

# Cross-tissue multicellular coordination and its rewiring in cancer

<https://doi.org/10.1038/s41586-025-09053-4>

Received: 11 April 2024

Accepted: 22 April 2025

Published online: 28 May 2025

Open access

 Check for updates

Qiang Shi<sup>1,4</sup>✉, Yihan Chen<sup>1,4</sup>, Yang Li<sup>1</sup>, Shishang Qin<sup>1</sup>, Yu Yang<sup>1,2</sup>, Yang Gao<sup>3</sup>, Linnan Zhu<sup>1</sup>, Dongfang Wang<sup>1</sup> & Zemin Zhang<sup>1,2</sup>✉

The multicellular coordination that underlies tissue homeostasis and disease progression is of fundamental interest<sup>1–5</sup>. However, how diverse cell types are organized within tissue niches for cohesive functioning remains largely unknown. Here we systematically characterized cross-tissue coordinated cellular modules in healthy tissues, uncovering their spatiotemporal dynamics and phenotypic associations, and examined their rewiring in cancer. We first compiled a comprehensive single-cell transcriptomic atlas from 35 human tissues, revealing substantial inter-tissue variability in cellular composition. By leveraging covariance in cellular abundance, we identified 12 cellular modules with distinct cellular compositions, tissue prevalences and spatial organizations, and demonstrated coordinated intercellular communication within cellular modules using *in situ* spatial and *in vivo* perturbation data. Among them, two immune cellular modules in the spleen showed contrasting chronological dynamics with ageing. Analysis of multicellular changes in the breast revealed a menopausal trajectory associated with fibroblast dynamics. Furthermore, interrogation across cancer types uncovered simultaneous rewiring of two types of multicellular ecosystem during tumour progression, including the loss of tissue-specific healthy organization and the emergence of a convergent cancerous ecosystem. These findings reveal fundamental organizing principles of multicellular ecosystems in health and cancer, laying a foundation for further investigations into tissue-level functional coordination across diverse contexts.

The human organism is composed of an intricate mosaic of approximately 37 trillion diverse cells, harmoniously arranged into specialized tissues, organs and interconnected systems. Over the past decade, advances in single-cell technologies and initiatives such as the Human Cell Atlas<sup>6</sup> have enabled comprehensive profiling of various human tissues, revealing a substantial number of previously uncharacterized cell types<sup>7–9</sup>. However, it remains largely unclear how multiple cell types precisely coordinate within their native niches to perform essential physiological functions, and which components are particularly constrained under pathological conditions. In the context of tumours, for instance, the complex network of cellular interactions that underlie pro- and anti-tumour responses is still not fully elucidated<sup>10</sup>. This knowledge gap hampers our understanding of tissue organization and function, impeding the development of more effective strategies for disease diagnosis, monitoring and intervention.

Previous studies aimed to establish connections among cell types by measuring cell–cell interactions based on known ligand–receptor pairs<sup>11,12</sup>. Although these efforts have provided valuable insights into intercellular communication within specific contexts, they fall short of systematically unravelling complex tissue-level signalling networks, in which multiple cell types are coordinated into higher-order functional modules. Additionally, spatially resolved technologies have been

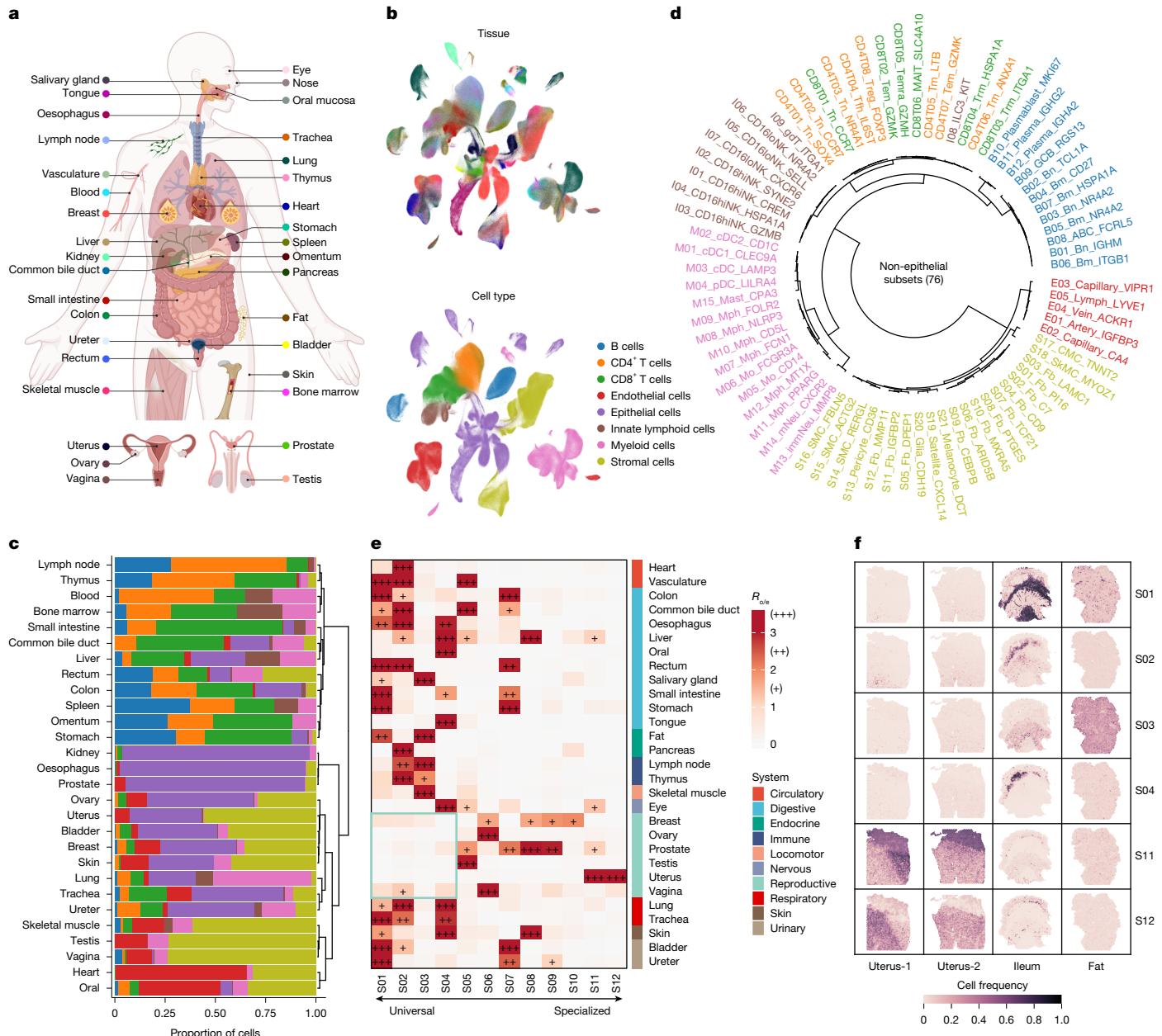
utilized to characterize cellular organization in tissues<sup>13–16</sup>; however, the limited number of genes that has been profiled and the lack of bona fide single-cell resolution restrict their ability to comprehensively portray those intricate multicellular structures. Notably, functionally related organs may share similar functional modules, as exemplified by mucosal immunity in both the small intestine and colon<sup>13</sup>, suggesting the cross-tissue potential of multicellular ecosystems. Therefore, a systematic investigation of multicellular coordination is essential to elucidate fundamental units across human tissues.

## A single-cell atlas of the human

To systematically elucidate the multicellular coordination, we assembled a comprehensive pan-tissue transcriptomic atlas at single-cell resolution. Following stringent quality control, we obtained a total of 2,293,951 high-quality cells from 706 healthy samples across 35 human tissues for downstream analyses (Fig. 1a, Extended Data Fig. 1 and Supplementary Table 1). To harmonize such extensive datasets, we evaluated several widely used data integration tools using the scIB platform<sup>17</sup>, with BBKNN<sup>18</sup> emerging as the top performer (Extended Data Fig. 2). Of note, uniform manifold approximation and projection (UMAP) embedding demonstrated distinct separation among various

<sup>1</sup>Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, and Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. <sup>2</sup>Chongqing Medical University, Chongqing, China. <sup>3</sup>Institute of Cancer Research, Shenzhen Bay Laboratory, Shenzhen, China. <sup>4</sup>These authors contributed equally: Qiang Shi, Yihan Chen. <sup>✉</sup>e-mail: shiq@pku.edu.cn; zemin@pku.edu.cn

# Article



**Fig. 1 | A single-cell atlas of the human.** **a**, Overview of the pan-tissue single-cell transcriptomic atlas. Created in BioRender; Shi, Q. (2025), <https://BioRender.com/g2chn78>. **b**, UMAP visualization of total cells (dots) coloured by tissue (top) or cell type (bottom). Tissue colours match those in **a**. **c**, Cell-type composition across tissues. Cell-type colours match those in **b**. Only sorting-free samples are included. **d**, Unsupervised hierarchical clustering of 76 non-epithelial cell subsets coloured by cell type. ABC, age-associated B cells; Bm, memory B cells; Bn, naive B cells; cDC, conventional dendritic cells; CMC, cardiac muscle cells; DC, dendritic cells; Fb, fibroblasts; GCB, germinal centre B cells; gdT, γδ T cells; immNeu, immature neutrophils; MAIT, mucosal-associated invariant T cells; mNeu, mature neutrophils; Mo, monocytes; Mph, macrophages;

pDC, plasmacytoid dendritic cells; SkMC, skeletal muscle cells; SMC, smooth muscle cells; Tem, effector memory T cells; Temra, terminally differentiated effector memory or effector T cells; TfH, follicular helper T cells; Tm, memory T cells; Tn, naïve T cells; Treg, regulatory T cells; Trm, tissue-resident memory T cells. **e**, Tissue prevalence of fibroblast subsets measured by  $R_{o/e}$  (the ratio of observed to expected cell numbers) (Methods). Tissues are categorized and ordered by body system. Fibroblast subsets are ordered by Shannon equitability from most universal (S01) to most specialized (S12). **f**, Spatial distribution of fibroblast subsets across tissues (Visium). Relative frequencies of each fibroblast subset among total fibroblasts are shown for individual spots.

cell types and effective integration across different sexes, tissues and datasets (Fig. 1b and Supplementary Fig. 1a).

We performed unsupervised cell clustering and annotation hierarchically. Initially, we established eight broad cell types using canonical markers, consistent with their annotations in the original studies (Fig. 1b and Supplementary Fig. 1b,c). Notably, approximately 45% of the cells in our atlas were attributed to the immune compartment (Supplementary Fig. 1c), providing a solid foundation for studying

the phenotypic diversification of immune cells<sup>19</sup>. The composition of these cell types exhibited noticeable variation among diverse tissues ( $P < 2.2 \times 10^{-16}$ , Chi-squared test) (Fig. 1c). For example, peripheral blood and immune organs or tissues, including the bone marrow, lymph nodes, omentum, spleen and thymus, were predominantly composed of immune cells, aligning with their roles in immune cell production, maturation or storage. By contrast, the reproductive tissues, such as the uterus, vagina and testis, exhibited a higher proportion of stromal cells.

Next, each cell type was further dissected into several distinct cell states or subsets (Fig. 1d). Despite sharing similar transcriptomic profiles, cell subsets within the same cell types displayed clear preferences for specific tissues (Supplementary Fig. 2). For instance, endothelial cells—which are critical components of the vasculature—exhibited unique delineations based on their sources, whether from blood or lymphatic vessels. Additionally, a rare cell subset called age-associated memory B cells (ABC, B08), which constituted less than 1% of total B cells, was prominently present not only in the liver, bone marrow and spleen—as previously reported<sup>7,20</sup>—but also in unexpected locations such as the ureter and skeletal muscle.

Among stromal cells, fibroblasts exhibit substantial diversity, encompassing both universal and specialized subsets<sup>21</sup>. We identified 12 fibroblast subsets across various tissues, with subsets S01 to S04 being the most widely distributed (Supplementary Fig. 3a–d and Supplementary Table 2). These universal subsets were considerably depleted in the reproductive system, whereas reproductive-system-specific subsets appeared more specialized to specific tissues rather than the entire reproductive system (Fig. 1e). This suggests unique architectural adaptations related to reproductive functions. Spatially resolved transcriptomics revealed distinct spatial distribution patterns for two uterus-specific fibroblast subsets, with S11 corresponding to the endometrium and S12 to the myometrium<sup>22</sup> (Fig. 1f). Moreover, while exhibiting the strongest specificity for particular tissues, epithelial cells within the same human body systems displayed higher similarities than those among different systems (Supplementary Fig. 3e). In sum, we constructed a comprehensive pan-tissue single-cell atlas (<http://cm.cancer-pku.cn>) and revealed substantial preferences of cell subsets across various tissues.

## Identification of cross-tissue CMs

A notable illustration of multicellular coordination is evident in gut-associated mucosal immunity, where diverse cell types such as lymphocytes, dendritic cells and epithelial cells collaborate to defend against pathogenic insults<sup>23</sup>. Our study aims to determine whether such coordinated multicellular ecosystems represent a recurrent theme across the human body, thereby addressing the fundamental question of how these ecosystems contribute to cohesive functional coordination among diverse cell types at the tissue level.

To investigate this, we translated the concept of multicellular coordination into representations of co-occurring cellular networks and developed a computational framework, CoVarNet. It reconstructs cellular module (CM) networks by leveraging covariance in cell subset frequencies across samples through two parallel modules (Fig. 2a and Methods). The first module utilizes non-negative matrix factorization (NMF) to identify a set of factors, prioritizing cell subsets on the basis of their weights. The top subsets from each factor serve as co-occurring nodes in the CM network. The second module determines specifically correlated subset pairs, which act as potential edges in the network. Multiple CM networks are then constructed by connecting these co-occurring nodes through the potential edges, followed by topological and statistical evaluations.

Applying this framework to our pan-tissue atlas, we identified a total of 12 CMs (Fig. 2b, Extended Data Fig. 3a–d and Supplementary Fig. 4). In light of our focus on cross-tissue multicellular coordination, epithelial cells, owing to their highly tissue-divergent nature, were excluded from this analysis. The vast majority (88.16%, 67 out of 76) of non-epithelial cell subsets participated in at least one CM, whereas the remaining subsets (9 out of 76) were excluded either owing to technical underrepresentation, including germinal centre B cells (B09) and neutrophils (M13 and M14), or owing to their highly specialized roles in a single tissue, as in the case of skeletal muscle cells (S18). In particular, one-quarter of these subsets (17 out of 67) were involved in multiple CMs, and none of the networks clearly exhibited the presence

of a hub node. These findings are consistent with fundamental biological principles, emphasizing that all cell types are functionally distinct, with some serving unique roles in different tissues. Additionally, we utilized the coefficient matrix from NMF as a measure of CM activities for each sample (Extended Data Fig. 3e). As expected, the activities of all CMs exhibited positive correlations with the frequencies of their component cell subsets across tissues (Supplementary Fig. 5). Given that each sample hosted a predominant CM, we categorized all samples into 12 CM types (CMTs) according to their most abundant CMs. Notably, all CMTs were composed of multiple tissues, indicating the cross-tissue nature of the identified CMs (Extended Data Fig. 3e,f).

