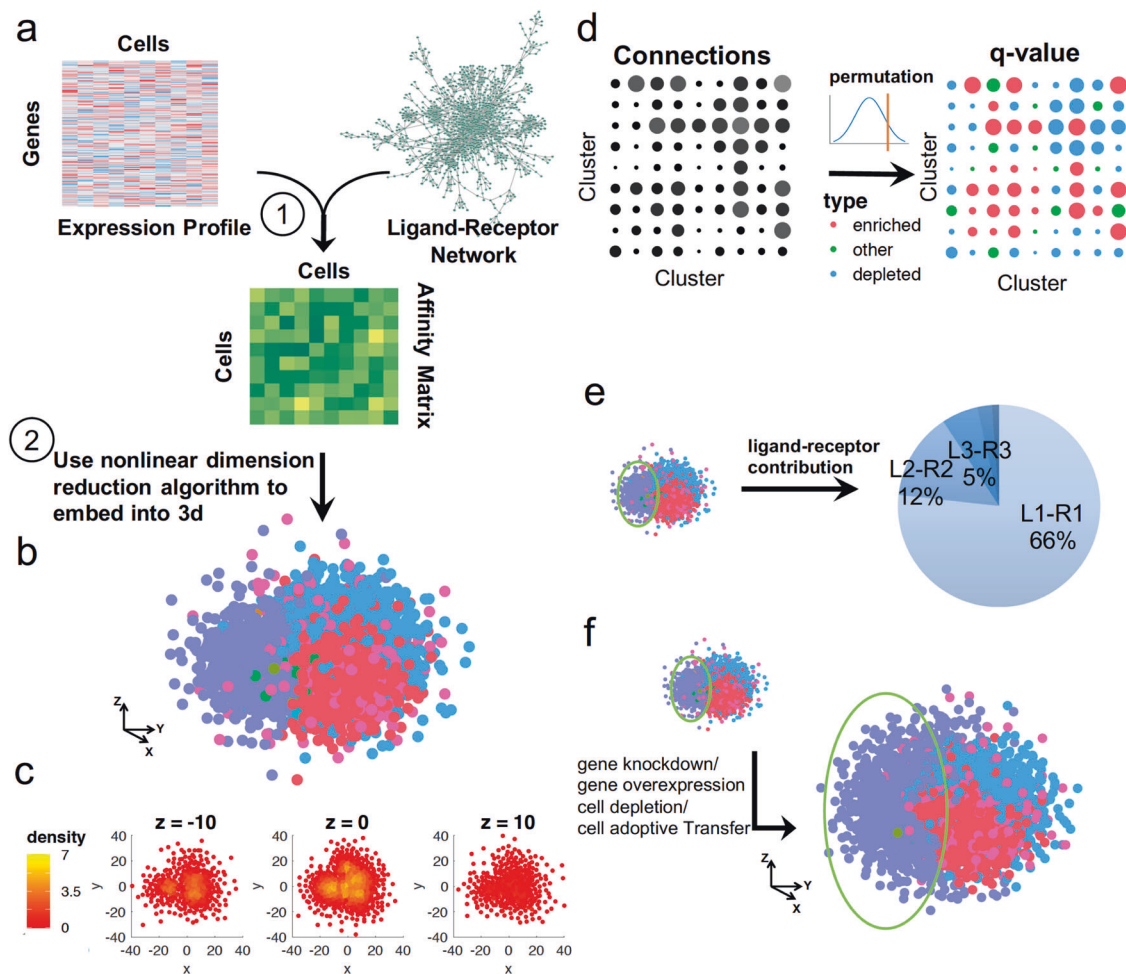


Fig. 1 Schematics of single-cell spatial reconstruction by CSOmap. a CSOmap takes the gene-by-cell expression matrix generated by scRNA-seq and the known ligand-receptor network as inputs, upon which a cell-by-cell affinity matrix is estimated. **b** The inherently high-dimensional cell-by-cell affinity matrix is embedded into a three-dimensional space via resolving cell competitions. **c** Density can be estimated for individual cells based on their three-dimensional coordinates obtained from **b**, which allows the identification of spatially-defined cell clusters. **d** Given the cell cluster labels, the number of connections among cell clusters and their statistical significance can be summarized and evaluated by CSOmap. **e** For a pair of cell clusters, the contributions of each ligand-receptor pair to their interactions can be calculated. **f** CSOmap allows *in silico* interference of the original dataset including gene knockdown/overexpression and cell depletion/adoptive transfer to examine the corresponding effects on cellular spatial organizations.

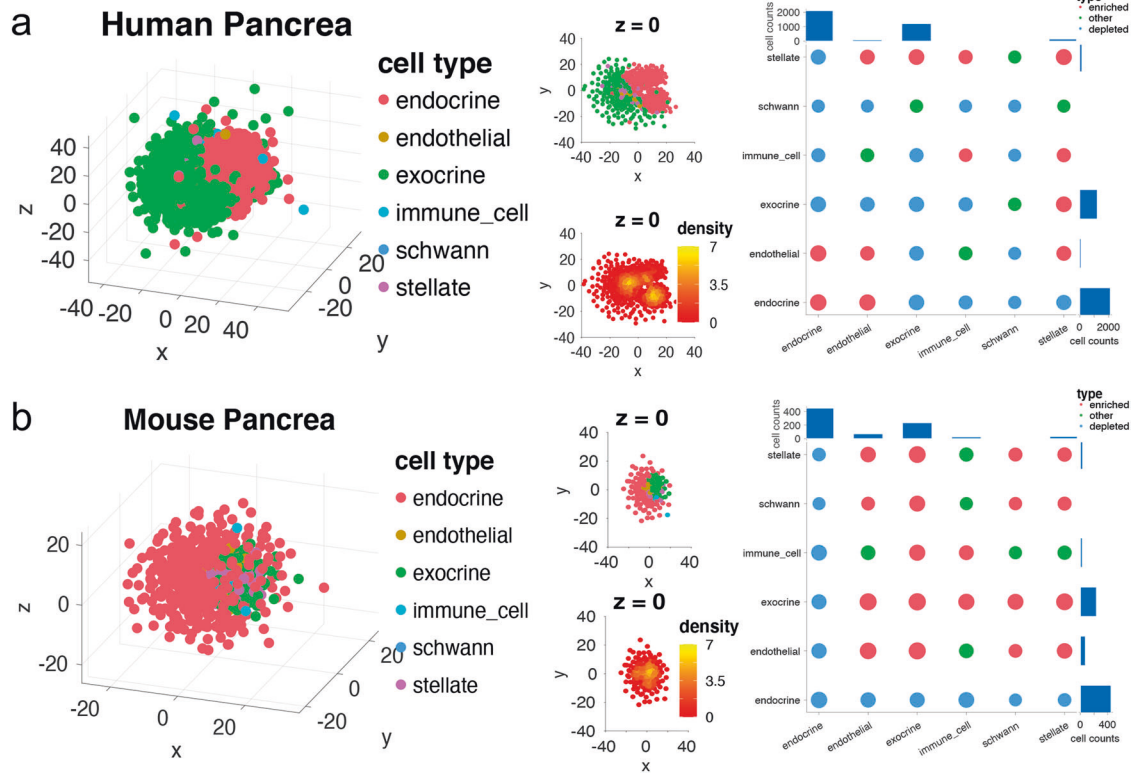


Fig. 2 The exocrine and endocrine compartments of pancreas can be recapitulated by ligand-receptor based inference with CSOmap. **a** The 3D visualization of CSOmap prediction of the human pancreatic scRNA-seq data (left), the cross-section of $z = 0$ of the 3D visualization (middle), and the statistical significance of interactions between different cell types (right). **b** The 3D visualization of CSOmap prediction of the mouse pancreatic scRNA-seq data (left), the cross-section of $z = 0$ of the 3D visualization (middle), and the statistical significance of interactions between different cell types (right). Enriched: cells of one cell type are enriched in the neighborhood of the other cell type, P (right tail) < 0.05 and $q < 0.05$; depleted: cells of one cell type are depleted in the neighborhood of the other cell type, P (left tail) < 0.05 and $q < 0.05$. Exocrine: acinar and ductal cells; endocrine: α , β , γ , δ and ϵ cells.

reference.⁸ Due to the inherent difficulty to dissociate endothelial cells from liver cells, paired-cell sequencing has been customized to resolve the spatial positions of endothelial cells within liver lobules.¹⁹ In silico predictions by CSOmap reached consistent results with paired-cell sequencing ($R = 0.73$, $P < 0.04$, Spearman correlation, Supplementary information, Fig. S1c). Systematic evaluation based on the Tabula Muris datasets²⁰ demonstrated that CSOmap could reproduce the organ-level separations by revealing significantly higher intra-organ cellular interactions than inter-organ interactions for almost all organs except tongue (15/16, 93.75%, Supplementary information, Fig. S2), of which 176 out of 199 interacting cell type pairs were from different cell types. Of 6 organs that have both epithelial and endothelial cells available in the Tabula Muris dataset (Supplementary information, Fig. S3), we observed that in almost all the organs except trachea the epithelial cells occupy the outside space (topologically equivalent to the organ edges) while the endothelial cells occupy the inner space (topologically equivalent to the organ basement), suggesting the spatial resolution below the organ level. Such successful applications clearly show the effectiveness, robustness, and wide applicability of CSOmap for multiple organs from different organisms and different technical platforms.

To further demonstrate the effectiveness of CSOmap to reconstruct the cell spatial organization de novo based on scRNA-seq data, we applied CSOmap to a human scRNA-seq dataset consisting of both normal and fibrotic lungs.²¹ CSOmap was applied individually for each healthy donor and patient with pulmonary fibrosis, and then the spatial characteristics of alveolar

cells were compared among donors and patients. Based on the scRNA-seq data of normal donors, CSOmap revealed that Type II alveolar cells disperse in the outer pseudo-space (topologically equivalent to the alveolar space) and Type I alveolar cells form compact basal structures together with endothelial, alveolar macrophages, and other cells (Fig. 3a). The visual characteristics were further confirmed by quantifying the distance of Type II alveolar cells to the center of the pseudo-space (Fig. 3b). Permutation-based statistical testing suggests that Type II alveolar cells are spatially exclusive to themselves and other cell types (i.e., depleted in the neighborhood of Type II alveolar cells), but Type I alveolar cells show significant interactions with themselves, endothelial cells, and macrophages (i.e., enriched in the neighborhood of Type I alveolar cells). These spatial characteristics agree with the histological observations of human alveoli,²² suggesting the validity of CSOmap.

However, samples from patients with pulmonary fibrosis show distinct spatial characteristics. For idiopathic pulmonary fibrosis (IPF), Type II alveolar cells do not disperse in the outer space but rather show significant interactions with other cells (Fig. 3c, d), consistent with the pathological observation of diffuse alveolar septal thickening and type II pneumocyte hyperplasia.²³ For systemic sclerosis-associated interstitial lung disease (SSc-ILD), although Type II alveolar cells still disperse in the outer space, Type I alveolar cells do not have significant interactions with endothelial/lymphatic cells or macrophages (Fig. 3e, f), agreeing with the pathological characteristics of injured alveolar epithelium.²⁴ By analyzing the dominant ligand-receptor pairs that mediate the spatial organizations of alveolar cells in normal lungs,

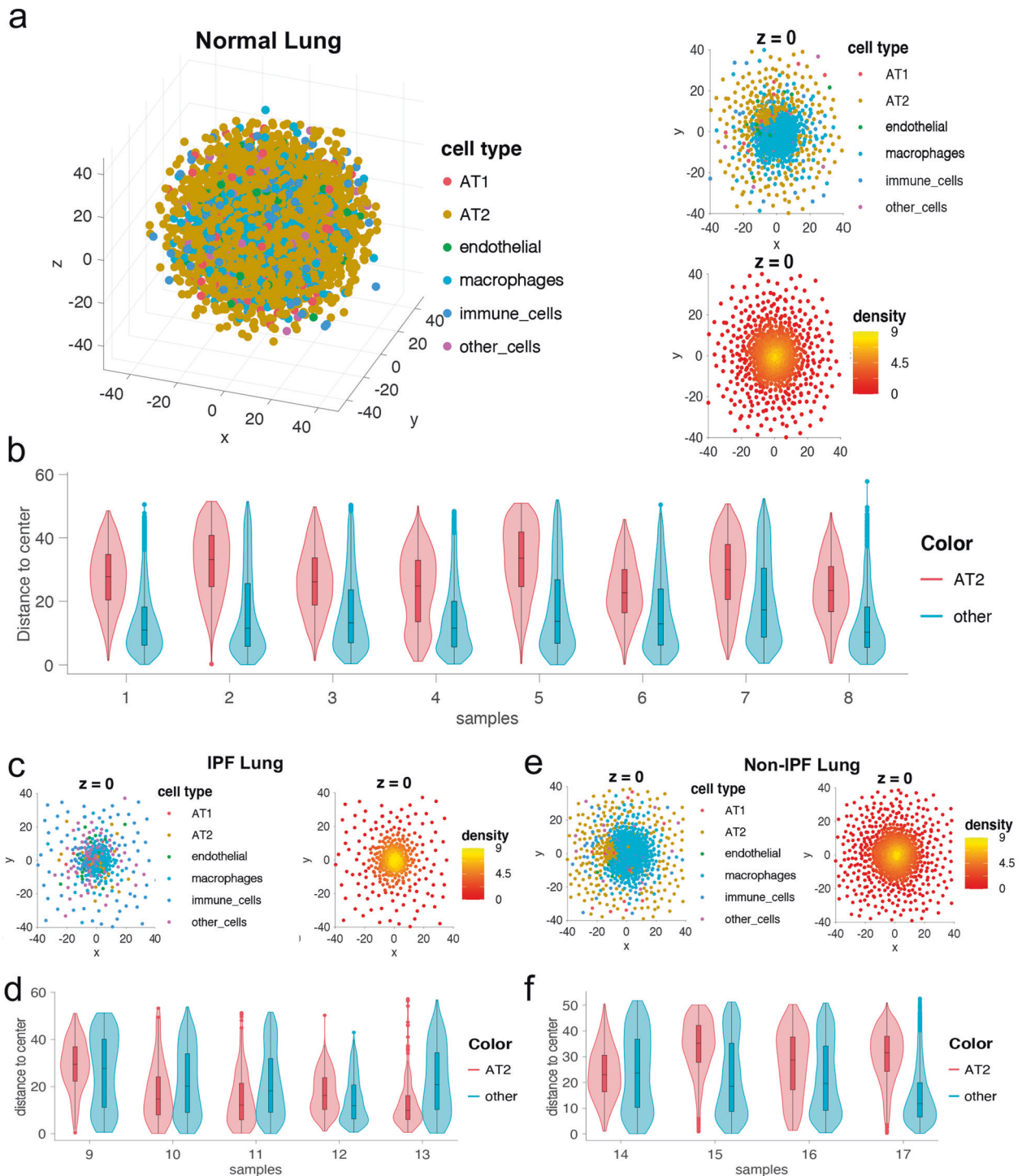


Fig. 3 CSOmap recapitulates the spatial characteristics of normal alveoli of human lungs and the pathological characteristics of pulmonary fibrosis. **a** The spatial organization of normal alveoli in the pseudo-space inferred by CSOmap based on the scRNA-seq data of donor 1. AT2: Type II alveolar cells; AT1: Type I alveolar cells. **b** The distance of AT2 cells to the center of the pseudo-space compared with other cells ($P < 0.05$, rank-sum test). **c** The section view at $z = 0$ of a patient with IPF. **d** Distance of AT2 cells to the center of the pseudo-space compared with other cells. **e** The section view at $z = 0$ of a patient with SSC-ILD. **f** Distance of AT2 cells to the center of the pseudo-space compared with other cells for one patient with hypersensitivity pneumonitis (Sample ID: 14) and three patients with SSC-ILD (Sample IDs: 15–17).

we found that SFTP1-TLR2 was ranked top in mediating interactions among Type I and II alveolar cells. However, for pulmonary fibrosis samples, the scores of SFTP1-TLR2 were significantly reduced (Supplementary information, Fig. S4). Since mutations or aberrant expression of SFTP1 and TLR2 have been

associated with pulmonary fibrosis,^{25,26} the identification of the critical role of SFTP1-TLR2 in maintaining the normal spatial organization of human alveoli by CSOmap via an unbiased approach further underscores the validity of CSOmap and the molecular insights that it may bring.

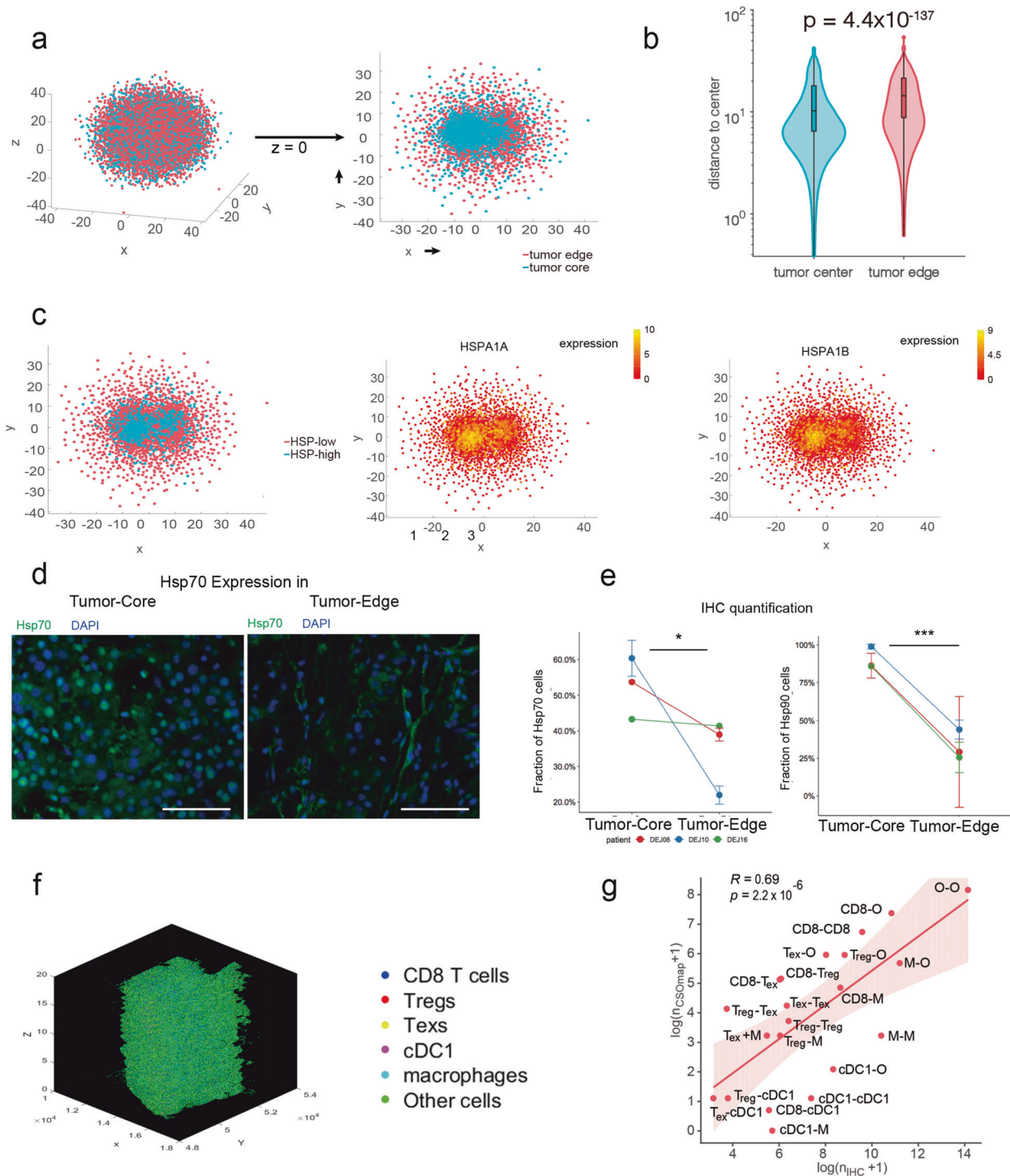


Fig. 4 Performance of CSOmap in reconstructing the spatial organization of a liver tumor sample. **a, b** Tumor core cells tend to locate in the center of the pseudo-space reconstructed by CSOmap. **c** The CSOmap reconstruction revealed that genes encoding HSPs show spatial preference. **d** IHC staining of independent liver tumor samples confirmed the spatial preference of Hsp70. Scale bar, 50 μ m. **e** Quantification based on IHC images confirmed the statistical significance of the spatial preference of Hsp70 and Hsp90 (Student's *t*-test, right tailed, **P* < 0.05; ****P* < 0.01). **f** 3D plot of the tumor sample by stacking 19 IHC images together after manual rotation, in which six major cell types were discriminated by the corresponding markers. **g** Spearman correlation between cell connections based on IHC images (X-axis) and the CSOmap prediction (Y-axis). Treg: regulatory T cells (Foxp3⁺); Tex: exhausted T cell (PD-1⁺); CD8: CD8⁺PD-1⁻ T cells; cDC1: CLEC9A⁺ dendritic cells; M: macrophages (CD68⁺); O: other cells. The median distance of the 3rd nearest neighbor of all cells was used as the cutoff to determine whether two cells were spatially connected or not. The overwhelming number of "other cells" highlights the fact that millions of cells can crowd in a compact piece of tissue, posing great challenges for staining/imaging-based analysis.