To validate the robustness of these results, we conducted an integrative analysis combining single-cell RNA sequencing (scRNA-seq) data with approximately 12,000 RNA-sequencing (RNA-seq) profiles from the Genotype-Tissue Expression (GTEx) project<sup>24</sup> (Extended Data Fig. 4a,b and Methods). Of note, all CMTs could be well recovered, and each bulk sample also had a predominant CM (Extended Data Fig. 4c,d). Furthermore, the expression patterns of markers for CM-related cell subsets demonstrated strong concordance between the two data types (Extended Data Fig. 4e), further confirming the validity of the CM landscape that we established.

## CM annotations

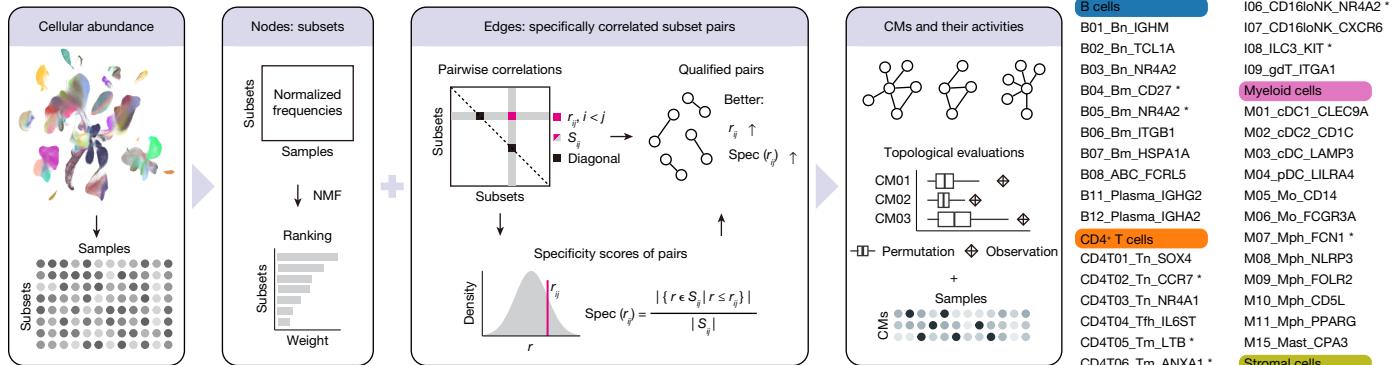
Most CMs exhibited notable preferences for specific human body systems (Fig. 2c), validated by external GTEx data (Extended Data Fig. 4f). Among the 12 identified CMs, CM04, CM05, CM06 and CM09 were enriched in primary immune organs (bone marrow and thymus), secondary immune organs (lymph nodes and spleen) and peripheral blood. This enrichment correlated with their cellular compositions, primarily comprising immune subsets prevalent in these tissues (Fig. 2b). Similarly, CM07 and CM12 demonstrated preferences for the reproductive system, aligning with the presence of those specialized fibroblasts predominantly found in the ovary, vagina, breast and prostate (Figs. 1e and 2b). CM02 and CM03 were mainly distributed in the urinary system and gastrointestinal tract, whereas CM08 was enriched in barrier tissues such as the skin, oral mucosa, tongue, vagina and trachea, indicating that these CMs may represent multicellular ecosystems within mucosa-associated lymphoid tissues. Additionally, CM10 appeared to function as a vascular unit, characterized by pericytes (S13), smooth muscle cells (S14) and vascular endothelial cells (E01 and E02) (Fig. 2b and Extended Data Fig. 3c), and was enriched in the vasculature, heart, skin and fat tissues. By contrast, CM11 showed enrichment in the lung, kidney, liver and fat, implying roles in metabolic processes. CM01, characterized by tissue-resident macrophages (M09), universal fibroblasts (S03 and S05) and lymphatic endothelial cells (E05) (Fig. 2b and Extended Data Fig. 3c), was broadly distributed across nearly all human body systems (Fig. 2c and Extended Data Fig. 3f), suggesting a universal multicellular organization. Collectively, these findings demonstrate that the identified CMs exhibit distinct preferences for multiple tissues, thus representing cross-tissue multicellular ecosystems (Fig. 2d).

## Spatial characteristics of CMs

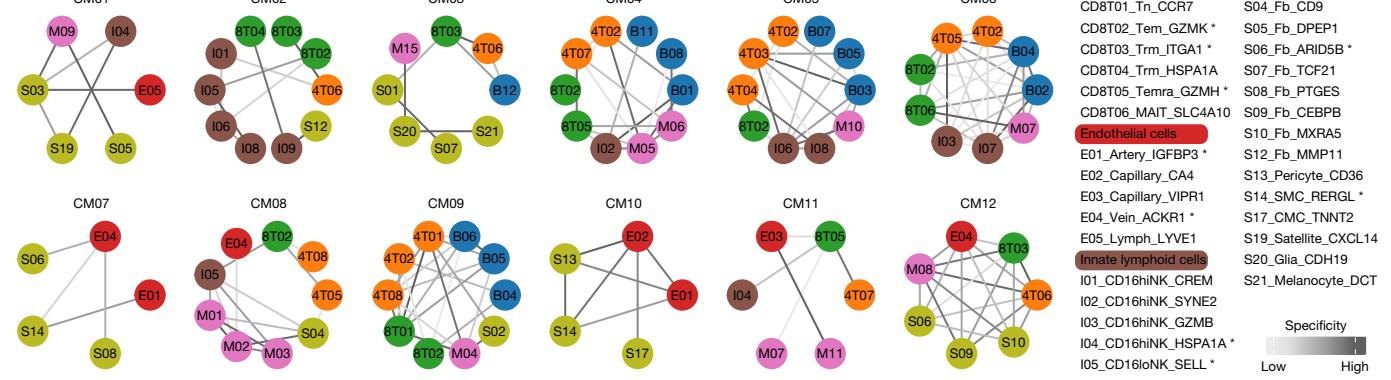
To further characterize CMs in a spatial context, we mapped them onto spatial transcriptomics data (Visium) using cell2location<sup>25</sup> (Supplementary Table 3). We first examined CM08 and CM10, both of which were enriched in the skin. Our analysis revealed prominent spatial colocalization of cell subsets within CMs (Extended Data Fig. 5a–c). CM08, which included various immune cells such as dendritic cells and T cells, was localized to the epidermis and adjacent dermis layers, where it is likely to have a key role in immune defence. By contrast, CM10 represented components of the blood vessel that are located

# Article

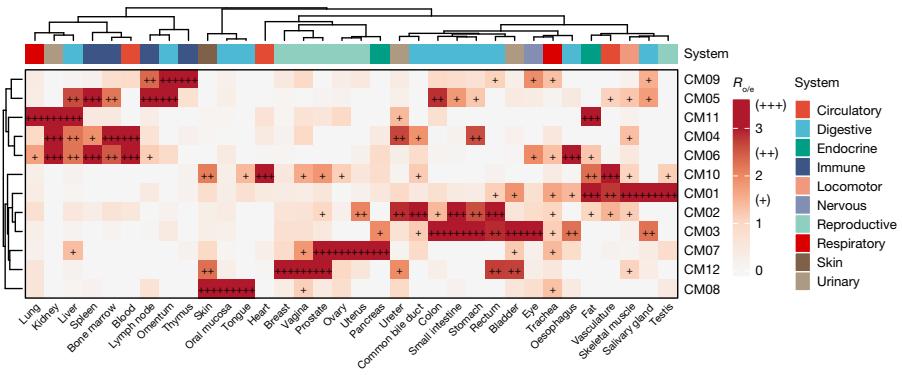
a



b



c



**Fig. 2 | Systematic identification of cross-tissue CMs.** **a**, Schematic of the CoVarNet workflow (Methods). Spec, specificity. **b**, Network plots of the 12 CMs identified from the pan-tissue analysis. Nodes represent cell subsets labelled with short names and coloured by cell type. Edge colour indicates correlation

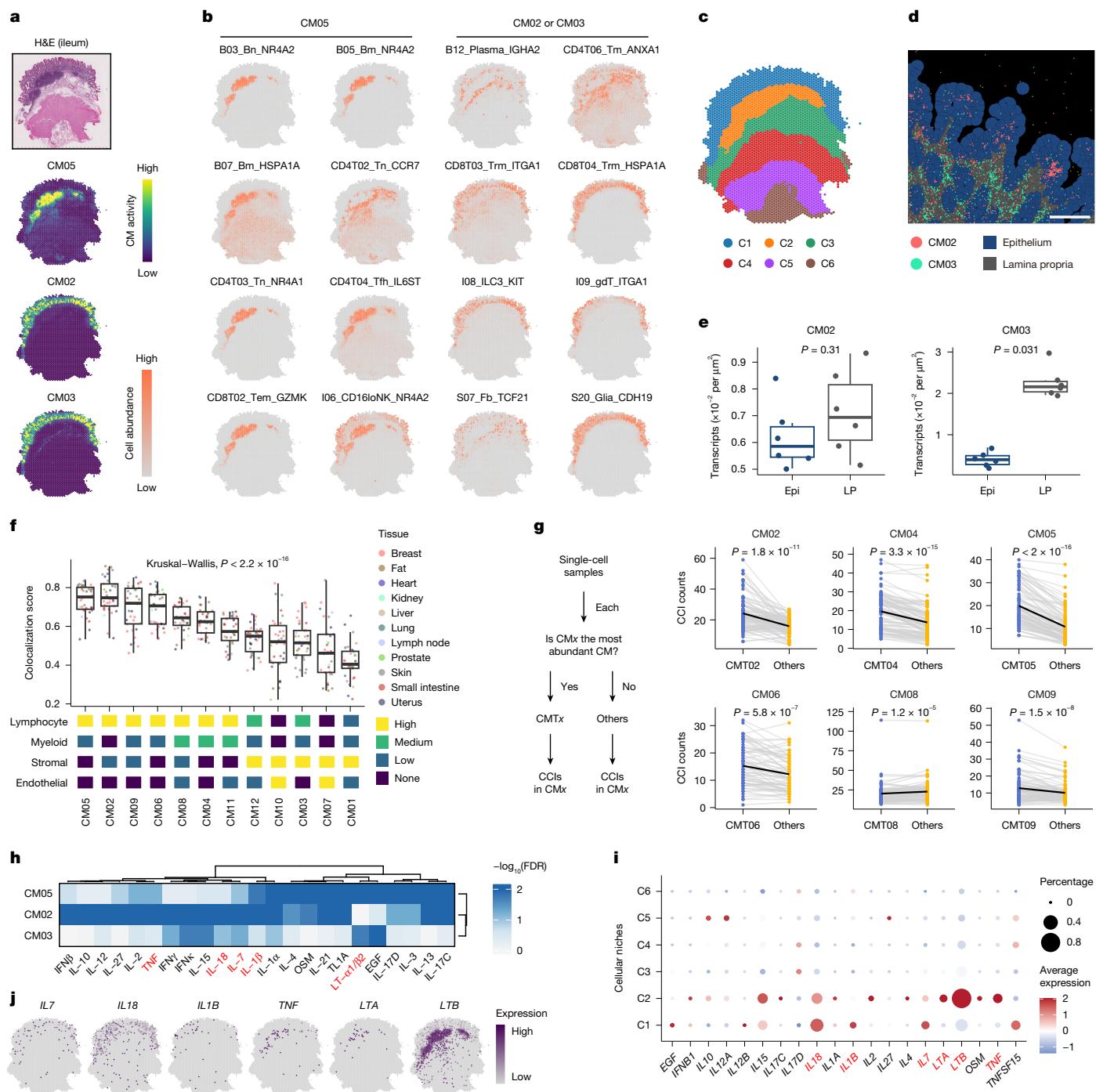
specificity. Asterisks mark cell subsets involved in multiple CMs. **c**, Tissue prevalence of CMs measured by  $R_{\text{ave}}$ . Tissues are categorized and coloured by body system. **d**, A brief summary of CM annotations.

primarily in the dermis layer. Of note, some CM08 subsets, such as vein endothelial cells (E04), also overlapped with CM10 subsets, highlighting the interconnection of different CMs within the tissue microenvironment. These findings were consistently observed across samples from six available donors (Extended Data Fig. 5d), underscoring the robust existence of CMs.

Out of the 12 identified CMs, CM02, CM03 and CM05 exhibited prominent enrichment in the small intestine (Fig. 2c), a tissue with distinct anatomical organization<sup>13</sup>. Consistent spatial patterns of these CMs were observed in four sequential ileum sections from a single donor (Fig. 3a,b and Extended Data Fig. 5e,f). Unsupervised spatial clustering using CellCharter<sup>26</sup> further validated their organization that CM05 aligned with the C2 niche, whereas CM02 and CM03 aligned with the C1 niche (Fig. 3c and Extended Data Fig. 5g). Notably, CM05

showed elevated spatial concentration in the Peyer's patch region (Extended Data Fig. 5h), where its cell subsets, including naive B cells (B03), naive T cells (CD4T03) and follicular helper T cells (CD4T04), displayed notable colocalization (Fig. 3a,b). By contrast, CM02 and CM03 were located primarily in the intestinal mucosa, in line with their composition of IgA-producing plasma cells (B12), memory (CD4T06) and tissue-resident memory (CD8T03 and CD8T04) T cells, as well as innate immune cells (I08 and I09) (Fig. 3a,b). These results suggest that CM05 and CM02/CM03, respectively, recapitulate the inductive and effector modules of mucosal immunity, underscoring their potential functional roles within tissue ecosystems.

To deepen our understanding of CM02 and CM03, we utilized high-resolution Xenium data, enabling precise *in situ* characterization of hundreds to thousands of genes within cells and tissues. Using



**Fig. 3 | Spatially resolved multicellular coordination in CMs.** **a–c**, Visium analysis of the ileum showing spatial distribution of CMs (**a**), their component cell subsets (**b**), and cellular niches identified by CellCharter (**c**). H&E, haematoxylin and eosin. **d,e**, Representative Xenium imaging (**d**) and quantification (**e**;  $n = 6$ ) of CM02 and CM03 in intestinal mucosal sections (Methods). Scale bar, 100  $\mu$ m. Two-tailed paired Wilcoxon tests. Epi, epithelium; LP, lamina propria. **f**, Spatial colocalization and cellular composition of CMs. Top, CM colocalization scores across Visium samples ( $n = 34$ ) coloured by tissue (Methods and Supplementary Table 3). CMs are ordered by the median of colocalization scores across samples. Bottom, proportion of different cell lineages in all cell-subset nodes within CM networks shown in Fig. 2b. None, 0; 0 < low < 0.2; 0.2 ≤ medium < 0.4; high ≥ 0.4. **g**, CellPhoneDB cell–cell interaction (CCI) counts within CMs, measured in single-cell samples with high (for example,

CMT02) or low (others) CM activities. Dots represent unique subset pairs within CMs and black lines link the average counts. CM02,  $n = 90$ ; CM04,  $n = 90$ ; CM05,  $n = 90$ ; CM06,  $n = 72$ ; CM08,  $n = 72$ ; CM09,  $n = 90$ . Two-tailed paired Wilcoxon tests. **h**, Cytokine responsiveness in CM02, CM03 and CM05, as inferred using the Immune Dictionary<sup>28</sup>. The CM-level false discovery rate (FDR) denotes the minimum FDR value among all subsets within the CM (Methods). **i**, Expression of cytokine genes that are enriched in CM02, CM03 or CM05 across the ileum cellular niches in **c**. Only cytokines detected by Visium data are shown. **j**, Spatial distribution of selected cytokines (labelled in red in **h,i**) in the Visium section of the ileum shown in **a–c**. For box plots in **e,f**, the centre line represents the median, the box limits delineate the top and bottom quartiles, and the whiskers extend to the highest and lowest values within  $1.5 \times$  the interquartile range.

intestinal Xenium data, we designed a gene panel to differentiate the multiple cell subsets within CM02 and CM03, with gene transcript density representing the intensity of individual CMs (Extended Data

Fig. 5i). Our analysis revealed a notable enrichment of CM03 in the lamina propria, whereas CM02 exhibited a more uniform distribution across the tissue (Fig. 3d,e). This subtle difference, which cannot be

# Article

detected from Visium data alone, highlights not only the importance of integrating single-cell and spatial data, but also the superior sensitivity of single cell-based approaches in dissecting multicellular ecosystems with greater precision.

## Multicellular communication within CMs

The formation and maintenance of CMs within tissue niches may rely on the local microenvironment. We hypothesized that diverse cell subsets within CMs are spatially organized to respond collectively, with cellular crosstalk potentially varying across different tissue microenvironments.

To explore this, we first investigated the spatial organization of CM components using Visium data across various tissues (Supplementary Table 3). Our analysis revealed a strong association between CM spatial organization and composition. The highest spatial concordance was observed in CMs with a greater proportion of lymphocytes, followed by myeloid cells, stromal cells and endothelial cells (Fig. 3f and Extended Data Fig. 6a,b). To further understand the implications of these patterns, we analysed ligand–receptor-mediated cell–cell communication using single-cell data with CellPhoneDB<sup>27</sup> (Methods). Notably, endothelial cells and stromal cells produced a wider variety of ligands compared with lymphocytes (B cells, T cells and innate lymphoid cells) (Extended Data Fig. 6c). This aligns with the notion that spatial proximity may enhance the specificity of intercellular communication, suggesting that lymphocyte-enriched CMs foster particular interactions within local niches. Conversely, CMs with a higher proportion of stromal and endothelial cells—often situated farther apart—generated a more diverse array of signalling molecules, facilitating broader signalling interactions over longer distances (Extended Data Fig. 6d,e). These insights highlight the relationship among spatial organization, cellular composition and intercellular communication.

To investigate the effect of tissue microenvironments on intercellular communication within CMs, we assessed cell–cell interactions within CMs using single-cell samples from different CMTs (Extended Data Fig. 3e and Methods). Only six CMs with high colocalization scores were analysed. Our analysis revealed enhanced cell–cell interactions within samples with high CM activities, except for CM08 (Fig. 3g), implying distinct cellular phenotypes shaped by tissue microenvironments in a CM-dependent manner. To further investigate such phenotypes, we utilized recent *in vivo* perturbation data from the Immune Dictionary<sup>28</sup> to identify cell-type-specific cytokines that induced these cellular phenotypes (Methods). We found that half of the cell subsets showed responsive associations with at least one cytokine and CMs with high colocalization scores tended to harbour more diverse cytokines (Extended Data Fig. 6f,g and Supplementary Table 4). Diverse cell subsets clustered together on the basis of both their cell types and CM identities (Extended Data Fig. 6f), suggesting that cellular phenotypes are collectively determined by intrinsic properties and local stimuli. For example, CD8<sup>+</sup> effector memory T cells (CD8T02) exhibited distinct cytokine responses among different CMs, with TNF identified in CM02, CM08 and CM09, but absent in CM04 and CM06 (Extended Data Fig. 6f). Additionally, we examined spatial distribution of cytokines using the intestine data (Fig. 3a). Despite the transient and low-level expression characteristics of many cytokine genes, our analysis successfully validated their spatial enrichment, such as *IL7* and *IL18* in the CM02/CM03 (C1) niche, as well as *LTA* and *LTB* in the CM05 (C2) niche (Fig. 3h–j). These results provide a comprehensive overview of the cytokine-mediated regulatory landscape within CMs.

Collectively, our analyses of spatial organization and intercellular communication demonstrate that CMs, as fundamental tissue configurations, effectively recapitulate the complexity of multicellular ecosystems. These findings emphasize the interplay among diverse cell types within tissue microenvironments, providing insights into the mechanisms underlying tissue homeostasis.

## Coordinated ageing dynamics in the spleen

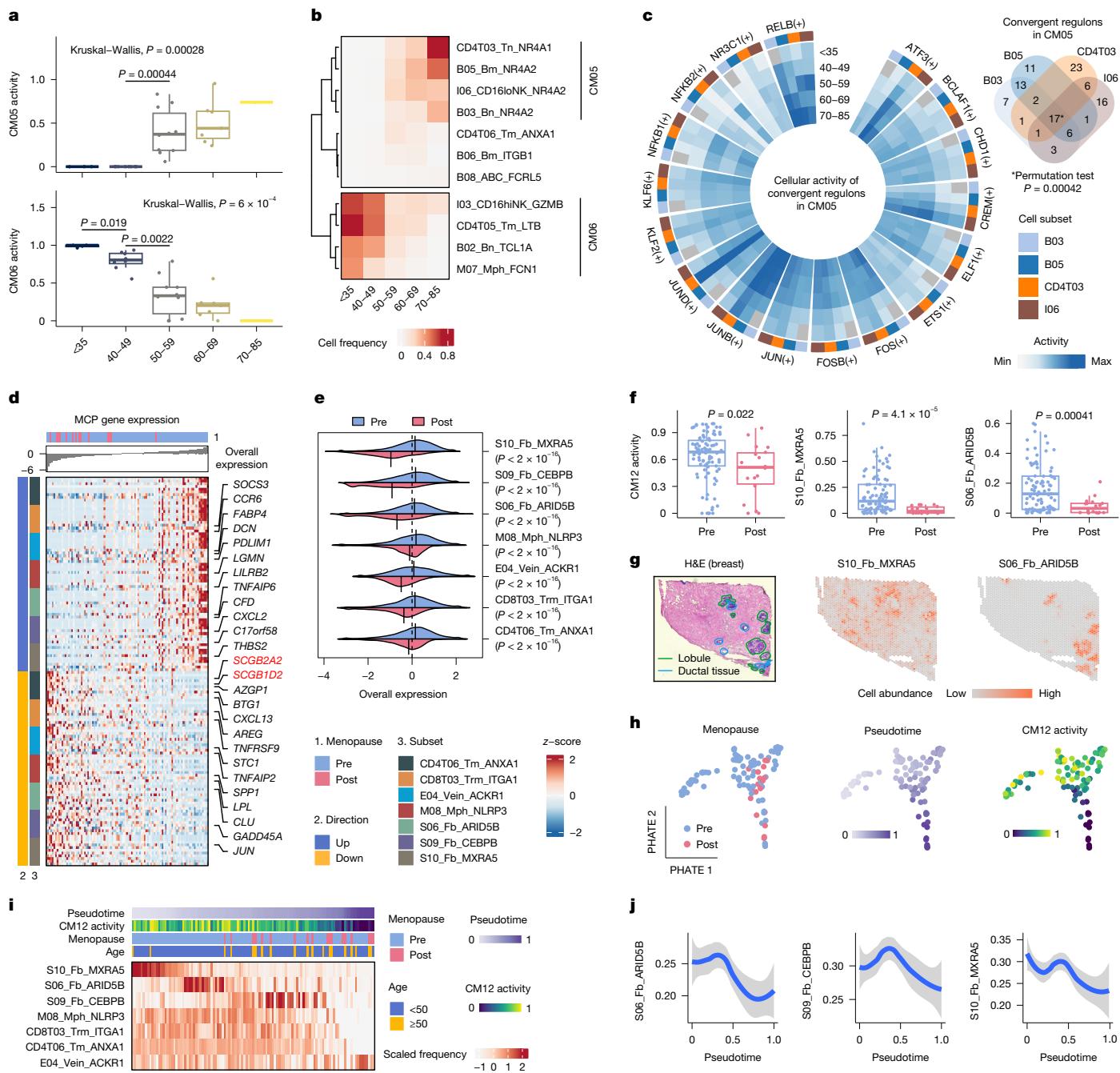
To further investigate the significance of CMs, we assessed CM associations with phenotypic data. As CM activities were primarily influenced by the tissue (Extended Data Fig. 7a), we conducted a systematic interrogation of individual CMs within tissue-specific contexts (Supplementary Fig. 6 and Methods). This analysis revealed comparable CM patterns between male and female individuals in nearly all non-reproductive tissues (Extended Data Fig. 7b), whereas the thymus exhibited notable variation across age groups (Extended Data Fig. 7c). Consistent with age-related thymic involution, the thymus-enriched CM09 showed reduced activity in older individuals, with components, such as naive T cells (CD8T01) and regulatory T cells (CD4T08) being more abundant in younger individuals (Extended Data Fig. 7d).

Another noteworthy age-related association was observed in the spleen, where CM05 increased chronologically and CM06 decreased (Fig. 4a and Extended Data Fig. 7c). As the spleen included various immune cells (Fig. 1c), we systematically examined all immune subsets in the spleen, identifying ten that varied across age groups (Extended Data Fig. 7e,f). Of these, 80% (8 out of 10) were components of CM05 or CM06 (Fig. 4b), highlighting that these cross-tissue CMs effectively captured tissue-specialized variations. Notably, the expansion of four CM05 subsets (B03, B05, CD4T03 and I06) was more pronounced with ageing than that of the previously reported ABCs (B08)<sup>7,20</sup> (Fig. 4b and Extended Data Fig. 7g), suggesting their potential roles in the ageing process. The accumulation of *NR4A1*-high CD4<sup>+</sup> T cells (CD4T03) in the spleen with ageing might be explained by the combined effects of immune tolerance<sup>29</sup> and thymic involution.

To gain deeper insights into these coordinated dynamics, we applied SCENIC<sup>30</sup> to uncover the regulators of these subsets. For each of the four CM05 subsets, we determined their specific regulons relative to other subsets within their cell types (Methods and Supplementary Table 5). Of note, these subsets exhibited notable overlap in regulons, and the convergent regulons in CM05 showed increased activity with ageing (Fig. 4c), shedding light on the common regulatory mechanisms that underlie multicellular coordination. Additionally, these subsets shared a set of signature genes, including 20 transcription factor genes such as *NR4A1* and *NR4A2* (Extended Data Fig. 8a–c). Further analysis identified eight key transcription factor genes (*ATF3*, *FOS*, *FOSB*, *JUN*, *JUNB*, *JUND*, *KLF6* and *NFKB1*) as both regulons and signature genes (Extended Data Fig. 8c). These key regulators tended to act as regulatory hubs (Extended Data Fig. 8d), targeting many common genes across different cell types (Extended Data Fig. 8e). These findings align with previous reports in mice that highlight the activation of *Jun* and *Fos* members of the AP-1 complex as a signature of immune ageing<sup>31</sup>. Of note, *NR4A1*, a key mediator of T cell dysfunction<sup>29</sup>, was regulated by multiple key regulators across cell types (Extended Data Fig. 8e), suggesting possible immune dysfunction associated with ageing. In summary, these results in the spleen, the largest peripheral immune organ in adults, underscore coordinated behaviour at the molecular, cellular and multicellular levels. Further investigation is warranted to identify the functional mechanisms underlying these dynamics in human ageing.

## Fibroblast-engaged menopausal trajectory

To explore the role of fibroblasts within multicellular ecosystems, we focused on breast-enriched CM12, which comprised three specialized fibroblast subsets (S06, S09 and S10) alongside other diverse cells (Fig. 2b,c). We first utilized DIALOGUE<sup>3</sup> to investigate whether local microenvironments in the breast triggered CM12-related coordinated multicellular programmes (MCPs), representing combinations of gene programmes across different cell subsets. Notably, this analysis identified an MCP that was upregulated in pre-menopausal samples compared with post-menopausal samples (Fig. 4d,e, Supplementary Fig. 7 and Supplementary Table 6). For instance, expression



**Fig. 4 | Multicellular dynamics in context.** **a**, CM05 and CM06 activities in spleen samples across age groups. Dots represent individual samples. <35,  $n = 3$ ; 40–49,  $n = 8$ ; 50–59,  $n = 10$ ; 60–69,  $n = 7$ ; 70–85,  $n = 1$ . Kruskal-Wallis tests and two-tailed unpaired Wilcoxon tests. **b**, Sample-averaged frequencies of cell subsets across age groups. **c**, Circular heatmap (left) showing cellular activity of 17 convergent regulons in CM05 (right) across age groups. Min, minimum; max, maximum. **d**, Sample-averaged expression of CM12 MCP genes (rows) across breast samples (columns). Genes are categorized by their direction (up- or down-regulation, left colour bar) and associated cell subsets (right colour bar). Samples are sorted by overall expression (top bar plot) and are labelled by menopausal status (top colour bar). **e**, Distribution of CM12 MCP overall expression in pre- and post-menopausal breast cells grouped by cell subsets. Vertical black line indicates the median of the distribution. Two-tailed unpaired Wilcoxon tests. **f**, CM12 activity (left) and frequencies of fibroblast

subsets S10 (middle) and S06 (right) in pre-menopausal ( $n = 93$ ) and post-menopausal ( $n = 18$ ) samples. Two-tailed unpaired Wilcoxon tests. **g**, Visium analysis of the breast showing histopathological regions annotated (left) and spatial distribution of fibroblast subsets S10 (middle) and S06 (right). **h**, PHATE (potential of heat diffusion for affinity-based trajectory embedding) visualization of breast samples (dots) coloured by menopausal status (left), pseudotime (middle) or CM12 activity (right) (Methods). **i**, Heat map showing frequencies of CM12 subsets (rows) across breast samples sorted by pseudotime. **j**, Trend lines show LOESS-smoothed fibroblast inflammatory scores along pseudotime in breast samples. The error bands show the 95% confidence intervals. For box plots in **a,f**, the centre line represents the median, the box limits delineate the top and bottom quartiles, and the whiskers extend to the highest and lowest values within  $1.5 \times$  the interquartile range.

of genes such as *SCGB2A2* and *SCGB1D2* increased in post-menopausal samples across most subsets (Fig. 4d), aligning with previous reports highlighting *SCGB2A2* as a promising biomarker for breast

detection<sup>32</sup>. Additionally, many inflammatory genes, including human leukocyte antigen genes, were found to be more highly expressed in pre-menopausal samples across cell subsets (Supplementary Table 6).