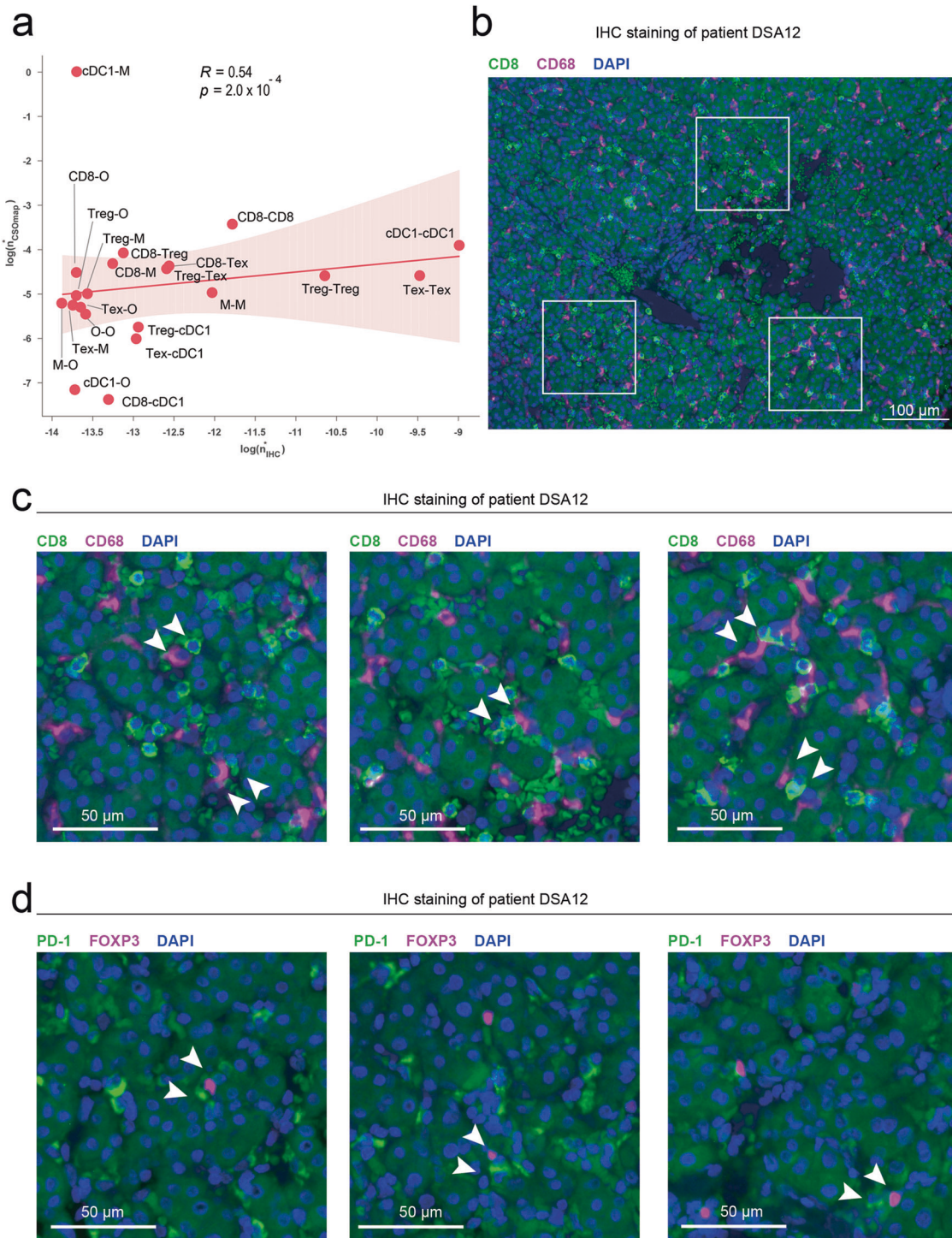


Fig. 5 Consistency between CSOMap prediction and IHC imaging of the tumor sample from HCC. a Spearman correlation between cell connections based on IHC images (X-axis) and the CSOMap prediction (Y-axis) after normalizing the biases introduced by uneven cell counts among different cell types. Treg: regulatory T cell (Foxp3⁺); Tex: exhausted T cell (PD-1⁺); CD8: CD8⁺PD-1⁻ T cells; cDC1: CLEC9A⁺ dendritic cells; M: macrophages (CD68⁺); O: other cells. **b** Example IHC image showing the interaction between CD8⁺PD-1⁻ T cells and macrophages. **c** Enlarged illustrations of the selected windows in **b** (from left to right in order). **d** Example IHC images showing the interaction between Tregs and Tregs.

Evaluating the validity of CSOMap based on experimental tissue dissection and imaging

We further validated the performance of CSOMap with experimental tissue dissection and imaging. First, we dissected a liver tumor sample into tumor edges and cores, and applied scRNA-seq separately. With the scRNA-seq data only, CSOMap

reconstructed a cell spatial organization with tumor core cells located interiorly and tumor edge cells exteriorly (Fig. 4a). The spatial separation was statistically significant by quantitatively evaluating the distance of tumor core and edge cells to the center of the pseudo-space (Fig. 4b). The spatial patterns of genes encoding heat-shock proteins (Hsps; Hsp40, Hsp70 and

Hsp90) revealed by CSOmap, i.e., with high expression level in the center while low expression level near the edge (Fig. 4c), were further experimentally confirmed by immunohistochemical (IHC) staining on independent liver tumor samples (Fig. 4d, e), suggesting the effectiveness and robustness of CSOmap in deriving new biological insights.

We then assessed the CSOmap results in a quantitative manner by simultaneously generating scRNA-seq and IHC staining data for a tumor sample derived from hepatocellular carcinoma (HCC). In brief, six cell types, including regulatory T cells (Tregs, marked by Foxp3⁺), exhausted CD8⁺ T cells (Texs, marked by PD-1⁺), CD8⁺PD-1⁻ T cells, type 1 dendritic cells (cDC1, marked by CLEC9A⁺), macrophages (marked by CD68⁺), and other cells, were labeled by specific antibodies on a 1 cm × 1 cm × 100 μm tumor tissue. This tumor tissue was consecutively sliced into 1 cm × 1 cm × 5 μm pieces for IHC staining, and then the cell types and positions of 1,181,790 cells were recorded to serve as the reference for evaluating the performance of CSOmap (Fig. 4f and Supplementary information, Fig. S5). With 1,329 scRNA-seq profiles based on SMART-seq2,²⁷ CSOmap reached high concordance with the results of IHC analysis (Fig. 4g, $R = 0.69$, $P = 2.2 \times 10^{-6}$, Spearman correlation) and recapitulated multiple cell-cell interactions exemplified by CD8 T cells-macrophages and Tregs-Texs pairing (Fig. 5). After removing the potential biases introduced by the uneven cell counts of different cell types, the consistency score between IHC results and the CSOmap prediction was still 0.54 ($P = 2.0 \times 10^{-4}$, Spearman correlation, Fig. 5) while the correlations based on random coordinate assignment and random gene pairs were -0.12 and 0.34 , respectively.

CSOmap reveals the critical role of CD63-TIMP1 interaction in tumor morphology

Besides the spatial reconstruction, CSOmap can also provide important insights into the underlying molecular mechanisms. We applied CSOmap separately to a head and neck cancer (HNC) scRNA-seq dataset²⁸ and a melanoma scRNA-seq dataset.²⁹ Based on the IHC images of the original report,²⁸ the spatial characteristics of HNC tumor microenvironment can be summarized as follows: (1) malignant cells not subject to partial epithelial to mesenchymal transition (p-EMT) were located close to each other and formed a loose structure; (2) malignant cells subject to p-EMT were located at the interface between malignant cells and cancer-associated fibroblasts (CAFs); (3) CAFs were connected to each other and formed a compact structure (Fig. 6a). CSOmap not only qualitatively recapitulated all these IHC characteristics (Fig. 6b), but also highlighted the distinct spatial patterns of malignant cells between HNC and melanoma, i.e., malignant cells in HNC tended to form a loose structure (adjusted $P > 0.05$, permutation-based test, Fig. 6c, d) while tumor cells in melanoma tended to form a compact structure (Fig. 6e, adjusted $P < 0.05$, permutation-based test). By analyzing the dominant ligand-receptor pairs contributing to this spatial organization, we identified that the interactions between CD63 and TIMP1 contributed ~66% to the cellular interaction potential of melanoma malignant cells (Fig. 6f) while HNC cells expressed CD63 and TIMP1 at much lower levels. Using CSOmap, we were able to readily perform “in silico perturbation” of CD63 and recalculate the spatial characteristics. Indeed, in silico knockdown of CD63 expression in melanoma malignant cells resulted in the transition from compact to loose structures while overexpression of CD63 in HNC malignant cells resulted in compact structure (Fig. 6g). The association of CD63 with the morphology of melanoma has been experimentally supported by a previous in vivo and in vitro study,³⁰ in which the mechanism underlying such association was attributed to the negative linkage between CD63 signaling and EMT. This notion is recapitulated by CSOmap since the p-EMT program was observed in the HNC dataset but absent in melanoma, suggesting the effectiveness of CSOmap in