# Article

Additional analysis confirmed a decrease in inflammatory scores after menopause (Extended Data Fig. 9a). Although this may seem at odds with literature suggesting a systemic increase in inflammation with ageing<sup>33</sup>, we hypothesize that breast tissue, as a reproductive organ, is more strongly influenced by levels of oestrogen, leading to localized reductions in inflammation. These results indicate coordinated phenotypic shifts among diverse subsets within CM12 in response to menopause.

Next, we systematically examined the association between all non-epithelial subsets with menopausal status in the breast. Fibroblasts showed a particularly strong association with menopause, with subsets S10 and S06 exhibiting the most notable decreases in post-menopausal samples (Fig. 4f and Extended Data Fig. 9b). Notably, S10 displayed high expression of collagen genes (*COL1A1*, *COL1A2* and *COL3A1*) (Extended Data Fig. 9c), consistent with previous reports demonstrating a reduction<sup>34</sup> in such fibroblasts in women over the age of 50. Spatial analysis revealed that S10 was uniformly distributed across the connective tissue, whereas S06 was located primarily in the lobules (Fig. 4g). By contrast, immune subsets and S09 fibroblasts within CM12 were found to colocalize in the ductal tissues of the breast (Extended Data Fig. 9d), with little change in their abundance observed after menopause. These findings underscore distinct functional roles of different fibroblast subsets in the breast.

In line with these molecular and cellular observations, post-menopausal women exhibited decreased CM12 activity compared with pre-menopausal women (Fig. 4f). Of note, such decrease was less pronounced in women over 50 (Extended Data Fig. 9e), suggesting that CM12 recapitulates menopause-associated biological changes rather than merely chronological variation. Further tailored studies are needed to fully elucidate the respective contributions of ageing and menopause, as well as their potential interactions.

Given that these changes are often progressive, we hypothesized that alterations in CM12 could serve as indicators of menopausal progression. Using the frequencies of CM12 cell subsets, we identified a menopausal trajectory that transitions from pre-menopausal to post-menopausal states (Fig. 4h). Specifically, the frequency of S10 exhibited a consistent decrease along this trajectory, whereas S06 tended to decrease later in the process (Fig. 4i). Notably, these fibroblasts (S06, S09, and S10) exhibited decreased inflammatory scores along the trajectory, particularly after menopause (Fig. 4j). Of note, we replicated the menopausal trajectory and fibroblast changes using an external breast dataset (Extended Data Fig. 9f,g), underscoring the robustness of our results. Together, these multicellular analyses highlight spatiotemporal dynamics of specialized fibroblasts in the breast.

## Multicellular rewiring in cancer

To systematically understand the multicellular ecosystems in cancer, we extended our analysis to the tumour microenvironment (TME), a pathological tissue niche in which diverse immune and stromal cells interact to form a complex network<sup>10</sup>. We first established a comprehensive pan-cancer single-cell transcriptomic atlas comprising 1,062 clinical samples from 29 cancer types, identifying 91 cell subsets including 15 cancer-associated subsets previously reported across various cancer types<sup>35–39</sup> (Fig. 5a,b, Extended Data Fig. 10, Supplementary Fig. 8 and Supplementary Tables 7–9) (<http://cm.cancer-pku.cn>).

Previous studies have shown that histologically normal tissues adjacent to tumours harbour genetic alterations and manifest a unique intermediate state between healthy and tumour tissues<sup>40,41</sup>. Thus, we used adjacent non-tumour samples as a surrogate for precancerous tissues. To examine multicellular dynamics during tumour progression, we focused on eight cancer types with matched healthy, tumour and adjacent non-tumour samples (Fig. 5c). A notable reduction in healthy CM activity was observed in tumour samples across cancer types, suggesting a widespread disruption of tissue-specific multicellular ecosystems (Fig. 5d,e and Extended Data Fig. 11a). Notably, CM08

maintained consistent activity across healthy, adjacent non-tumour and tumour samples, suggesting that the multicellular ecosystem in healthy tissues was relatively well-preserved in cutaneous squamous cell carcinoma (Extended Data Fig. 11a). This finding aligns with the superior response to immunotherapy observed in skin cancers, such as melanoma and cutaneous squamous cell carcinoma, compared with other cancer types<sup>42</sup>, emphasizing the significance of contextualizing healthy ecosystems within the framework of cancer research (Supplementary Figs. 9 and 10 and Supplementary Table 10).

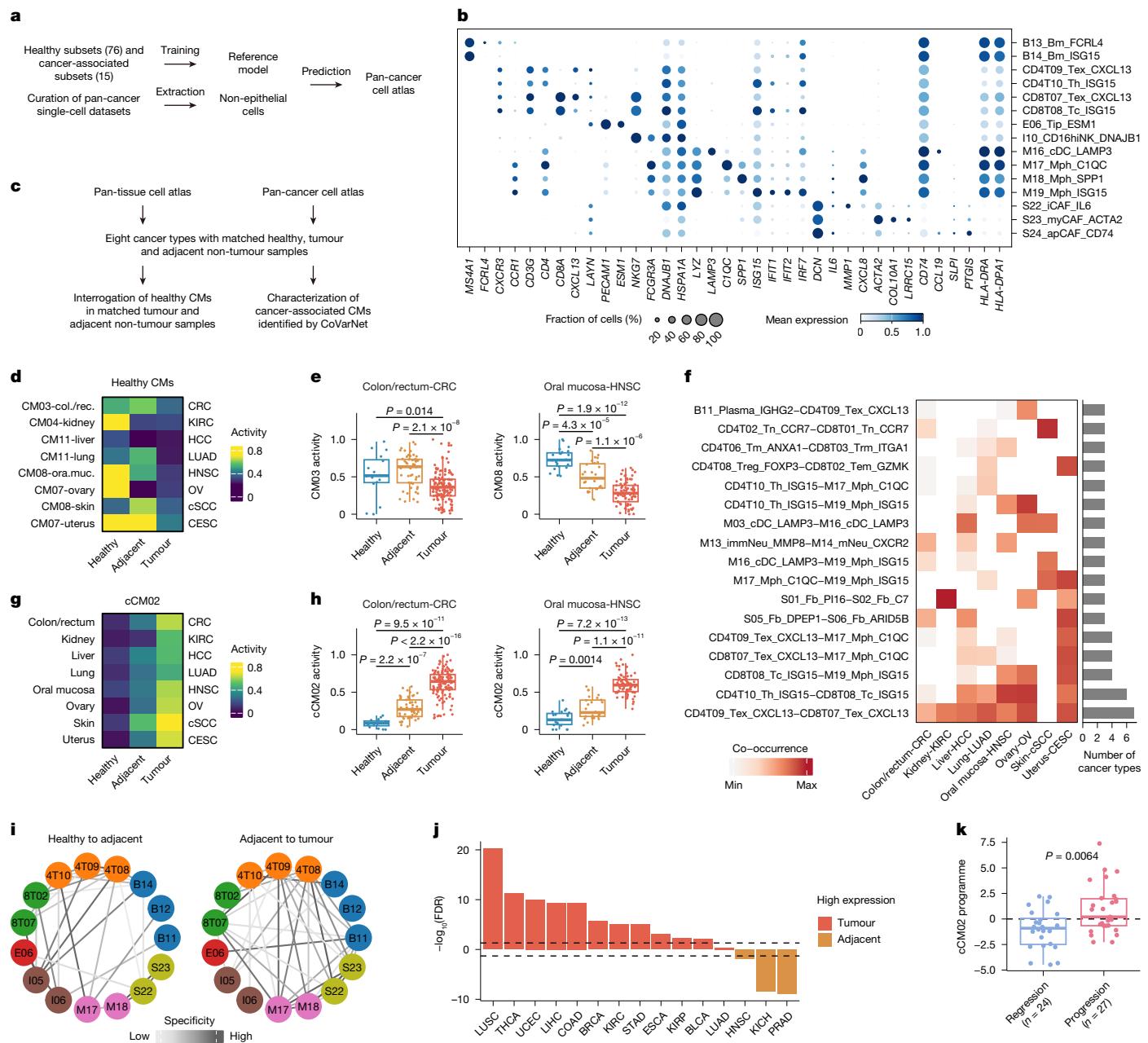
Given the extensive remodelling of multicellular ecosystems in cancer, we next examined the co-occurrence of cell subsets across individual cancer types. Our analysis revealed that cancer-associated subsets frequently co-occurred across multiple cancer types (Fig. 5f), indicating the emergence of a convergent multicellular ecosystem shared across cancers. To further explore this, we applied CoVarNet to the eight cancer types and identified four cancer-associated CMs (cCMs) (Extended Data Fig. 11b–d). Among these, cCM02, composed primarily of cancer-associated cell subsets, was enriched in tumour samples from most cancer types (Extended Data Fig. 12a,b), representing a cancer-associated TME ecosystem. Notably, cCM02 activity progressively increased from healthy to adjacent non-tumour, and then to tumour samples across cancer types (Fig. 5g,h and Extended Data Fig. 12c), underscoring its role as an indicator of tumour progression. We also observed increased co-occurrence of cell subsets within cCM02 in tumour samples compared with adjacent non-tumour samples, providing discernible evidence of tumour progression (Fig. 5i). Together, these findings highlight simultaneous rewiring of two types of multicellular ecosystem during tumour progression—marked by the loss of tissue-specific healthy organizations and the emergence of a convergent cancerous ecosystem (Extended Data Fig. 12d).

Further cytokine analysis for cCM02 revealed that key mediators of intercellular regulation included interferon, IL-18 and IL-15 (Extended Data Fig. 12e and Supplementary Table 11), aligning with a recent study highlighting CD8<sup>+</sup> T cell-derived IFN $\gamma$  as a critical modulator of the TME compared with TNF<sup>43</sup>. Additionally, DIALOGUE analysis identified an MCP associated with increased cCM02 activity in tumour samples, characterized by upregulation of S100 family member genes (*S100A2*, *S100A9* and *S100A8*)<sup>44</sup> across most cell subsets (Extended Data Fig. 12f and Supplementary Table 12). This TME programme was validated using external datasets from The Cancer Genome Atlas (TCGA) (Fig. 5j). To assess its clinical significance, we examined its expression in pre-invasive lung lesions from 51 individuals with known outcomes<sup>45</sup>. Remarkably, pre-invasive lesions that progressed to invasive lung cancer exhibited higher expression of this programme compared with those undergoing spontaneous regression (Fig. 5k), suggesting its potential for early diagnosis of cancer.

## Discussion

Understanding how diverse cell types coordinate to maintain tissue homeostasis and contribute to disease progression remains a fundamental challenge in biology. Here we present a computational framework for systematically identifying cross-tissue, co-occurring CMs and their rewiring in cancer. The pan-tissue and pan-cancer single-cell atlases that we curated represent valuable resources for the community. CoVarNet bridges the gap between well-characterized cellular diversity and the complex organization and function of tissues. By linking CMs to phenotypic data, we uncovered fundamental biological insights, highlighting CMs as a scaffold for studying multicellular organization across diverse contexts. Together, these findings illuminate core principles of multicellular ecosystems and advance our understanding of tissue-level coordination in health and disease, opening avenues for future research and potential therapeutic insights.

Methodologically, CoVarNet offers several advantages over existing strategies for identifying cellular niches<sup>26</sup> or spatial domains<sup>46</sup> using



**Fig. 5 | Rewiring of multicellular ecosystems in cancer.** **a**, Overview of the pan-cancer single-cell atlas. **b**, Expression of signature genes for cancer-associated cell subsets. **c**, Overview of multicellular analysis in cancer. **d**, Sample-averaged activities of dominant healthy CMs in healthy, adjacent non-tumour, and tumour samples across cancer types (Methods). CESC, cervical squamous cell carcinoma; col., colon; CRC, colorectal cancer; cSCC, cutaneous squamous cell carcinoma; HCC, hepatocellular carcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; ora.muc., oral mucosa; OV, ovarian cancer; rec., rectum. **e**, CM03 activities in CRC (left) and CM08 activities in HNSC (right) across healthy, adjacent non-tumour and tumour samples. Dots represent individual samples. Colon/rectum-CRC: healthy,  $n = 16$ ; adjacent,  $n = 51$ ; tumour,  $n = 132$ . Oral mucosa-HNSC: healthy,  $n = 23$ ; adjacent,  $n = 27$ ; tumour,  $n = 73$ . Two-tailed unpaired Wilcoxon tests. **f**, Left, heat map showing the co-occurrence of cell subset pairs in individual cancer types (Methods). Right, bar plot showing the number of cancer types in which individual pairs are detected. Only pairs detected in at least three cancer types are shown. **g**, Sample-averaged cCM02

activities in healthy, adjacent non-tumour and tumour samples across cancer types (Methods). **h**, cCM02 activities in CRC (left) and HNSC (right) across healthy, adjacent non-tumour and tumour samples. Dots represent individual samples, matching those in **e**. Two-tailed unpaired Wilcoxon tests. **i**, Dynamics of the co-occurring cCM02 network during tumour progression (Methods). **j**, Differential expression of the cCM02 programme between tumour and adjacent non-tumour samples across TCGA cancer types. BLCA, bladder urothelial carcinoma; BRCA, breast cancer; COAD, colon adenocarcinoma; ESCA, oesophageal carcinoma; KICH, kidney chromophobe; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUSC, lung squamous cell carcinoma; PRAD, prostate adenocarcinoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma. **k**, cCM02 programme expression in pre-invasive lesions with spontaneous regression or progression to cancer outcomes. Two-tailed unpaired Wilcoxon test. For box plots in **e,h,k**, the centre line represents the median, the box limits delineate the top and bottom quartiles, and the whiskers extend to the highest and lowest values within  $1.5 \times$  the interquartile range.

spatial data. First, spatial datasets are often limited by gene coverage or resolution, impeding comprehensive profiling of multicellular ecosystems. By contrast, our approach leverages single-cell transcriptomes

to define fine-grained CMs, which can then be mapped onto spatial data to capitalize on the strengths of both modalities. For instance, spatial mapping of CM02 and CM03 revealed distinct distributions

that were not apparent from spatial data alone. This framework enables integrative, multimodal analysis across a wide range of biological contexts. Second, many existing methods rely on spatial proximity to infer intercellular relationships, potentially overlooking broader coordination. By contrast, our approach—based on co-occurrence patterns—captures both local and distal multicellular interactions, which may be essential for understanding complex networks such as systemic immunity and cross-tissue regulation.

This study has several limitations. Our current framework does not explicitly incorporate epithelial cells, the extracellular matrix or microbiome components. Including these essential elements in future analyses will provide a more holistic perspective on tissue-level functional coordination. Additionally, integrating coordinated intercellular networks with intracellular regulatory circuits<sup>47</sup> holds promise for a more nuanced understanding of tissue functions. Expanding the analysis to larger cohorts and a broader range of phenotypes will further advance our understanding of multicellular ecosystems and their implications for translational medicine.

## Online content

Any methods, additional references, Nature Portfolio 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/s41586-025-09053-4>.

1. Luca, B. A. et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* **184**, 5482–5496.e5428 (2021).
2. Ramirez Flores, R. O., Lanzer, J. D., Dimitrov, D., Velten, B. & Saez-Rodriguez, J. Multicellular factor analysis of single-cell data for a tissue-centric understanding of disease. *eLife* **12**, e93161 (2023).
3. Jerby-Arnon, L. & Regev, A. DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data. *Nat. Biotechnol.* **40**, 1467–1477 (2022).
4. Mitchel, J. et al. Coordinated, multicellular patterns of transcriptional variation that stratify patient cohorts are revealed by tensor decomposition. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-024-02411-z> (2024).
5. Green, G. S. et al. Cellular communities reveal trajectories of brain ageing and Alzheimer's disease. *Nature* **633**, 634–645 (2024).
6. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
7. Dominguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
8. Eraslan, G. et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* **376**, eabl4290 (2022).
9. The Tabula Sapiens Consortium. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
10. Wang, D., Liu, B. & Zhang, Z. Accelerating the understanding of cancer biology through the lens of genomics. *Cell* **186**, 1755–1771 (2023).
11. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
12. Armingol, E., Baghdassarian, H. M. & Lewis, N. E. The diversification of methods for studying cell-cell interactions and communication. *Nat. Rev. Genet.* **25**, 381–400 (2024).
13. Hickey, J. W. et al. Organization of the human intestine at single-cell resolution. *Nature* **619**, 572–584 (2023).
14. Madissoon, E. et al. A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. *Nat. Genet.* **55**, 66–77 (2023).
15. Ji, A. L. et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497–514.e422 (2020).
16. He, S. et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).
17. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
18. Polanski, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
19. Sender, R. et al. The total mass, number, and distribution of immune cells in the human body. *Proc. Natl Acad. Sci. USA* **120**, e2308511120 (2023).
20. Blanco, E. et al. Age-associated distribution of normal B-cell and plasma cell subsets in peripheral blood. *J. Allergy Clin. Immunol.* **141**, 2208–2219.e2216 (2018).
21. Buechler, M. B. et al. Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
22. Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711 (2021).
23. Agace, W. W. & McCoy, K. D. Regionalized development and maintenance of the intestinal adaptive immune landscape. *Immunity* **46**, 532–548 (2017).
24. Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
25. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
26. Varrone, M., Tavernari, D., Santamaria-Martinez, A., Walsh, L. A. & Ciriello, G. CellCharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. *Nat. Genet.* **56**, 74–84 (2024).
27. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
28. Cui, A. et al. Dictionary of immune responses to cytokines at single-cell resolution. *Nature* **625**, 377–384 (2024).
29. Liu, X. et al. Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature* **567**, 525–529 (2019).
30. Albar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
31. Karakaslar, E. O. et al. Transcriptional activation of Jun and Fos members of the AP-1 complex is a conserved signature of immune aging that contributes to inflammaging. *Aging Cell* **22**, e13792 (2023).
32. Larson, M. H. et al. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat. Commun.* **12**, 2357 (2021).
33. Bonafe, M., Storci, G. & Franceschi, C. Inflamm-aging of the stem cell niche: breast cancer as a paradigmatic example: breakdown of the multi-shell cytokine network fuels cancer in aged people. *Bioessays* **34**, 40–49 (2012).
34. Kumar, T. et al. A spatially resolved single-cell genomic atlas of the adult human breast. *Nature* **620**, 181–191 (2023).
35. Cheng, S. et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792–809.e723 (2021).
36. Tang, F. et al. A pan-cancer single-cell panorama of human natural killer cells. *Cell* **186**, 4235–4251.e4220 (2023).
37. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).
38. Gao, Y. et al. Cross-tissue human fibroblast atlas reveals myofibroblast subtypes with distinct roles in immune modulation. *Cancer Cell* **42**, 1764–1783.e1710 (2024).
39. Yang, Y. et al. Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. *Cell* **187**, 4790–4811.e4722 (2024).
40. Aran, D. et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat. Commun.* **8**, 1077 (2017).
41. Jakubek, Y. A. et al. Large-scale analysis of acquired chromosomal alterations in non-tumor samples from patients with cancer. *Nat. Biotechnol.* **38**, 90–96 (2020).
42. Yarchoan, M. et al. PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. *JCI Insight* **4**, e126908 (2019).
43. Hoekstra, M. E. et al. Distinct spatiotemporal dynamics of CD8<sup>+</sup> T cell-derived cytokines in the tumor microenvironment. *Cancer Cell* **42**, 157–167.e159 (2024).
44. Bresnick, A. R., Weber, D. J. & Zimmer, D. B. S100 proteins in cancer. *Nat. Rev. Cancer* **15**, 96–109 (2015).
45. Teixeira, V. H. et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. *Nat. Med.* **25**, 517–525 (2019).
46. Zhao, E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).
47. Badia, I. M. P. et al. Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet.* **24**, 739–754 (2023).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

## Methods

### Single-cell data collection and preprocessing of healthy samples

To assemble a comprehensive pan-tissue cell atlas, we collected scRNA-seq datasets and conducted quality control procedures via the Scanpy<sup>48</sup> toolkit, as detailed in subsequent sections (Extended Data Fig. 1 and Supplementary Table 1). Default parameters were used unless otherwise specified.

**Data collection.** We included scRNA-seq datasets from adult samples that met the following criteria: (1) utilization of fresh, not frozen, samples; (2) inclusion of samples based on cell-type enrichment: (a) no cell-type enrichment; (b) a mixture of immune, epithelial, endothelial and stromal compartments; (c) enrichment for either immune or non-immune cell populations; and (3) generation of single-cell, not single-nucleus, data using the 10x Genomics platform. These criteria were implemented to minimize batch effects across the datasets<sup>49</sup>. Ultimately, a total of 33 datasets from 26 cohorts were included, collectively representing a cell atlas across 35 human tissues.

**Quality control.** To standardize datasets annotated with different versions of the human genome assembly, we limited the transcriptome to the common set of 21,812 genes found in the three most widely used 10x Genomics gene annotations, specifically GRCh38 (Ensembl 84), GRCh38 (Ensembl 93) and GRCh38 (GENCODE v32/Ensembl 98). Cells identified as low-quality or germ line in the original studies were excluded, and only cells meeting the following criteria were retained: 500–8,000 genes, 1,000–100,000 gene counts, and less than 20% mitochondrial gene counts. We applied Scrublet<sup>50</sup>, integrated into Scanpy, to each cohort and removed cells with a doublet score exceeding the 90th percentile across all cohorts. We then excluded samples with fewer than 50 high-quality cells. In the end, the analysis comprised a total of more than 700 samples that passed the stringent quality control measures.

**Preprocessing.** Beginning with the combined gene count matrix across all datasets, we derived the normalized gene expression matrix by normalizing total counts per cell (library size) using a scale factor of 10,000 followed by logarithmic transformation. Highly variable genes (HVGs) were then selected using the function `scanpy.pp.highly_variable_genes` with the following parameters: (`n_top_genes=2000`, `flavor = "cell_ranger"`, `batch_key = "datasetID"`). Notably, HVG selection was performed after removing specific genes, including immunoglobulin genes, T cell receptor genes, ribosome protein-coding genes, heat shock proteins-associated genes, and mitochondrial genes. Several confounding effects, including total gene counts per cell, the percentage of mitochondrial gene counts, and cell cycle were addressed, using the function `scanpy.pp.regress_out`. Finally, HVGs were centred and scaled among all cells.

### Single-cell data integration and annotation

To integrate these extensive datasets, we used the Scanpy toolkit with default parameters unless otherwise specified.

**Benchmarking integration methods.** To determine the best integration method for our datasets, we utilized scIB<sup>17</sup> to benchmark several widely used Python-based tools: BBKNN<sup>18</sup>, Harmony<sup>51</sup>, Scanorama<sup>52</sup> and deep learning-based scVI<sup>53</sup>, scANVI<sup>54</sup>, and SCALEX<sup>55</sup>. Among the 14 metrics in scIB, biological conservation for HVG and trajectory were not applicable, and the kBET metric was excluded owing to memory requirements exceeding 2 TB. Overall scores were calculated as a weighted mean (40/60) of batch correction and biological variance conservation. Importantly, we conducted two independent benchmarking analyses on the entire atlas and a subset atlas, respectively. In the end, BBKNN emerged as the top performer and was used for the integration of the pan-tissue datasets (Extended Data Fig. 2).

**Dataset integration.** Principal component analysis was performed on the centred and scaled HVG expression matrix to extract 50 principal components. BBKNN, integrated into Scanpy, was then executed with the dataset as the batch variable. The batch-corrected graph was then utilized to perform UMAP<sup>56</sup> for visualizing cells on a two-dimensional layout.

**Cell clustering and annotation.** We performed at least two levels of unsupervised cell clustering and annotation. The first level of clustering was performed using the function `scanpy.tl.leiden` with resolution = 0.1 followed by identification of differentially expressed genes (DEGs; log<sub>2</sub>-transformed fold changes >1, FDR < 0.05, Student's *t*-test). The eight broad cell types were identified on the basis of canonical markers and DEGs. We also received assistance with cell annotation from CellTypist<sup>7</sup>, an automated cell-type annotation tool, using the `Immune_All_High` and `Immune_All_Low` models. Subsequently, further clustering (second or more levels) was performed using context-specific resolutions to obtain several distinct cell subsets for each cell type. Epithelial cells were excluded from further clustering owing to their highly tissue-specific nature. In total, 2,293,951 high-quality cells from 706 samples across 317 donors were annotated into 76 non-epithelial subsets and 26 epithelia cell types.

**Hierarchical clustering of cell subsets.** We generated pseudo-bulk profiles for 76 non-epithelial cell subsets by averaging the gene expression of all cells within the same subsets. Next, unsupervised hierarchical clustering was performed using correlation distance and the `hclust` function (method = "ward.D"). The results were visualized using the `dendextend` R package.

### CoVarNet framework

We introduced CoVarNet, a computational framework designed to systematically unravel coordination among multiple cell types. CoVarNet identifies co-occurring CM networks by analysing the covariance in cell subset frequencies across various samples.

**CoVarNet overview.** CoVarNet uses input data on cell subset frequencies within each cell type and sample. It utilizes two parallel modules to jointly determine CM networks by connecting co-occurring subsets (nodes) through edges. The first module applies NMF to the cell subset frequency matrix, identifying factors that prioritize subsets based on their weights. The top subsets of each factor act as co-occurring nodes in a single CM network. The second module identifies specifically correlated subset pairs, which act as potential edges. Multiple CM networks are then constructed to interconnect co-occurring nodes via these potential edges, followed by topological and statistical evaluations.

**Input frequency matrix.** To ensure comparability of cell-subset frequencies across tissues and clinical specimens, we included only samples without cell-type enrichment or those from mixtures of the four cell compartments. Samples with fewer than 50 high-quality cells were excluded. For each eligible sample, we computed the frequencies of cell subsets within their corresponding cell types. Min-max normalization was applied to correct the frequency matrix, mitigating the impact of varying numbers of cell subsets across different cell types. Thus, a corrected frequency matrix, ranging from 0 to 1, was utilized in the CoVarNet procedure. Specifically, we generated a frequency matrix consisting of 76 subsets (rows) and 510 samples (columns) for the pan-tissue atlas.

**NMF.** NMF has been used in the analysis of single-cell<sup>57–60</sup> and spatial<sup>61,62</sup> expression data to extract gene expression programmes. In this study, CoVarNet applies NMF to the frequency matrix to decipher cellular co-occurring programmes, specifically using the nsNMF method with ranks from 2 to 20, as implemented in the NMFR package<sup>63</sup>. To ensure

# Article

robustness, we conducted 30 runs to derive a consensus output, consisting of  $k$  factors and their activities in each sample. Specifically, the top ten subsets of each factor were used as co-occurring node candidates in a single CM network for the pan-tissue analysis.

**Rank selection.** To determine the optimal rank for NMF analysis, we used the cophenetic correlation coefficient (CCC) as the evaluation index, in accordance with practices from previous reports<sup>1</sup>. CCC is used to quantify classification stability, with values ranging from 0 to 1 and 1 indicating maximum stability<sup>64</sup>. We denoted the CCC at rank  $k$  as  $\rho_k$  and established a procedure tailored to this context for consistent stability based on the following criteria: (1)  $\rho_{k-2} < \rho_{k-1} < \rho_k$ ; (2)  $\rho_k > \rho_{k+1}$ . Among a set of ranks meeting these criteria, the optimal rank was then identified as the one at which CCC is maximized. The optimal rank selected for the pan-tissue atlas was 12 (Extended Data Fig. 3a,b).

**Specifically correlated subset pairs.** CoVarNet utilizes Pearson correlation coefficients to assess whether any two cell subsets co-occur. For a given set of  $s$  cell subsets, pairwise correlation tests are performed based on the frequency matrix, resulting in an  $s \times s$  correlation coefficient matrix (denoted  $R$ ). To quantify the specificity of correlations, an indicator is defined. For each element  $r_{ij}$  ( $i < j$ ) in  $R$ , its background set  $S_{ij}$  and specificity index  $\text{Spec}(r_{ij})$  are defined as:

$$S_{ij} = \{r_{ik} | k \neq i\} \cup \{r_{kj} | k \neq j\}$$
$$\text{Spec}(r_{ij}) = \frac{|\{r \in S_{ij} | r \leq r_{ij}\}|}{|S_{ij}|}$$

In other words, the specificity index is defined as the fraction of elements in the background set that do not exceed  $r_{ij}$ . The specificity cutoff is determined by an automatic method. If  $n$  and  $N$  represent the assumed number of subsets in each CM and the total number of subsets, then the specificity cutoff Cutoff( $n, N$ ) will be determined as:

$$\text{Cutoff}(n, N) = 1 - \frac{(n-1) \times 2 - 1}{(N-1) \times 2 - 1}$$

This approach enables a balanced assessment of the number of subsets within CMs and their co-occurrence. Specifically correlated subset pairs are determined jointly by the correlation (coefficient and FDR) and specificity. We generated 147 pairs for the pan-tissue atlas. These pairs were visualized as a global network (Supplementary Fig. 4).