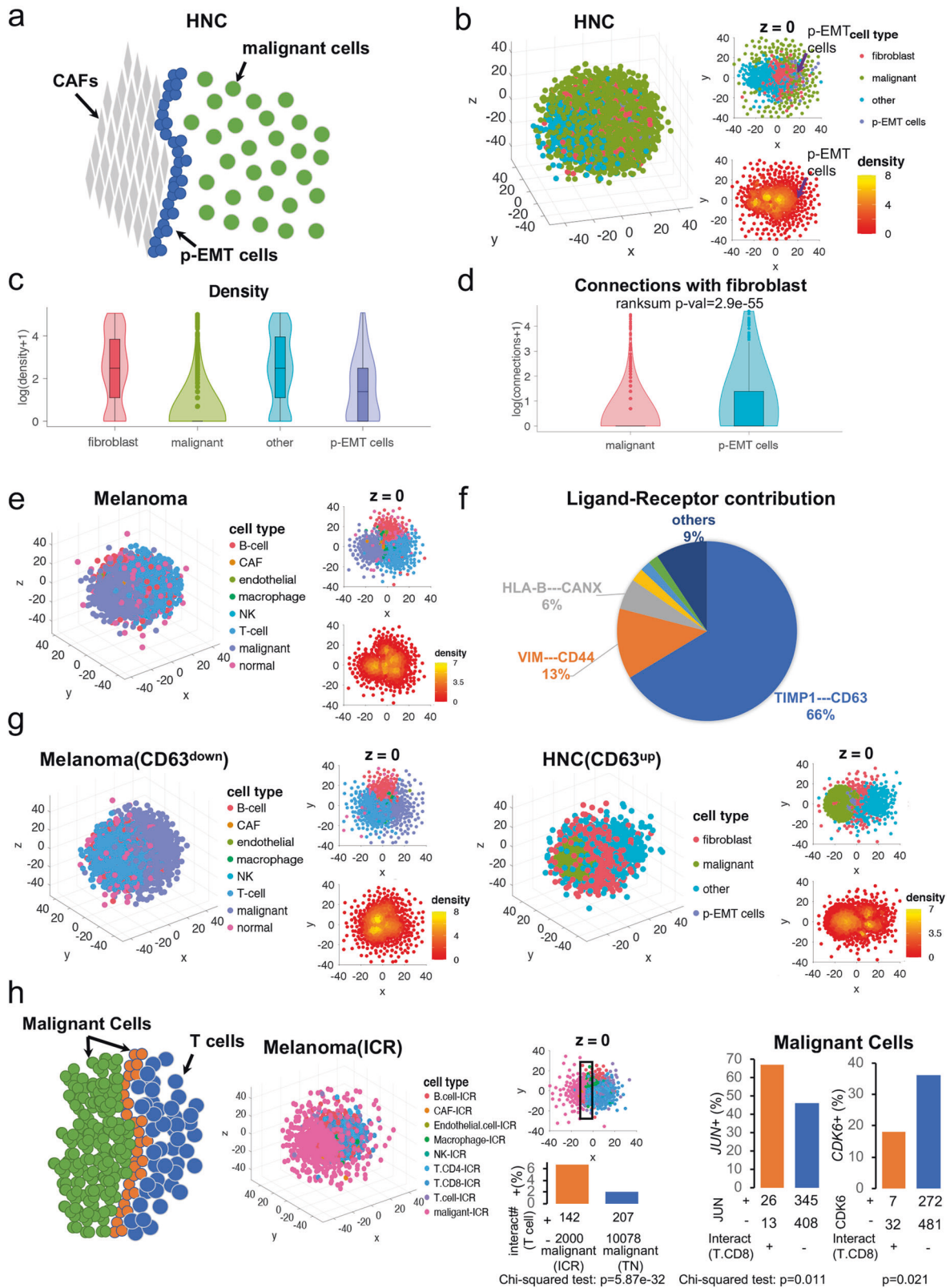
spatial reconstruction and the potential in revealing the underlying molecular mechanism.

Further biological insights were also provided by the spatial reconstruction of CSOmap based on the melanoma and HNC scRNA-seq data. First, the malignant cells of both melanoma and HNC did not show significant interactions with T cells according to our CSOmap analyses, which may partially explain the immune evasion of these tumors. By contrast, while the CAFs, endothelial cells and macrophages account for much smaller fraction of the datasets, they showed statistically significant interactions with almost all the other cell types. These results recapitulate the critical roles of these cells in the spatial organization of tumors and in the regulation of tumor-infiltrating lymphocytes.³¹ Second, comparison of the spatial organizations between melanoma samples that are treatment-naïve (TN)²⁹ and on immunotherapy³² highlights the tumor and T cell compartments observed by IHC (Fig. 6h). Upon treatment, increased tumor-T interactions were observed by IHC (also revealed by CSOmap), indicating the potential effects of immunotherapy. Differential gene expression analysis based on the CSOmap prediction indicated that malignant cells not interacting with T cells show lower levels of class I major histocompatibility complex (MHC) molecules and JUN but higher level of CDK6. These results recapitulate the cancer cell program contributing to resistance of immune checkpoint blockade in melanoma identified recently,³² suggesting the effectiveness of CSOmap in generating valid biological insights.

CSOmap provides new insights into the roles of regulatory T cells in tumor immunity

Since CSOmap also allows in silico cellular perturbation, we applied CSOmap to three scRNA-seq datasets based on T cells from the peripheral blood, tumors and tumor-adjacent normal tissues of patients with HCC,³³ non-small cell lung cancer (NSCLC),³⁴ or colorectal cancer (CRC).³⁵ Since blood, tumor and normal tissues have distinct morphologies, and tertiary lymphoid structures (TLSs) are frequently found in tumors,³⁶ we hypothesize that T cells infiltrating into different tissues may also demonstrate distinct spatial organization characteristics. CSOmap analyses on all three datasets suggest that T cells from tumors tend to have significantly more interactions with themselves while T cells from the peripheral blood tend to disperse from each other (Fig. 7a), confirming our hypothesis. Cellular density analysis clearly indicated the existence of tightly-linked structures (Fig. 7b), with tumor-derived T cells forming the major part of these structures (Fig. 7c). It has been reported that Tregs tend to trap tumor-infiltrating CD8⁺ T cells into TLSs or draining lymph nodes.³⁷ Consistently, our analysis indicated that Tregs and tumor-infiltrating Texs are the major parts of these compact structures (Fig. 7d), which we speculate to correspond to TLSs in or near tumors. Interestingly, although blood-derived T cells did not show significant interacting potential to each other compared to tumor-derived T cells, those compact structures composed of blood-derived T cells were also observed in the HCC and NSCLC datasets, supporting the colony-forming capacity of T lymphocytes from peripheral blood as reported previously.³⁸

Among T cells, Tregs and Texs exhibited significant interaction, and the ligand-receptor pair CCL4-CCR8 drove such interaction (Fig. 7e). While CCL4 was highly expressed in most activated CD8⁺ T cells, its expression level in Texs was two-fold of that in other cells. CCR8 was specifically and highly expressed in tumor Tregs. According to the spatial organization reconstructed by CSOmap, Texs could be further divided into two subgroups: Texs interacting and not interacting with Tregs. Consistent with a previous report,³⁹ MKi67 was depleted in Texs interacting with Tregs (Fig. 7f), suggesting reduced proliferation by Tregs. Since Texs are characterized by high expression of T cell exhaustion markers including PDCD1, CTLA4, HAVCR2, TIGIT and LAG3, we examined the expression levels of their ligands in Tregs. CD274⁺ (or PDL1,



gene for the PDCD1 ligand) and $CD80^+$ (gene for CTLA-1 ligand) Tregs were significantly enriched in Tregs interacting with Tregs while $CD86^+$ Tregs were enriched in Tregs not interacting with Tregs in CRC (Fig. 7g). These results suggest that Tregs might suppress $CD8^+$ T cells via PD-1 and CTLA-4-mediated co-inhibitory

axes. Similar trends were found in HCC and NSCLC despite varying significance. In addition to Tregs, Tregs also showed significant interactions with a set of $CD8^+$ effector memory T cells (Tems). In CRC, T cell receptor (TCR)-based tracking suggests frequent state transitions between Tems and Tregs in tumor.³⁵ The ratio of Tems

Fig. 6 CSOmap reveals CD63-TIMP1 as a critical ligand-receptor pair in determining the spatial characteristics of HNC and melanoma malignant cells. **a** Cartoon illustration of spatial characteristics observed in IHC images of HNC patients.²⁸ **b** Global (left) and cross-section (right) views of reconstructed spatial organization of HNC cells. **c** Compactness of different cell classes of HNC estimated by the density of each cell. **d** P-EMT cells showed significantly higher interactions with fibroblasts than other malignant cells. **e** Global (left) and cross-section (right) views of reconstructed spatial organization of melanoma cells. **f** Contributions of ligand-receptor pairs to the interactions among melanoma malignant cells. **g** Spatial characteristics of melanoma (left) and HNC (right) after altering the expression levels of CD63. **h** Cartoon illustration of spatial characteristics of melanoma on immunotherapies observed by IHC staining (left), reconstructed spatial organization (middle), comparison of interactions with T cells between treatment-naïve (TN) melanomas and melanomas with resistance to immune checkpoint blockade (ICR) (bottom right in the middle), and differential usage of JUN and CDK6 between T cell-interacting and not interacting malignant cells in the ICR dataset.

to Tregs in tumor has also been associated with better survival in lung adenocarcinoma patients³⁴ and better response to immunotherapies in melanoma recently.⁴⁰ The notable interaction between Tregs and Tems in tumor may suggest a role of Tregs in the early stage of immune evasion of tumors. Similar to Tregs, Tems interacting with Tregs showed higher expression levels of *IFNG* and *TNF* than those not interacting, and the expressions of *IFNG* and *TNF* showed significant correlations with *CCL4* in Tregs/Tems interacting with Tregs rather than those not interacting, suggesting that functional CD8⁺ T cells were prone to be targeted by Tregs due to high level of *CCL4* secretion. IHC analysis of HCC and CRC samples confirmed the interactions of Tregs with Tregs and Tems based on the colocalization of Tregs and CD8⁺ T cells in tumor (Fig. 5 and Supplementary information, Fig. S6).