**Construction, evaluation and visualization of CM networks.** For each NMF factor, the top subsets are designated as potential nodes, and edges connect specifically correlated subset pairs, removing isolated nodes. In each constructed CM network, the connectivity score is calculated as the ratio of observed edges to the total possible edges among all nodes within that network. The statistical significance of this score is assessed using a permutation test ( $n = 10,000$ ) on the node labels. We used the igraph R package to visualize the CM networks, with nodes colour-coded by cell type and edge colour gradients scaled to reflect specificity.

**CMT classifications of samples.** The CM activities in individual samples are measured by the coefficient matrix from the NMF procedure, with the sum of activities for all CMs equalling 1 for each sample. Each sample was assigned a CMT label based on its most abundant CM. For instance, if a sample exhibited the highest activity of CM01 among all CMs, it was labelled as CMT01. All healthy single-cell samples used across tissues were stratified into 12 CMT groups (Extended Data Fig. 3e).

## Integrative analysis of scRNA-seq and GTEx RNA-seq

We utilized GTEx<sup>24</sup> RNA-seq datasets to validate the CMs defined by single-cell data (Extended Data Fig. 4).

**RNA-seq data preprocessing.** We retrieved gene transcripts per million (TPM) and metadata for 17,382 bulk RNA-seq samples from the GTEx Portal (V8 release)<sup>65</sup>. Samples derived from cell lines were excluded, and the ‘Cervix uteri’ category was merged into the ‘Uterus’ category for consistency. To ensure consistency, only tissues represented in the single-cell cohort were retained, narrowing down to a total of 12,240 samples spanning 23 tissues for further analysis. Gene expression data were re-normalized to a uniform library size of 10,000 and log-transformed for comparability with single-cell data.

**CMT classifications of RNA-seq samples.** We began by identifying DEGs among pseudo-bulk CMT samples. The top ten DEGs, ranked by fold change, were designated as CMT signature genes. Utilizing the Seurat R package<sup>66</sup>, we applied the AddModuleScore function to calculate scores for individual RNA-seq samples based on these CMT signature sets. All negative scores were adjusted to zero, and 2.3% (278 out of 12,240) of samples with the highest score less than 0.2 were excluded to ensure robust classification. Ultimately, the remaining 11,962 samples were categorized into 12 distinct CMTs, facilitating a detailed examination of CM representation across the analysed tissues.

## Tissue prevalence of cell subsets and CMs

To assess the prevalence of cell subsets across tissues, we compared observed (o) to expected (e) cell numbers for each subset-tissue combination, expressed as  $R_{o/e} = \text{observed/expected}$ , following established methods<sup>35,38,39</sup>. Expected cell numbers for each subset-tissue combination were derived from the Chi-square test, with enrichment defined as  $R_{o/e} > 1$  (Fig. 1e and Supplementary Fig. 2). For the assessment of each CM, we computed tissue-level CM activities by averaging its activity across all samples within each tissue. The  $R_{o/e}$  ratio indicated the tissue distribution of CM profiles (Fig. 2c). To compare CM enrichment across 23 overlapping tissues between bulk and single-cell analyses, we independently calculated  $R_{o/e}$  ratios for each data type and combined them for comparison (Extended Data Fig. 4f). Results were visualized using the ComplexHeatmap R package<sup>67</sup>.

## Analysis of spatial transcriptomics data

**Data collection.** We gathered published spatially resolved transcriptomics datasets (Visium and Xenum) of various human tissues and cancer types. Detailed accession numbers and references for these datasets are provided (Supplementary Table 3).

**Cell subset identification.** For deconvoluting spatial transcriptomics data, we utilized cell2location<sup>25</sup>, a Bayesian model capable of accurately resolving fine-grained cell types within spatial data. Utilizing both healthy and cancer datasets, we used the corresponding integrated scRNA-seq data as a reference to obtain cell-type signatures. Prior to this process, the scRNA-seq data was subsampled to 1,000 cells per cell subset. In cases where cell types comprised fewer than 1,000 cells, all available cells were included. Following the recommended guidelines of cell2location, we set  $N = 5$  as the expected cell abundance per spot and  $\alpha_y = 20$  to regularize within-experiment variation in RNA detection sensitivity. The output yielded the expected cell abundance per cell subset in each spot.

**CM activities.** To quantify and visualize CMs in spatial transcriptomics, we aggregated the abundance of the component cell subsets within each CM, applying weights derived from NMF results. The resulting CM activities were then scaled to a uniform standard, with the 99th percentile value across all CMs being set to 1, allowing for a direct comparison of their relative magnitudes.

**Colocalization scores.** To assess the colocalization of cell-subset components within CMs across Visium spots, we calculated the colocalization score for individual spatial sections. For each CM, we calculated

Spearman correlation coefficients between subset pairs where at least one subset is within this CM, resulting in a set of correlation coefficients denoted as  $S$ . The median correlation coefficient within the CM is termed  $r$ . The colocalization score for each CM is defined as the proportion of correlations in  $S$  that are less than or equal to  $r$ , providing a measure of colocalization relative to global contexts.

**Aggregation scores.** To assess the regional aggregation of cell-subset components within CMs in spatial transcriptomics, we utilized the global bivariate Moran's  $I^8$  using the *spdep* R package. Similar to the colocalization score, the aggregation scores were calculated using global Moran's  $I$  instead of the correlation coefficient.

**Cellular niches.** To provide orthogonal validation for the identified CMs, we used CellCharter<sup>26</sup> to identify cellular niches, clustering Visium spots based on both gene expression and spatial information to enable spatially informed niche categorization. This analysis was performed for each sample independently to mitigate batch effects, allowing cross-validation of results across samples from the same tissues.

**Xenium analysis.** The Xenium platform enables *in situ* characterization of hundreds to thousands of genes in cells and tissues with ultra-precise single-cell spatial imaging. Using published intestinal Xenium data, we characterized the spatial locations of CM02 and CM03. We first designed a gene panel to distinguish multiple cell subsets within these CMs, with gene transcript density serving as a proxy for CM intensity (Extended Data Fig. 5i). Spatial regions of the tissue (epithelium or lamina propria) were identified on the basis of  $k$ -means clustering ( $k = 2$ ) in the original dataset. To assess CM distribution across tissue sub-regions, we selected six different areas within the intestinal mucosa and measured CM intensities, providing a spatially resolved, single cell-resolution distribution of CMs.

### Cell–cell communication analysis

To disentangle complex cellular crosstalk within and across CMs, we performed ligand–receptor-mediated cell–cell communication analysis using single-cell data with the CellPhoneDB Python package<sup>11,27</sup>.

**Global analysis agnostic to CMs.** Considering the extensive number of cells, we conducted subsampling to equalize the contribution of each cell subset. Specifically, we subsampled the number of cells to 1,000 cells for each cell subset. However, if the total cell count for certain subsets did not exceed 1,000, all cells were included in the analysis. This approach aimed to ensure that the null distribution accurately represented all cell subsets, avoiding bias towards cell subsets with larger cell numbers. Subsequently, the CellPhoneDB procedure was used for statistical inference of cell–cell interaction specificity, allowing the derivation of cell–cell interaction counts among different cell types (Extended Data Fig. 6c) or CMs (Extended Data Fig. 6d) by averaging the results of corresponding cell subsets. We also validated these results using an alternative tool, CellChat<sup>69</sup> (Extended Data Fig. 6e).

**Comparative analysis informed by CMs.** To explore the effect of tissue microenvironments (CMs) on cellular crosstalk within CMs, we conducted CellPhoneDB analysis using two groups of samples with high or low CM activities, respectively. For each CM, the ‘High’ group included all samples in which that CM showed the highest activity among all CMs, while other samples were used as the ‘Low’ group. Following the previously defined CMTs in ‘CMT classifications of samples’, taking CM01 as an example, all samples labelled as CMT01 were used as the High group, while other samples were used as the Low group. Subsequently, comparative analysis between the two groups focused on cell subsets that were components of each CM (Fig. 3g).

### Cytokine response analysis

The recently reported Immune Dictionary provides a comprehensive overview of cell-type-specific responses to 86 cytokines<sup>28</sup>. Leveraging this foundational knowledge, we explored intercellular crosstalk within CMs through cytokine responses. CM07 and CM10 were excluded as they lack immune cell subsets, while CM12 had no significant cytokine outputs.

**CM-dependent DEGs of cell subsets.** The tissue microenvironments exerted a broad influence on cell-subset phenotypes in a CM-dependent manner. Specifically, for each immune subset, we identified its DEGs ( $\log_2$ -transformed fold changes  $>\log_2(1.2)$ ,  $FDR < 0.05$ , Student's  $t$ -test) in samples from corresponding CMTs when compared with those from other CMTs, utilizing the Scanpy toolkit. These identified DEGs were interpreted as responses to cytokines within the tissue microenvironments associated with CMs.

**Cytokine response inference.** First, we built a database of cytokine signatures for each cell type, using supplementary table 3 from the Immune Dictionary publication<sup>28</sup>. Of note, we performed one-to-one conversion of mouse and human ortholog genes based on the Mouse Genome Informatics (MGI) database<sup>70</sup> ([http://www.informatics.jax.org/downloads/reports/HMD\\_HumanPhenotype.rpt](http://www.informatics.jax.org/downloads/reports/HMD_HumanPhenotype.rpt)). Subsequently, we conducted cell-type-aware immune response enrichment analysis using the hypergeometric test ( $FDR < 0.05$ ) through the enricher function in the clusterProfiler R package<sup>71</sup>.

**CM-specific cytokine networks.** To visualize cytokine-mediated multicellular regulation, we constructed cytokine networks by considering both cytokine production and response in a CM-specific manner. For cytokines to be considered in a CM, the normalized expression value of the cytokine gene needed to exceed 0.1 in at least one of non-responsive cell subsets. In the case of heteromeric cytokines or cytokine complexes with two subunits, each subunit was separately represented. The igraph R package was used to generate visual representations of the cytokine networks.

### Association analysis of CM activity and phenotypic factors

**Tissues.** We first examined the association between CM activities and tissues, excluding tissues with fewer than five samples. For each CM, we fitted a linear model, with adjusted  $R^2$  indicating the proportion of variance explained. The FDR of the  $F$  test was reported to ensure the robustness of the results. Given the strong tissue preferences of CMs, subsequent analyses focused on tissue-specific associations.

**Sexes and age groups.** In non-reproductive tissues, CM activities were compared between samples from male and female participants, with FDR-adjusted significance. Samples lacking age information were excluded from the analysis. Age data were categorized into groups: <35, 35–39, 40–49, 50–59, 60–69 and 70–85 years. For the breast dataset (D03) with over 100 samples (Supplementary Table 1), age was categorized as <50 or  $\geq 50$  years. Associations between CM activities and age groups were assessed, with adjustments for statistical significance. Specifically, we also examined associations between immune cell-subset frequencies and age groups in the spleen.

**Additional phenotypes.** We performed further association analyses between CMs and specific phenotypic factors. We analysed CM09 in relation to alcohol consumption in the lymph node, CM06 and CM11 with childhood tuberculosis in the lung, CM12 with menopause in the breast, and CM07 with menstrual cycle phases in the uterus.

# Article

## CM05 regulators in the spleen

**Inferring regulons.** We used the pySCENIC<sup>30,72</sup> pipeline to infer regulons for the four subsets (B03, B05, CD4T03, and I06) within CM05, performing the analysis separately for the three cellular lineages (B cells, CD4<sup>+</sup> T cells, and innate lymphoid cells). For each subset, regulons were ranked on the basis of their regulon specificity scores (RSS), and the top 50 regulons with the highest RSS were selected for each cell subset. Seventeen regulons were shared among the four subsets.

**Activities and target genes of shared regulons.** Sample-level activities for shared regulons were determined by averaging cellular activities across all cells in each sample. The mean activity of each regulon was then compared across age-stratified sample groups. Target genes of shared regulons were compared across lineages, and a regulatory network was constructed to illustrate their interrelationships.

## MCP analysis

**MCP identifications.** As different cell types within the same CMs tend to be exposed to similar tissue microenvironments, we hypothesized that they might exhibit coordinated responses. To investigate this, we utilized a method called DIALOGUE<sup>3</sup> to map MCPs for CMs. This procedure involved setting the parameter ( $k = 3$ ) and assessing the association between the MCPs identified and other phenotypes. This analysis was applied to CM12 in the breast and CM08, and resulted MCPs were termed as CM12 programme and CM08 programme.

## Comparison between CM08 programme and other signatures.

To compare CM08 programme and inflammatory and cytotoxic signatures, we calculated their overall expression in external RNA-seq data, as described previously<sup>73,74</sup>. Specifically, RNA-seq datasets of samples from individuals with advanced melanoma following anti-PD-1 therapies<sup>75</sup> were downloaded from <https://github.com/ParkerICI/MORRISON-1-public>. The inflammatory signature genes of immune cells are defined as *CD3D*, *IDO1*, *CIITA*, *CD3E*, *CCL5*, *GZMK*, *CD2*, *HLA-DRA*, *CXCL13*, *IL2RG*, *NKG7*, *HLA-E*, *CXCR6*, *LAG3*, *TAGAP*, *CXCL10*, *STAT1* and *GZMB*<sup>76</sup>. The cytotoxic signature genes are defined as *IFNG*, *GZMA*, *GZMB*, *PRF1*, *GZMK*, *ZAP70*, *GNYL*, *FASLG*, *TBX21*, *EOMES*, *CD8A*, *CD8B*, *CXCL9*, *CXCL10*, *CXCL11*, *CX3CL1*, *CCL3*, *CCL4*, *CX3CR1*, *CCL5*, *CXCR3*, *NKG7*, *CD160*, *CD244*, *NCR1*, *KLRC2*, *KLRK1*, *CD226*, *GZMH*, *ITK*, *CD3D*, *CD3E*, *CD3G*, *TRAC*, *TRBC1*, *TRBC2*, *CD28*, *CD5*, *KIRDL4*, *FGFBP2*, *KLRF1*, *SH2D1B* and *NCR3* (ref. 77).

## Inflammatory scores of CM12 subsets in the breast

We calculated the sample-level inflammatory scores for immune and fibroblast subsets within CM12. Fibroblasts and immune subsets were scored using the corresponding inflammatory gene sets. The inflammatory signature genes of fibroblasts are defined as *PLAU*, *CHI3L1*, *MMP3*, *IL1RI*, *IL13RA2*, *TNFSF11*, *MMP10*, *OSMR*, *IL11*, *STRA6*, *FAP*, *WNT2*, *TWIST1* and *IL24* (ref. 78). The inflammatory signature genes of immune cells are defined as above. Specifically, we first calculated the average gene expression among all cells of individual subsets for each sample. Subsequently, we used the R package AUCCell<sup>30</sup> to calculate the sample-level inflammatory scores.