In silico Treg depletion by CSOmap revealed that a subset of blood-enriched recently activated effector memory T cells (Temras) demonstrate significantly increased interactions with Tregs via the CXCR3-CCL5 axis (Supplementary information, Fig. S7), which is different from the CCR8-CCL4 axis mediating the Treg-Tem interactions. It has been recently reported in murine and human melanoma that, compared with CCR2 and CCR5, CXCR3 is necessary for the successful trafficking of tumoricidal T cells across tumor vascular checkpoints,⁴¹ consistent with our finding that CXCR3 might mediate the migration of blood T cells into tumor via the CCL5 gradient. Treg depletion also increased the interactions of CD4⁺CXCR6⁺ tissue-resident helper T cells (Ths) with Tregs via the CXCR3-CCL5 axis (Supplementary information, Fig. S7), supporting the role of T cell competition in immune regulation as revealed previously.⁴²

In silico cell transfer reveals phenotypic determinants in T cell-based tumor cell killing

CSOmap also enables computational simulation of adoptive cell transfer, which has proven to be effective immunotherapy for cancer treatment.⁴³ It is currently difficult to experimentally evaluate the phenotypic outcome of adoptively transferred T cells. We used the HNC and melanoma datasets as foundations to simulate their tumor microenvironments and used blood- and tumor-derived T cells for in silico adoptive transfer. We simulated a gradient of TCR-pMHC affinity between the adoptively transferred T cells and the malignant cells and quantified the numbers of tumor-T interactions, tumor-infiltrating T cells and targeted malignant cells. Interestingly, while the numbers of interactions between T and malignant cells increased in a linear function of the TCR-pMHC affinity, those infiltrating T cells and targeted malignant cells increased in a logarithmic function (Fig. 8). Visually, an interface formed between T cells and malignant cells (Fig. 8). This phenomenon observed in silico might recapitulate and explain the morphological patterns frequently observed in tumor microenvironment by IHC and multiplexed ion beam imaging.⁴⁴ While the TCR-pMHC affinity was the dominant determinant of T-malignant cell interactions, the phenotypes of T cells and malignant cells also contributed significantly (Fig. 8 and Supplementary information, Fig. S8). In particular, tumor-derived T cells showed significantly higher efficiency in tumor infiltration than blood-derived T cells (Fig. 8) while malignant cells of melanoma

were more prone to be targeted than those of HNC (Supplementary information, Fig. S8) according to ANOVA analysis with repeated measures. Computationally, these results recapitulated the variations of T cell transfer-based therapies across cancers and the predictive values of immunophenotypic characterization of infused T cell product in engraftment and responses observed in various clinical trials.⁴⁵

Robustness of CSOmap to its technical parameters

We used the HNC and melanoma datasets to show the robustness of CSOmap to its parameter selections. First, we show that it is necessary to embed the cell-by-cell affinity matrix estimated by ligand-receptor interactions into 3D space. As shown by the HNC dataset (Supplementary information, Fig. S9), eight cluster pairs showed different statistical significance before and after 3D embedding (Supplementary information, Fig. S9a), and the changes before and after 3D embedding are nonlinear (Supplementary information, Fig. S9b), with the spatial configuration inferred with 3D embedding showing significantly higher consistency with the IHC images (Supplementary information, Fig. S9c). This result suggests that spatial conflicts may play key roles in shaping cell spatial organization, which is further confirmed by the low correlations between 2D and 3D embeddings (Supplementary information, Fig. S10). Determining the neighborhood of each cell is challenging. However, we show that our results are robust to the selection of the number of cells in the neighborhood of a cell. Using the median distance of the 3rd and 5th nearest neighbors as the cutoff to determine the neighborhood of each cell, CSOmap generates highly consistent cell-cell interaction maps ($R > 0.99$, $P < 10^{-80}$, Spearman correlation, Supplementary information, Fig. S11). We also evaluated the impacts of the comprehensiveness of ligand-receptor pairs on spatial inference by including the data in CellPhoneDB.¹⁸ The results suggest that the cell spatial organizations inferred by CSOmap are highly reproducible with or without ligand-receptor pairs in CellPhoneDB ($R > 0.97$, $P < 10^{-50}$, Spearman correlation, Supplementary information, Fig. S12), suggesting that the current ligand-receptor pairs may be adequate for cell spatial organization inference.

DISCUSSION

CSOmap provides a computational tool to reconstruct cellular spatial organization de novo from scRNA-seq data. The underlying assumption is that cells can compete and self-assemble into specific spatial patterns via ligand-receptor interactions. A wide collection of factors may hinder such computational prediction, including the incomplete nature of known ligand-receptor interactions and their affinity parameters (particularly for TCR-pMHCs), the dropout issue of scRNA-seq, biases or errors of estimating the protein abundance of ligands and receptors by transcriptomic data, the distinction of capacity to infer realistic interactions, and the unavailability of other physical, chemical, and nutrient factors involved in cell organizations. Despite these difficulties, evaluations on multiple scRNA-seq datasets spanning human and mouse physiological and diseased conditions

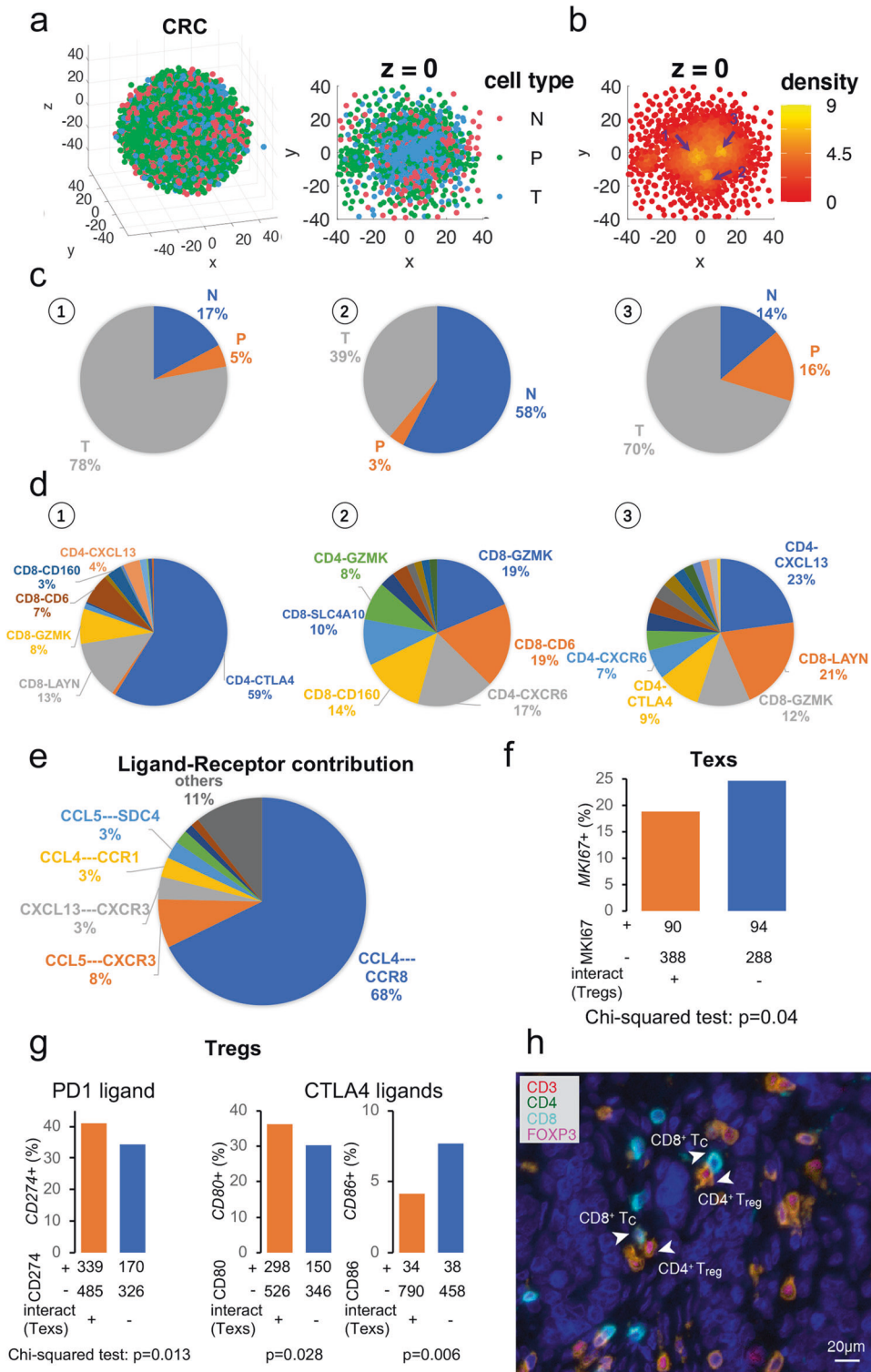


Fig. 7 Spatial and molecular characteristics of CRC T cells. **a** Global (left) and cross-section (right) views of the reconstructed spatial organization of CRC T cells (colored by the tissue origins, N: normal tissue; P: peripheral blood; T: tumor). **b** Compactness of individual cells estimated by density, with 1, 2, and 3 indicating three observed compact structures. **c** Tissue compositions of the three observed compact structures in **b**. **d** Cluster compositions of the three observed compact structures in **b**, where the cluster names follow the nomenclature of the original paper with CD4-CTLA4 indicating tumor Tregs, CD8-LAYN indicating Tregs, and CD8-GZMK indicating Tregs. **e** Contributions of ligand-receptor pairs to the interactions between Tregs and Tregs. **f** Tregs interacting with Tregs showed depletion of MKi67⁺ cells. **g** Tregs interacting with Tregs showed enrichment of CD274⁺ and CD80⁺ cells and depletion of CD86⁺ cells. **h** Co-localization of Tregs and CD8⁺ T cells in CRC tumor samples revealed by IHC staining.