## Menopausal trajectory analysis

**Discovery cohort.** We constructed a menopausal trajectory in the breast (dataset D03; Supplementary Table 1) based on the frequencies of cell subsets within CM12, following the methodology described in a recent study<sup>5</sup>. To mitigate the impact of frequency differences between cell subsets, we applied z-score normalization to correct the frequency matrix. Subsequently, we computed the k-neighbourhood and performed clustering for breast samples using the function scanpy.

pp.neighbours and scanpy.tl.leiden with default parameters. To model trajectories along menopause, we performed PHATE<sup>79</sup> with  $a = 40$ , followed by pseudotime analysis using the Palantir<sup>80</sup> standard pipeline. The starting point was defined as the cluster with a high proportion of pre-menopausal samples.

**Validation cohort.** The Reed dataset<sup>81</sup> was used to validate the menopausal trajectory. For the epithelial-enriched ('organoid') samples in the dataset, cell subset annotation was performed with CellTypist<sup>7</sup>, using previously annotated gene expression profiles of breast tissue as a reference. The cell subset frequency matrix was subsequently input into the trajectory analysis as described above.

## Pan-cancer single-cell atlas

To disentangle the rewiring of CMs along malignant progression, we constructed a pan-cancer transcriptomic atlas at single-cell resolution (Fig. 5a).

**Data collection and preprocessing.** Following the criteria set for healthy datasets, we selectively included scRNA-seq datasets from fresh (not frozen) samples without cell-type enrichment, generated using 10x Genomics single-cell (not single-nucleus) platforms. One exception is the ESCC\_GSE160269 cohort, where samples represent a mixture of immune, epithelial, endothelial, and stromal compartments. Quality control and other preprocessing procedures were also applied consistently with healthy datasets. In total, more than 1,000 samples from 48 datasets were incorporated, collectively forming a cell atlas across 29 human major cancer types (Extended Data Fig. 10 and Supplementary Table 7).

**Data integration, cell clustering and initial annotation.** Following the methodology applied to healthy datasets, we performed dataset integration using BBKNN. Unsupervised clustering for all cells was carried out using the scanpy.tl.leiden function with a resolution of 0.1. Subsequently, we identified the eight broad cell types based on canonical markers.

**Supervised cell annotation.** To accurately determine cell identities in cancerous samples, we utilized a transfer-learning-based strategy for cell subset annotation. Initially, we curated a single-cell reference dataset that encompassed 76 non-epithelial cell subsets identified in healthy data and 15 cancer-associated subsets identified in various cancer types<sup>35–39</sup> (Supplementary Table 9). The number of cells of each subset was subsampled to 1,000, unless the total cell count did not exceed 1,000. Subsequently, a transformer-based reference model was trained on the reference dataset. Following this, non-epithelial cells from the pan-cancer atlas were annotated using the reference model. These procedures were executed using TOSICA, a multi-head self-attention model, enabling interpretable cell-type annotation<sup>82</sup>. The epoch number for prediction was selected as 15 (Supplementary Fig. 8c). Cells with predicted probabilities <0.5 were removed (Supplementary Fig. 8d). In the end, a total of 3,038,535 high-quality cells from 1,062 samples of 717 donors were well annotated to be 91 non-epithelial subsets.

## Rewiring of multicellular ecosystems in cancer

**Samples used.** To compare multicellular dynamics during tumour progression, we included eight cancer types that had at least three samples from each condition (healthy, adjacent non-tumour and tumour).

**Interrogation of healthy CMs.** To quantify healthy CMs across different conditions, we aggregated the abundance of the component cell subsets within each CM, applying weights derived from NMF results. The resulting CM activities were then rescaled to range from 0 to 1. For each cancer type, only the most dominant CM was considered.

**Co-occurrence of cell subsets in individual cancer types.** For each cancer type, we derived all specifically correlated cell-subset pairs using the correlation analysis module of CoVarNet. Compared with pan-tissue or pan-cancer analysis, the following more stringent cutoffs are used: coefficients  $>0.5$ , FDR  $<0.05$  and specificity  $>0.95$ . Only identified specifically correlated subset pairs are used to perform comparison across eight cancer types.

**Identification of cCMs using CoVarNet.** Specifically, we generated a frequency matrix consisting of 91 subsets and 955 samples for the pan-cancer atlas. Specifically, the top 15 subsets of each factor were used as co-occurring node candidates in a single CM network for the pan-cancer analysis.

**Interrogation of cCM02.** To quantify cCMs across different conditions, we measured cCM02 activities using the same procedure as in healthy CMs.

### cCM02 analysis

**Co-occurring network.** To explore the dynamics of cCM02 during tumour progression, we constructed two co-occurring networks of which nodes are the same as original nodes in cCM02, while edges were recalculated in two scenarios. One used healthy and adjacent non-tumour samples, while the other used tumour and adjacent non-tumour samples. Thus, CM networks were constructed using unaltered nodes and new edges.

**Cytokine analysis.** For each cell-subset component of cCM02, we identified its DEGs ( $\log_2$ -transformed fold changes  $>\log_2(1.2)$ , FDR  $<0.05$ , Student's *t*-test) in tumour samples compared with adjacent non-tumour samples. These identified DEGs were used to perform cytokine analysis as described above.

**MCP analysis.** We conducted MCP identification using only tumour and adjacent non-tumour samples. Overall expression of MCP were calculated for the RNA-seq datasets from TCGA portal and microarray data of pre-invasive lung lesions<sup>45</sup>. TCGA datasets were downloaded using the TCGAbiolinks<sup>83</sup> R package. Only projects with more than ten tumour or adjacent non-tumour samples were included. Datasets of pre-invasive lung lesions were downloaded from <https://github.com/ucl-respiratory/preinvasive>.

### Reporting summary

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

### Data availability

This study conducted a comprehensive analysis based on curated datasets from previously published studies without generating new primary data. Curated pan-tissue and pan-cancer scRNA-seq datasets can be downloaded from Zenodo (<https://zenodo.org/records/15169362>)<sup>84</sup> and can be interactively explored at <http://cm.cancer-pku.cn>. Spatially resolved transcriptomics datasets used are listed in Supplementary Table 3. Source data are provided with this paper.

### Code availability

The source code of CoVarNet has been uploaded to the Github repository <https://github.com/QiangShiPKU/CoVarNet>. Other analysis code has been uploaded to Zenodo (<https://zenodo.org/records/14175262>)<sup>85</sup>.

48. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
49. Mereu, E. et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* **38**, 747–755 (2020).
50. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e289 (2019).
51. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
52. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
53. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
54. Xu, C. et al. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).
55. Xiong, L. et al. Online single-cell data integration through projecting heterogeneous datasets into a common cell-embedding space. *Nat. Commun.* **13**, 6118 (2022).
56. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 37–44 (2018).
57. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e1624 (2017).
58. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e1817 (2019).
59. Barkley, D. et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat. Genet.* **54**, 1192–1201 (2022).
60. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
61. Townes, F. W. & Engelhardt, B. E. Nonnegative spatial factorization applied to spatial genomics. *Nat. Methods* **20**, 229–238 (2023).
62. Rodrigues, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
63. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinf.* **11**, 367 (2010).
64. Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl Acad. Sci. USA* **101**, 4164–4169 (2004).
65. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
66. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
67. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
68. Thioulouse, J., Chessel, D. & Champely, S. Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environ. Ecol. Stat.* **2**, 1–14 (1995).
69. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
70. Blake, J. A. et al. Mouse Genome Database (MGD): knowledgebase for mouse-human comparative biology. *Nucleic Acids Res.* **49**, D981–d987 (2021).
71. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
72. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
73. Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997.e924 (2018).
74. Jerby-Arnon, L. et al. Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. *Nat. Med.* **27**, 289–300 (2021).
75. Campbell, K. M. et al. Prior anti-CTLA-4 therapy impacts molecular characteristics associated with anti-PD-1 response in advanced melanoma. *Cancer Cell* **41**, 791–806.e794 (2023).
76. Ayers, M. et al. IFN- $\gamma$ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
77. Bagaev, A. et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* **39**, 845–865.e847 (2021).
78. Smillie, C. S. et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* **178**, 714–730.e722 (2019).
79. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
80. Setty, M. et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
81. Reed, A. D. et al. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. *Nat. Genet.* **56**, 652–662 (2024).
82. Chen, J. et al. Transformer for one stop interpretable cell type annotation. *Nat. Commun.* **14**, 223 (2023).
83. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
84. Shi, Q. & Chen, Y. Data for “Cross-tissue multicellular coordination and its rewiring in cancer” [Dataset]. Zenodo <https://doi.org/10.5281/zenodo.15169362> (2025).
85. Shi, Q. & Chen, Y. Code for “Cross-tissue multicellular coordination and its rewiring in cancer”. Zenodo <https://doi.org/10.5281/zenodo.14175262> (2024).

**Acknowledgements** The authors thank members of the Z.Z. laboratory for insightful discussion. This work was supported by the National Key Research and Development Program of China (2023YFF1204700), the National Natural Science Foundation of China (62203019, 92459001 and L2424217) and the Fundamental Research Funds for the Central Universities (Peking University). Q.S. was supported in part by the National Postdoctoral Fellowship Program of China (GZB20230023), the China Postdoctoral Science Foundation (2021TQ0012, 2022M720246 and 2023M740093), the Peking University Boya Postdoctoral Fellowship, the Postdoctoral Fellowship of Peking-Tsinghua Center for Life Sciences and the Boehringer-Ingelheim Postdoctoral Fellowship Program. Part of the analysis was performed on the High Performance Computing Platform of the Center for Life Sciences (Peking University).

# Article

**Author contributions** Z.Z. and Q.S. conceived and designed the project. Q.S. and Y.C. carried out data curation with the help of S.Q., Y.Y. and Y.G. Q.S., Y.C. and Y.L. conducted the computational analysis with the help of S.Q. D.W., L.Z., Y.G. and Y.Y. contributed valuable discussions. Q.S. and Z.Z. wrote the manuscript with input from all authors.

**Competing interests** Z.Z. is a founder of Analytical BioSciences and is a board member for InnoCare Pharma. The other authors declare no competing interests.

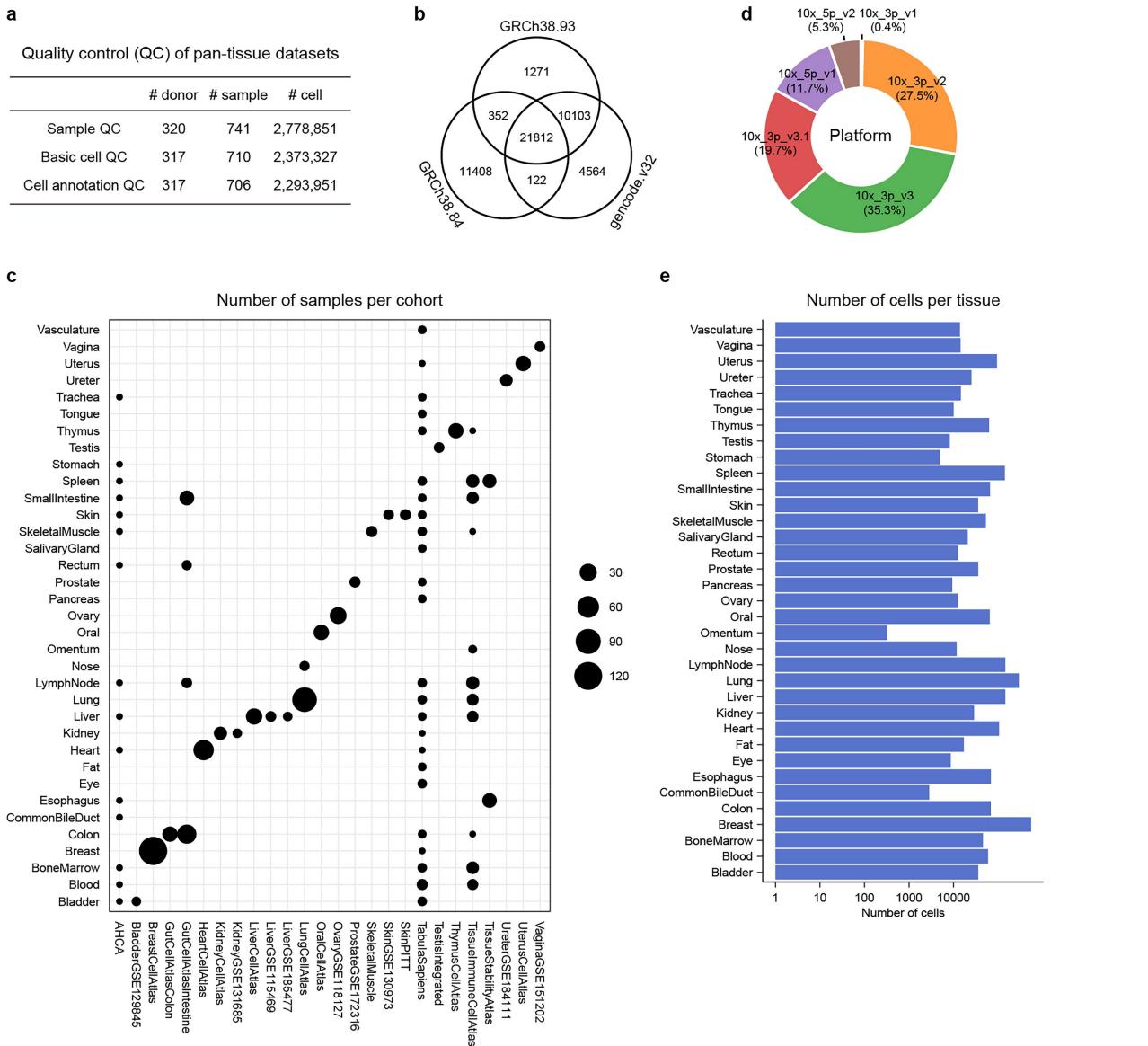
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-09053-4>.