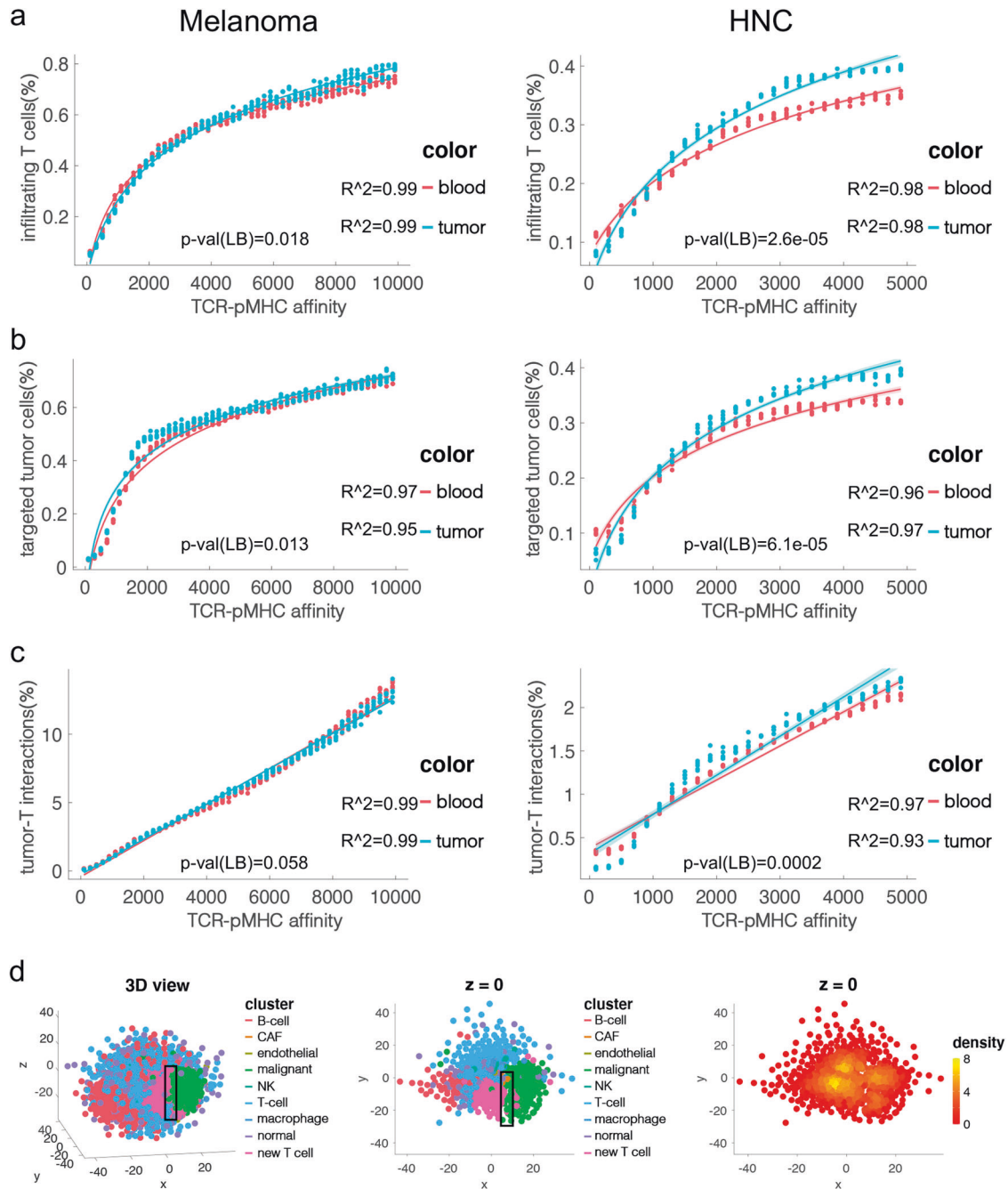


Fig. 8 Tumor-T interactions with varying TCR-pMHC affinity after adoptive T cell transfer. **a** The percentage of infiltrating T cells of the melanoma (left) and HNC (right) datasets. “Blood” and “tumor” indicate the tissue origins of the adoptively transferred T cells. **b** The percentage of targeted tumor cells of the melanoma (left) and HNC (right) datasets. **c** The percentage of tumor-T interactions relative to the theoretical numbers of the melanoma (left) and HNC (right) datasets. The R^2 values in **a** and **b** indicated the goodness of fitting a logarithmic function to the observed values of each series. The R^2 values in **c** indicated the goodness of fitting a linear function to the observed values of each series. The statistical significance of TCR-pMHC affinity and the tissue origins of adoptively transferred T cells to the percentages of infiltrating T cells, targeted tumor cells, and tumor-T interactions was evaluated by ANOVA analysis with repeated measures (by the ranova function of Matlab R2016b), and the lower bound (LB) of the P -values of the tissue origin was displayed. **d** The different dynamics of tumor-T interactions from infiltrating T cells and targeted tumor cells explain the visual patterns of tumor-T interfaces and tumor and T compartments frequently observed in IHC and multiplexed ion beam images. New T cell: adoptively transferred T cells.

demonstrated that CSomap can recapitulate critical spatial characteristics qualitatively and quantitatively. Such validity of CSomap in reconstructing cell spatial organizations de novo from scRNA-seq data supports the hypothesis that cell morphology may be inherently encoded by cell identity, and suggests that

ligand-receptor mediated cellular self-assembly may play key roles in tissue morphogenesis.

Compared with the recently proposed method Novosparc,¹⁰ which computationally maps single cells into predefined tissue shapes based on the assumption that similar single cells

have similar spatial locations, CSOmap makes four conceptual advances. First, CSOmap is a de novo spatial reconstruction method. In contrast, Novosparc is reference-based although the reference is a predefined geometric shape. For scRNA-seq datasets without available tissue shapes, only de novo inferring tools can be used to reconstruct cell spatial organizations in a pseudo-space. Second, CSOmap is built on the assumption that ligand-receptor interactions mediate cell self-assembly. Based on this assumption, it becomes feasible to reconstruct cell spatial organizations de novo, and the applications of CSOmap to almost all the evaluated instances suggest that similar single cells often, but not always (due to spatial conflicts), have similar spatial locations. However, the assumption behind Novosparc that similar single cells have similar spatial locations cannot indicate the roles of ligand-receptor interactions in cell morphogenesis and spatial inference. Third, CSOmap enables the evaluation of statistical significance of cell-cell interactions and the roles of individual ligand-receptor pairs in dictating such interactions while Novosparc does not provide such mechanistic insights. Finally, CSOmap allows in silico simulations of gene/cell perturbations to evaluate the roles of specific ligand/receptors and cell types in shaping/re-shaping a targeted tissue due to its nature of de novo inference. However, with predefined tissue shapes, it is hard for Novosparc to evaluate the roles of such changes in tissue shapes. This feature is important especially when perturbation experiments are hard to conduct.

In summary, CSOmap is able to generate profound hypotheses into the molecular mechanisms underlying cell spatial formation by using its several key features: the de novo reconstruction nature of CSOmap for cell spatial organization, the quantification ability for the statistical significance of cell-cell interactions and the roles of individual ligand-receptor pairs in shaping such interactions, and the convenience of in silico manipulation including gene overexpression, knockdown, cell adoptive transfer and depletion. Such computational modeling can provide important insights into various biological questions including development, immune response, and tumor immune escape. CSOmap will be applicable to interrogation of cellular organizations in pseudo-space from scRNA-seq data for various tissues in diverse systems, and it can be greatly enhanced when more complete knowledge of ligand-receptor interactions and other critical factors are available.

MATERIAL AND METHODS

Overview of CSOmap

CSOmap reconstructs cellular spatial organization of individual cells from scRNA-seq data based on three principles: (1) cellular spatial organization is determined by ligand-receptor mediated cellular self-assembly, with cells having high affinity spatially close to each other; (2) the affinity of cells can be defined by the abundance of ligands and receptors and their interacting potentials; (3) cells compete with each other to form spatial structures. Hence, the core algorithm of CSOmap to reconstruct spatial organization of single cells from RNA-seq data includes two steps: (1) estimating the cellular affinity matrix based on the gene expression profiles of individual cells and known ligand-receptor interactions; (2) embedding the inherently high-dimensional cellular affinity matrix into three-dimensional pseudo-space resembling the realistic biological tissues during which cell competitions are sufficiently considered. CSOmap also includes additional algorithms for analyzing the resultant three-dimensional coordinates of single cells, including estimating the density of each cell, identifying spatially-defined cell clusters/structures, evaluating the number of connections and statistical significance between two cell clusters defined by expression profiles or other characteristics, calculating the contributions of each ligand-receptor pair to the interaction potential of two cell

clusters, and in silico molecular and cellular interference. The details of the core and additional algorithms of CSOmap are depicted as follows separately.

Estimating cell-cell affinity by ligand-receptor interactions

The estimation of cell-cell affinity is critical to the performance of CSOmap. To define a valid function of cell-cell affinity, we assume that the affinity of two cells equals to the affinity summation of all the protein complexes formed by the proteins from the surfaces or extracellular matrices of two cells. We applied a series of approximations to facilitate computation at the genome scale. Details are stated as follows.

In biological reality, the number of components of a protein complex varies from two to tens. For computational convenience, we converted all the interactions of more than two components to binary interactions regardless of the complicated nonlinear effects. Given a binary interaction, i.e., one ligand A and one receptor B, according to the law of mass action in chemistry, the concentration of the complex AB can be calculated according to the following formula:

$$[AB] = k[A]^a[B]^b \quad (1)$$

where $[AB]$ is the concentration of the complex AB, $[A]$ is the concentration of the ligand A, $[B]$ is the concentration of the receptor B, k is the reaction constant, and a and b are the stoichiometric coefficients of A and B, respectively. The parameters k , a and b vary according to the chemical natures of A and B. For similarity, we approximate Formula (1) by the following formula to handle thousands of pairs of ligand and receptor:

$$[AB] \propto w_{A,B}[A][B] \quad (2)$$

where $w_{A,B}$ is introduced to summarize the total effects of the parameters k , a and b . Upon this approximation, the cell-cell affinity is defined by the following formula:

$$\begin{aligned} A_{C1,C2} &\propto \sum_{i=1}^l ([A_{C1}B_{C2}] + [A_{C2}B_{C1}]) \\ &\propto \sum_{i=1}^l w_{A,B}([A_{C1}][B_{C2}] + [A_{C2}][B_{C1}]) \end{aligned} \quad (3)$$

where $A_{C1,C2}$ denotes the affinity of Cell C1 and Cell C2, $[A_c]$ or $[B_c]$ denotes the concentration of the A or B molecule on Cell c , i is the index of ligand-receptor pairs, and there are a total of l pairs. Because the ligand and receptor can be simultaneously expressed by both of the cells, a symmetric term of the concentrations of the complex is added. Furthermore, we use the mRNA abundance of the ligand and receptor to approximate their protein concentrations, and thus Formula (3) can be updated as follows:

$$A_{C1,C2} \propto \sum_{i=1}^l w_{A,B}(A_{C1}^{TPM} \times B_{C2}^{TPM} + A_{C2}^{TPM} \times B_{C1}^{TPM}) \quad (4)$$

where A_c^{TPM} or B_c^{TPM} is the mRNA level of the ligand A or receptor B in Cell c estimated by the Transcripts Per Million (TPM) measure. Since there are thousands of ligand-receptor interactions, for which most of the parameters of their interacting dynamics (summarized by $w_{A,B}$) are not available, we set $w_{A,B} = 1$ in the current version of CSOmap while providing $w_{A,B}$ as a parameter of the software for incorporating users' knowledge of the chemical natures of the ligand-receptor interactions. According to (4), the computational method is thus established for estimation of cell-cell affinity based on scRNA-seq data and ligand-receptor interactions. In practice, we used the human ligand-receptor interaction database FANTOM5 with incorporation of immune-relevant chemokines, cytokines, co-stimulators, co-inhibitors and their receptors for estimating the cell-cell affinity matrix^{14,46} (Supplementary information, Table S1). Some

of these ligands such as chemokines and cytokines are not membrane-located but secreted proteins. We included these ligands into the estimation of cell-cell affinity similar to those membrane proteins because they often form gradients to affect the migration of other cells, particularly for chemokines. Interactions involved in B2M were manually filtered because of its housekeeping nature. Ligand-receptor interactions in the CellphoneDB¹⁸ database were also included to show the robustness of CSOmap predictions (Supplementary information, Table S2). Because of the potential noises introduced in the estimation of cell-cell affinity due to noises in gene expression levels and various approximations, we further discretize the cell-by-cell affinity matrix by retaining the top k highest-affinity neighbors for each cell to reduce noise ($k = 50$ by default).

Embedding the high-dimensional cell-cell affinity matrix into three-dimensional space

When the discretized cell-by-cell affinity matrix is obtained, this inherently high-dimensional matrix is further embedded into a three-dimensional space. The principle behind this operation is that the realistic biological space is of only three dimensions and that positive affinity values in the affinity matrix only indicate the interacting potentials rather than realistically occurred facts. Cells having interacting potentials with common targets need to compete the space to change potentials to reality. Considering this factor, we introduce three constraints to build the computational model. First, a minimum distance between cells is pre-defined because all cells have positive sizes and cannot be ultimately squeezed. Second, the total available space is also pre-defined by a parameter named as space radius to simulate the limited realistic space. Finally and most importantly, Student's t -distribution is introduced to resolve the crowding issues of cell-cell interactions, motivated by the visualization algorithm t -SNE, which allows cells to compete with others to form the spatial organization. Taken all these considerations together, we propose the following computational model as the core of CSOmap:

$$\min \sum_{i=1}^n \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (5)$$

subject to:

$$p_{ij} = \frac{1}{Z} \sum_{k=1}^K w_{L_k, R_k} (e_i^{L_k} \times e_j^{R_k} + e_i^{R_k} \times e_j^{L_k}) \text{ for } i \neq j \quad (6)$$

$$q_{ij} = \frac{1}{Z} \frac{1}{1 + d_{ij}^2} \text{ for } i \neq j \quad (7)$$

$$d_{ij}^2 = \sqrt{\sum_{k=1}^3 (y_i^k - y_j^k)^2} \text{ for } i \neq j \quad (8)$$

$$d_{ij} \geq r \text{ for } i \neq j \quad (9)$$

$$|y_i^k| \leq R \text{ for } i = 1 \dots n \text{ and } k = 1, 2, 3 \quad (10)$$

where $e_i^{L_k}$ or $e_i^{R_k}$ is the TPM values of the k th ligand or receptor in the i th cell, w_{L_k, R_k} is the weight summarizing the chemical nature of the k th pair of ligand-receptor, p_{ij} is the cell-cell affinity between i and j estimated by the aforementioned method, y_i^k is the k th coordinate of the i th cell, d_{ij} is the Euclidean distance between the i th and j th cells in the embedded space, and q_{ij} is the probability of the j th cell locating in the neighborhood of the i th cell. Constraints (6)–(8) give out the definitions of p_{ij} , q_{ij} , and d_{ij} while constraints (9) and (10) impose the spatial

limitations. Kullback-Leibler divergence is used to define the loss function (5). This optimization model is highly similar to the model used to implementing non-linear dimensional reduction in the frequently used visualization algorithm t -SNE except that constraints (9) and (10) are imposed to consider the spatial limitations. Similarly, a gradient-descent algorithm is used to solve this optimization problem with random initialization and then by updating the solution with the guidance of the gradient. When the maximum number of iterations was reached, the resultant three-dimensional coordinates were reported for subsequent analyses. In principle, large r and small R will reduce the volume of cell space and thus provide repulsive forces while the cell-cell affinity provides attracting forces. The repulsive and attracting forces together guide the self-assembly of cells and finally determine the cellular spatial organizations. Particularly, large r and small R will introduce fluctuations into the cellular spatial organizations. To obtain a stable organization, we set $r = 1$ and $R = 50$ in practice. The initializing solution is randomly assigned in a $50 \times 50 \times 50$ cube, and the maximum number of iterations is set to 1000. When the three-dimensional coordinates are obtained, a rotation of the coordination system is made by principal component analysis to guarantee the X and Y axes to capture most spatial variations. By default, we set the number of dimensions as 3 because biological tissues/organs are in 3D space, but the users can tune this parameter to 2 or 1 to examine specific tissue models.

Density analysis and clustering spatially compact cell clusters

Given the three-dimensional coordinates of all cells, a straightforward analysis is to check what spatial structures are formed, which can be examined visually and quantitatively. CSOmap implements a series of visualization method to facilitate the recognition of spatially organized structures, including global three-dimensional views with various rotation angles, cross-section views with various slicing methods, and even dynamic views to show how cells self-assemble into organizations via ligand-receptor mediated interactions. Categorical or numerical features can be used to color cells to highlight the patterns. Quantitatively, the compactness of the neighborhood of a cell, named as density, can be calculated by counting the number of cells within a predefined radius. By default, the radius is set to the median distance of each cell to its third nearest neighbor because the number of neighbors of a cell cannot be too large due to limited space. When the density of individual cells are defined, clustering based on fast searching and finding density peaks¹⁶ can then be applied to identify spatially compact cell clusters and dissociative cells. Sensitivity analysis suggested that the identification of compact structures is robust to the selection of the radius.

Evaluating the statistical significance of cellular interactions between cell clusters

The resultant three-dimensional coordinates of CSOmap also allow us to examine whether two cell clusters tend to interact with each other significantly and thus locate close to each other. Given a threshold defining the neighborhood radius of a cell, e.g., the median of the third nearest distance, a pair of cells can be assumed to "directly" interact with each other if their distance is less than the cutoff. Therefore, the total number of cell-cell interactions between two clusters can be counted. The statistical significance of the observed number of cell-cell interactions can be further evaluated by random permutation of the cluster labels of individual cells. With 1000 random permutations, the distribution of the randomly expected cell-cell interaction numbers of the given two cell clusters can be constructed, and thus right-tailed and left-tailed tests can be conducted, respectively, to calculate the P -values for the hypotheses that the observed interaction number was larger than that randomly expected (enrichment) and that the observed number was smaller than that randomly

expected (depletion). When P -values for enrichment between all clusters are obtained, the Benjamini-Hochberg procedure is used to estimate the q -values. If the enrichment (depletion) q -value for a given pair of cell clusters is less than 0.05, these two cell clusters are claimed to significantly interact with (disperse away from) each other. Otherwise, if the enrichment and depletion q -values are both larger than 0.05, the cell-cell interactions are assigned to the “other” type.

Determining the dominant ligand-receptor pairs underlying cell-cell interactions

Given a pair of cells, the contribution of a specific ligand-receptor interaction to the cell-cell interacting affinity can be calculated by the following formula:

$$c_k^{ij} = \frac{w_{L_k, R_k} (e_i^{L_k} \times e_j^{R_k} + e_i^{R_k} \times e_j^{L_k})}{\sum_{k=1}^K w_{L_k, R_k} (e_i^{L_k} \times e_j^{R_k} + e_i^{R_k} \times e_j^{L_k})} \quad (11)$$

where c_k^{ij} denotes the contribution of the k th ligand-receptor interaction to the cell-cell affinity of the i th and j th cells. Therefore, given a pair of cell clusters, the contribution of a specific ligand-receptor pair to the interactions of the two clusters can be calculated by the following formula:

$$c_k^{c_a, c_b} = \frac{1}{N} \sum_{\substack{i \in c_a, j \in c_b \\ d_{ij} \leq T}} c_k^{ij} \quad (12)$$

where $c_k^{c_a, c_b}$ denotes the contribution of the k th ligand-receptor interaction to the interactions of two clusters c_a and c_b , and N is the total number of cell pairs between c_a and c_b that conform to the definition of “direct” cell-cell interaction, i.e., the distance between two cells i and j should be less than the predefined threshold T ($d_{ij} \leq T$). When the contributions of all the ligand-receptor pairs are calculated, the ligand-receptor pairs with the highest scores are assumed to be the dominant molecular contributors underlying the cell clusters.

Evaluating the spatial effects of individual genes or cell clusters by in silico interference

CSOmap also provides functions to analyze the effects of in silico interfering genes or cell clusters on the cellular organization. In reality, cell-cell interactions form a highly nonlinear system, and thus it is hard to predict the spatial effects of gene alterations or cell interference. By simulating the cellular spatial organization via ligand-receptor mediated self-assembly, CSOmap provides an easy way to interrogate the nonlinear effects of ligand/receptor or cellular changes on the cellular organizations, and thus can provide important insights into the true biological mechanisms that are too expensive or even impossible to obtain by experimental methods. Currently, the in silico interference types of CSOmap include in silico gene knockdown, gene overexpression, adoptive cell transfer, and cell depletion. When cellular spatial organization with in silico interference is obtained, it can then be compared to the original organization to identify the significant differences. Although the current in silico interference can only examine the effects of ligand and receptors, it can be further enhanced by incorporating gene-gene interactions in the future to introduce dynamics for the expression levels of ligands and receptors. Since the output coordinates of CSOmap are in virtual spaces, it is now not possible to directly compare the changes of cellular spatial organizations at the coordinate level. All the comparisons in the current manuscript were conducted after abstracting the coordinate results into cell-cell interacting graphs. For the HNC dataset, in silico CD63 overexpression was conducted through changing all the original CD63 expression values to TPM 5000.

For the melanoma dataset, in silico CD63 knockdown was implemented by resetting the original expression values to zero. For the CRC T cell dataset, Treg depletion was implemented by removing all the cells belonging to the CD4-CTLA4 cluster.

Performance evaluations of CSOmap

The mouse liver lobule and paired-cell sequencing datasets were downloaded from the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>) with accession numbers GSE84498 and GSE108561, respectively. The human and mouse pancreas datasets were downloaded with the accession number GSE84133. The 10× genomics-based human placenta dataset was downloaded from <http://data.teichlab.org> (maternal-fetal interface). The 10× genomics-based Tabula Muris datasets were downloaded with the accession number GSE109774, with neuron and immune cells removed to investigate the interactions within and between organs. The scRNA-seq dataset of human lungs from healthy donors and patients with pulmonary fibrosis was downloaded with the accession number GSE122960. The HNC and melanoma (TN) datasets were downloaded with accession numbers GSE103322 and GSE72056, respectively. The melanoma (ICR) dataset was downloaded through the Single Cell Portal (https://portals.broadinstitute.org/single_cell/study/melanoma-immunotherapy-resistance). The T cell datasets of HCC, NSCLC and CRC were downloaded from the Gene Expression Omnibus with accession numbers GSE98638, GSE99254 and GSE108989. All the expression values of the original datasets were converted to TPM for CSOmap analysis. The reconstructed spatial organizations of HNC and melanoma datasets were compared to the IHC images published in their original papers by examining the spatial characteristics of various cell groups. The reconstructed spatial organizations of the three T cell datasets were validated by comparing their biological corollaries with published literature. Because cellular spatial organization is the result of many factors including cell identity, cellular environment, cell developmental history, and many physical and chemical ingredients, it is currently impossible to directly compare the precise structures of the reconstructed organizations with experimentally obtained images. Comparison of the spatial characteristics and biological corollaries is acceptable. The IHC and scRNA-seq experiments of the HCC sample was conducted in house, with the scRNA-seq operations following the protocol of SMART-seq2 and the IHC experiment and image analysis following the protocols provided by Nghiem et al.⁴⁷ and the PerkinElmer Vectra automated multispectral microscope. The scRNA-seq data were deposited into EGA with accession ID EGAS00001003449. The antibodies used in the IHC staining were from Abcam: PD1 (EPR4877(2), ab137132), CD8 (144B, ab17147), CD68 (EPR20545, ab213363), FOXP3 (236A/E7, ab20034), CLEC9A (8F9, ab104910).

T cell adoptive transfer analysis

In the in silico T cell adoptive transfer experiments, the HNC and melanoma (TN) datasets were used to simulate the corresponding tumor microenvironments. For fair comparison, the adoptively transferred T cells were sampled from the HCC T cell dataset. For each in silico adoptive transfer experiment, the number of adoptively transferred T cells was set to be the same as the number of T cells of the original datasets. To simulate the TCR-pMHC interactions between adoptively transferred T cells and the malignant cells, a pair of pseudo-ligand and -receptor was added into the ligand-receptor network. The pseudo-ligand (pMHC) was set to be only expressed in malignant cells with TPM 5000 and the pseudo-receptor (TCR) was set to be only expressed in adoptively transferred T cells with varying expression levels. The TCR-pMHC affinity was represented by the varying levels of the pseudo-receptor. ANOVA with repeated measures (implemented by the *ranova* function of Matlab

R2016b) was used to evaluate the effects of TCR-pMHC affinity, origins of T cells, and cancer types.

DATA AVAILABILITY

The mouse liver lobule and paired-cell sequencing datasets were downloaded from the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>) with accession numbers GSE84498 and GSE108561, respectively. The human and mouse pancreas datasets were downloaded with the accession number GSE84133. The 10x genomics-based human placenta dataset was downloaded from <http://data.teichlab.org> (maternal-fetal interface). The 10x genomics-based Tabula Muris datasets were downloaded with the accession number GSE109774, with neuron and immune cells removed to investigate the interactions within and between organs. The HNC and melanoma (TN) datasets were downloaded with accession numbers GSE103322 and GSE72056. The melanoma (ICR) dataset was downloaded through the Single Cell Portal (https://portals.broadinstitute.org/single_cell/study/melanoma-immunotherapy-resistance). The T cell datasets of HCC, NSCLC and CRC were downloaded from the Gene Expression Omnibus with accession numbers GSE98638, GSE99254 and GSE108989. The newly added HCC scRNA-seq data were deposited into EGA with accession ID EGAS00001003449. The scRNA-seq dataset of human lungs from healthy donors and patients with pulmonary fibrosis was downloaded with accession number GSE122960.

SOFTWARE AVAILABILITY

The software implementation of our CSOmap method is available at <https://doi.org/10.24433/CO.8641073.v1>.

ACKNOWLEDGEMENTS

We thank Boxi Kang for the IT assistance. We thank the Computing Platform of the CLS (Peking University) for providing computing resource. This project was supported by Beijing Advanced Innovation Centre for Genomics at Peking University, Key Technologies R&D Program (2016YFC0900100), and the National Natural Science Foundation of China (31991171, 31530036 and 91742203).

AUTHOR CONTRIBUTIONS

X.R. and Z.Z. conceived this study. X.R. designed the mathematical model and algorithm. G.Z. developed the software. X.R. and G.Z. performed the analysis. Q.Z., L.Z. and Y.S. conducted the validation experiment. X.R., G.Z. and Z.Z. wrote the manuscript with inputs from all the authors.

ADDITIONAL INFORMATION

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41422-020-0353-2>.

Competing interests: Z.Z. is a founder of Analytical Biosciences Limited. Other authors declare no competing interests.

REFERENCES

- Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
- Papalexis, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).
- Lee, J. H. Quantitative approaches for investigating the spatial context of gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **9**, e1369 (2017).
- Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
- Coons, A. H., Creech, H. J. & Jones, R. N. Immunological properties of an antibody containing a fluorescent group. *Proc. Soc. Exp. Biol. Med.* **47**, 200–202 (1941).
- Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
- Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).
- Achim, K. et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
- Nitzan, M., Karaiskos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).
- Beccari, L. et al. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018).
- Toda, S., Blauch, L. R., Tang, S. K. Y., Morsut, L. & Lim, W. A. Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science* **361**, 156–162 (2018).
- Ramilowski, J. A. et al. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
- Chen, L. & Flies, D. B. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* **13**, 227–242 (2013).
- van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Rodriguez, A. & Laio, A. Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014).
- Baron, M. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* **3**, 346–360 (2016).
- Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
- Halpern, K. B. et al. Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nat. Biotechnol.* **36**, 962–970 (2018).
- Tabula Muris Consortium. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
- Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
- Kierszenbaum, A. & Tres, L. *Histology and Cell Biology: An Introduction to Pathology* (Elsevier, 2019).
- Wolters, P. J., Collard, H. R. & Jones, K. D. Pathogenesis of idiopathic pulmonary fibrosis. *Annu. Rev. Pathol.* **9**, 157–179 (2014).
- Perelas, A., Silver, R. M., Arrossi, A. V. & Highland, K. B. Systemic sclerosis-associated interstitial lung disease. *Lancet Respir. Med.* **8**, 304–320 (2020).
- Takezaki, A. et al. A homozygous SFTPA1 mutation drives necroptosis of type II alveolar epithelial cells in patients with idiopathic pulmonary fibrosis. *J. Exp. Med.* **216**, 2724–2735 (2019).
- Yang, H. Z. et al. Targeting TLR2 attenuates pulmonary inflammation and fibrosis by reversion of suppressive immune microenvironment. *J. Immunol.* **182**, 692–702 (2009).
- Ramskold, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
- Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624 (2017).
- Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
- Lupia, A. et al. CD63 tetraspanin is a negative driver of epithelial-to-mesenchymal transition in human melanoma cells. *J. Invest. Dermatol.* **134**, 2947–2956 (2014).
- Kalluri, R. The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* **16**, 582 (2016).
- Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997 (2018).
- Zheng, C. et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* **169**, 1342–1356 (2017).
- Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
- Zhang, L. et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).
- Dieu-Nosjean, M. C., Goc, J., Giraldo, N. A., Sautès-Fridman, C. & Fridman, W. H. Tertiary lymphoid structures in cancer and beyond. *Trends Immunol.* **35**, 571–580 (2014).
- Zhou, X., Zhao, S., He, Y., Geng, S., Shi, Y. & Wang, B. Precise spatio-temporal interruption of regulatory T cell-mediated CD8⁺ T cell suppression leads to tumor immunity. *Cancer Res.* **79**, 585–597 (2019).
- Foa, R. & Catovsky, D. T-lymphocyte colonies in normal blood, bone marrow and lymphoproliferative disorders. *Clin. Exp. Immunol.* **36**, 488–495 (1979).
- Fujimoto, H. et al. Deregulated mucosal immune surveillance through gut-associated regulatory T cells and PD-1(+) T cells in human colorectal cancer. *J. Immunol.* **200**, 3291–3303 (2018).
- Sade-Feldman, M. et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998–1013 (2018).

41. Mikucki, M. E. et al. Non-redundant requirement for CXCR3 signalling during tumoricidal T-cell trafficking across tumour vascular checkpoints. *Nat. Commun.* **6**, 7458 (2015).
42. Stockinger, B., Barthlott, T. & Kassiotis, G. The concept of space and competition in immune regulation. *Immunology* **111**, 241–247 (2004).
43. Rosenberg, S. A. & Restifo, N. P. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* **348**, 62–68 (2015).
44. Keren, L. et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion Beam imaging. *Cell* **174**, 1373–1387 (2018).
45. Milone, M. C. & Bhoj, V. G. The pharmacology of T cell therapies. *Mol. Ther. Methods Clin. Dev.* **8**, 210–221 (2018).
46. Human chemokine and chemokine receptors. <https://www.ncbi.nlm.nih.gov/books/NBK6294/table/A13516/?report=objectonly>.
47. Nghiem, P. T. et al. PD-1 blockade with pembrolizumab in advanced merkel-cell carcinoma. *N. Engl. J. Med.* **374**, 2542–2552 (2016).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020