**Correspondence and requests for materials** should be addressed to Qiang Shi or Zemin Zhang.

**Peer review information** *Nature* thanks Sarah Teichmann, who co-reviewed with Simon Koplev; Itai Yanai; and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



#### Extended Data Fig. 1 | Overview of pan-tissue single-cell datasets.

(a) Statistics of donor, sample, and cell numbers before and after QC for scRNA-seq datasets. (b) Statistics of gene numbers. The intersecting gene sets of the three commonly used versions of 10x Genomics gene annotations are used for

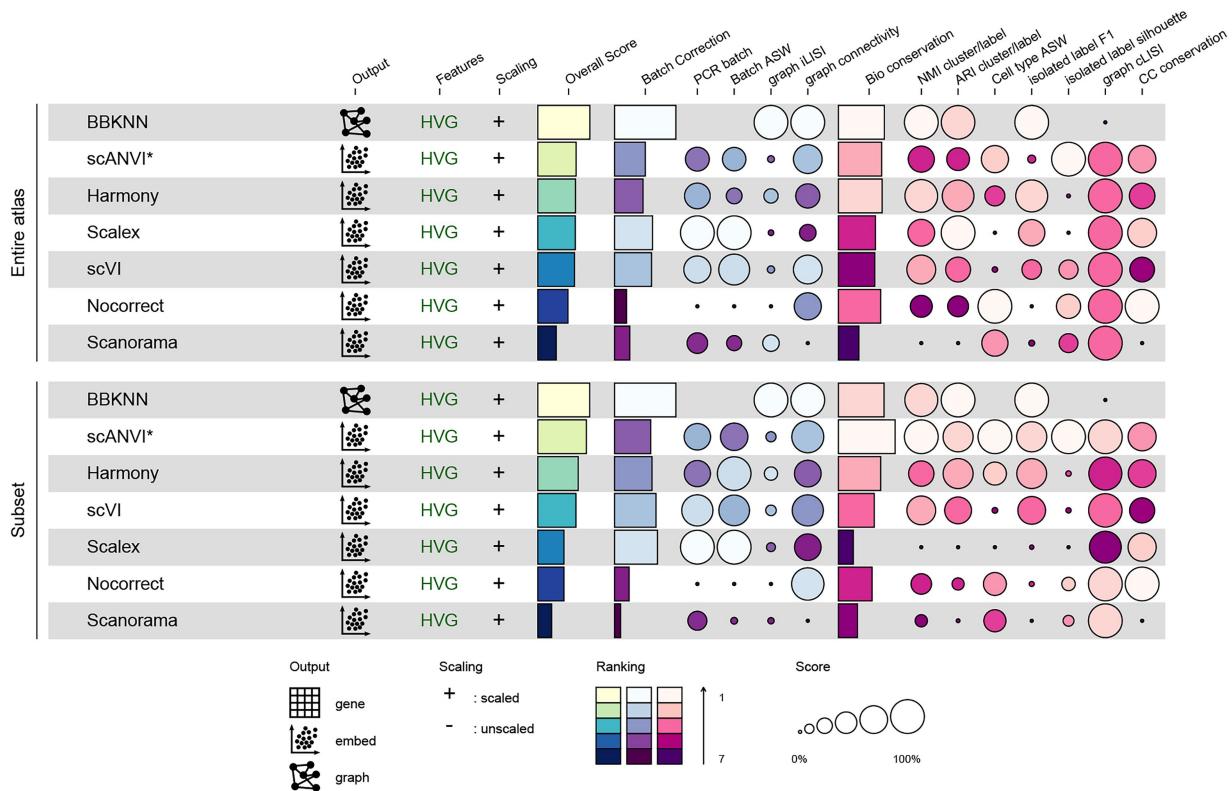
data integration. (c) Dot plot showing the number of samples per cohort.

(d) Pie chart showing the percentages of cells profiled by different 10x Genomics platforms. (e) Bar plot showing the number of cells per tissue.

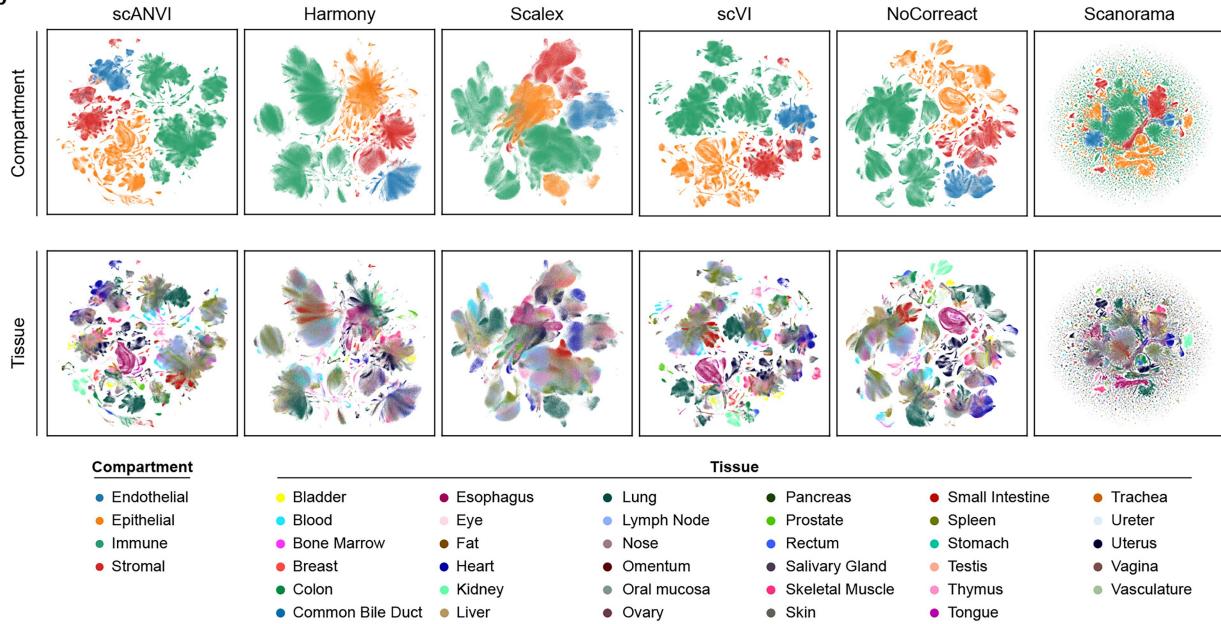
# Article

a

Benchmarking of integration methods using the scIB platform

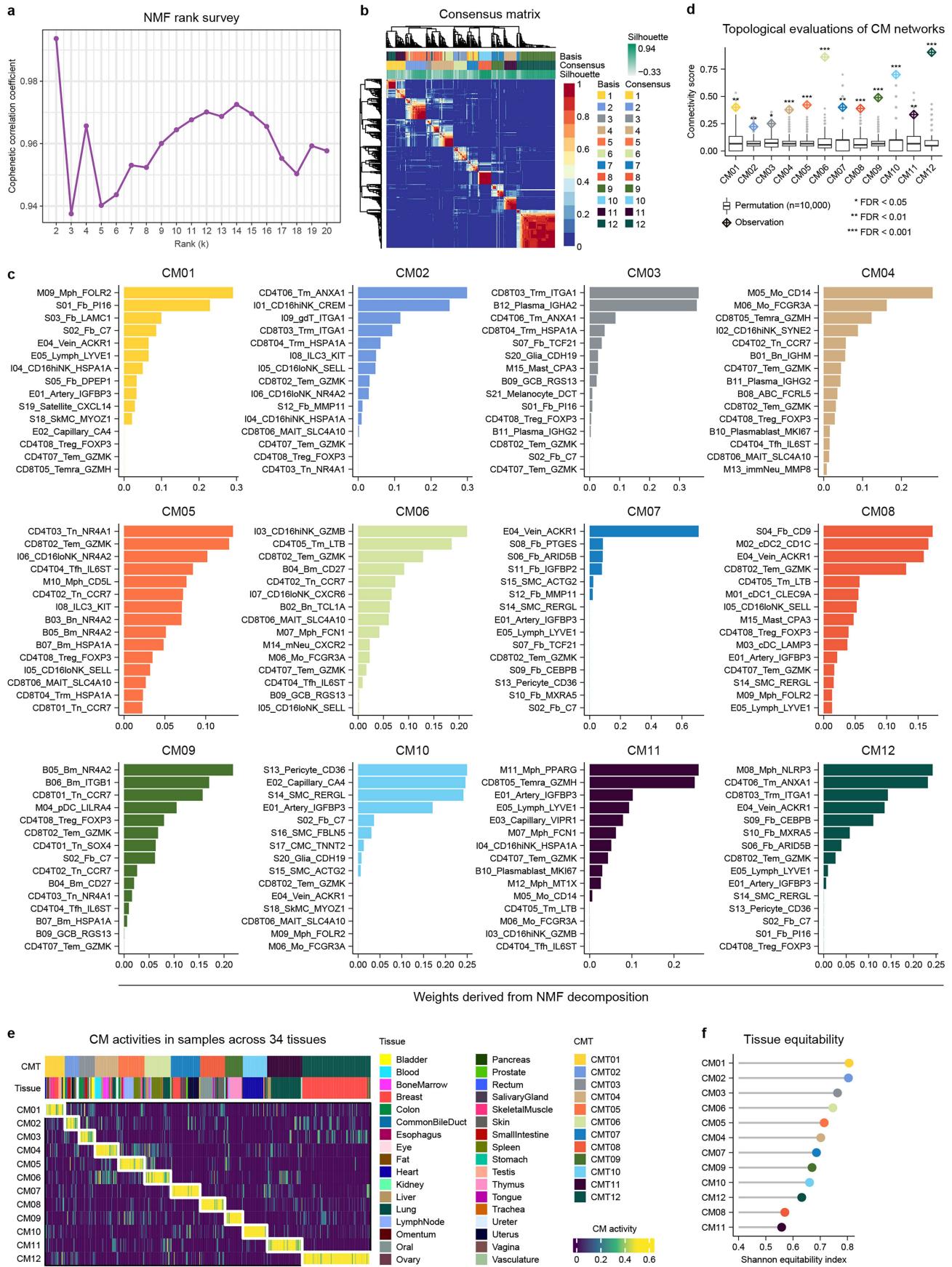


b



**Extended Data Fig. 2 | Benchmarking data integration methods.** (a) Summary of benchmarking for several widely used Python-based integration methods on the total atlas and a subset atlas according to the scIB platform. Overall scores are calculated using a weighted mean of batch correction (40%) and

bio-conservation (60%) scores. Methods are ranked by overall scores. (b) UMAP visualization of total cells colored by compartments (top) or tissues (bottom) across different integration methods.

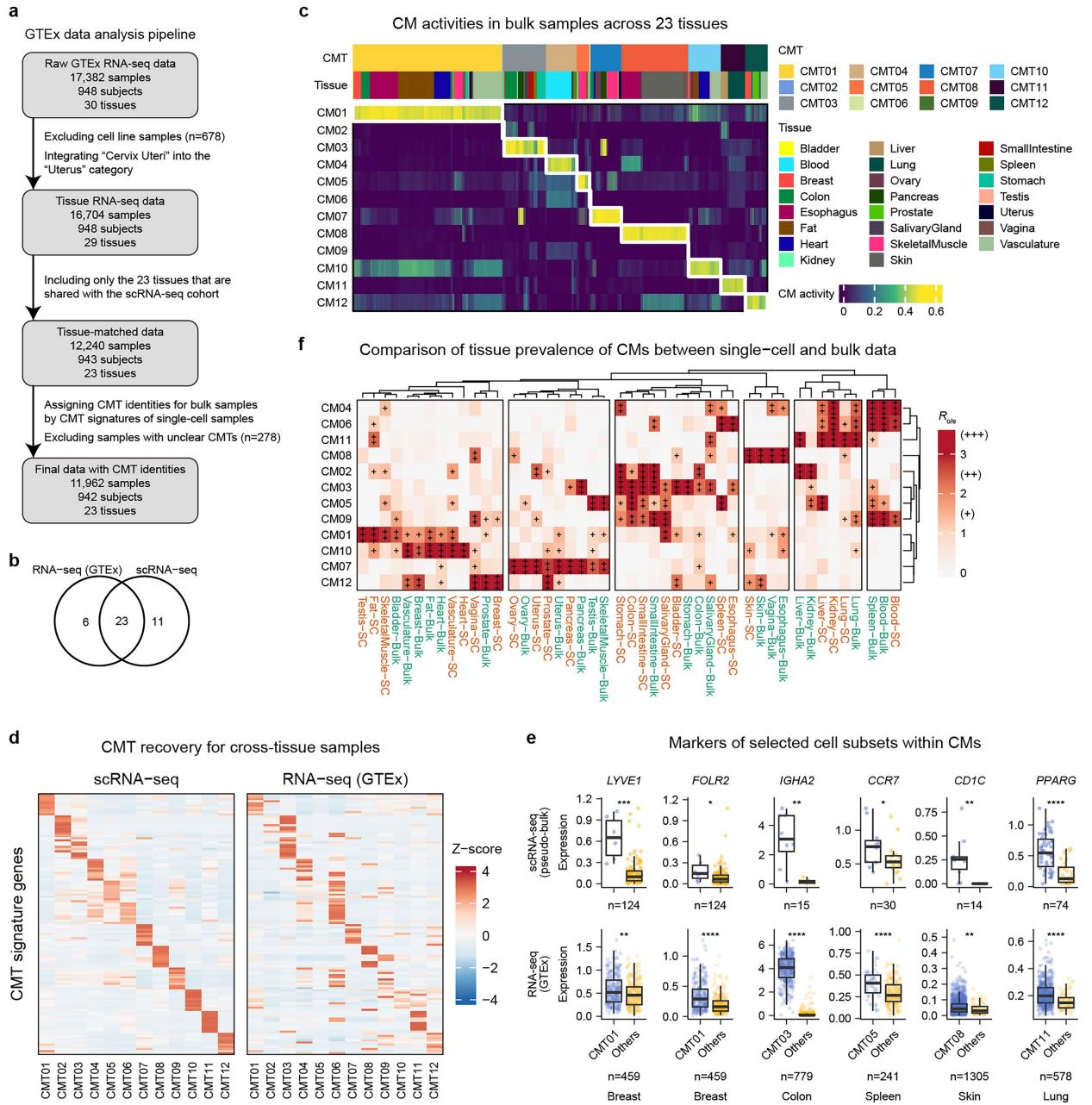


Extended Data Fig. 3 | See next page for caption.

## Article

**Extended Data Fig. 3 | NMF and cell subset ranking.** (a) NMF rank selection based on cophenetic correlation coefficients and their consistency, with the final rank chosen as 12 (Methods). (b) Consensus plot for NMF with a rank of 12. (c) Bar plots showing the top 15 cell subsets ranked by NMF weights for each factor (CM). (d) Topological and statistical evaluations of CMs, with the connectivity score defined as the ratio of observed edges to the total possible

edges among all nodes within the CM network. (e) CM activities in scRNA-seq samples across 34 tissues. Each sample is assigned a CMT based on its dominant CM, with samples ordered by CMT identities and tissues. (f) CM Shannon equitability across tissues. For box plot in d, the center line represents the median, the box limits delineate the upper and lower quartiles, and the whiskers extend to the highest and lowest values within 1.5× the interquartile range.

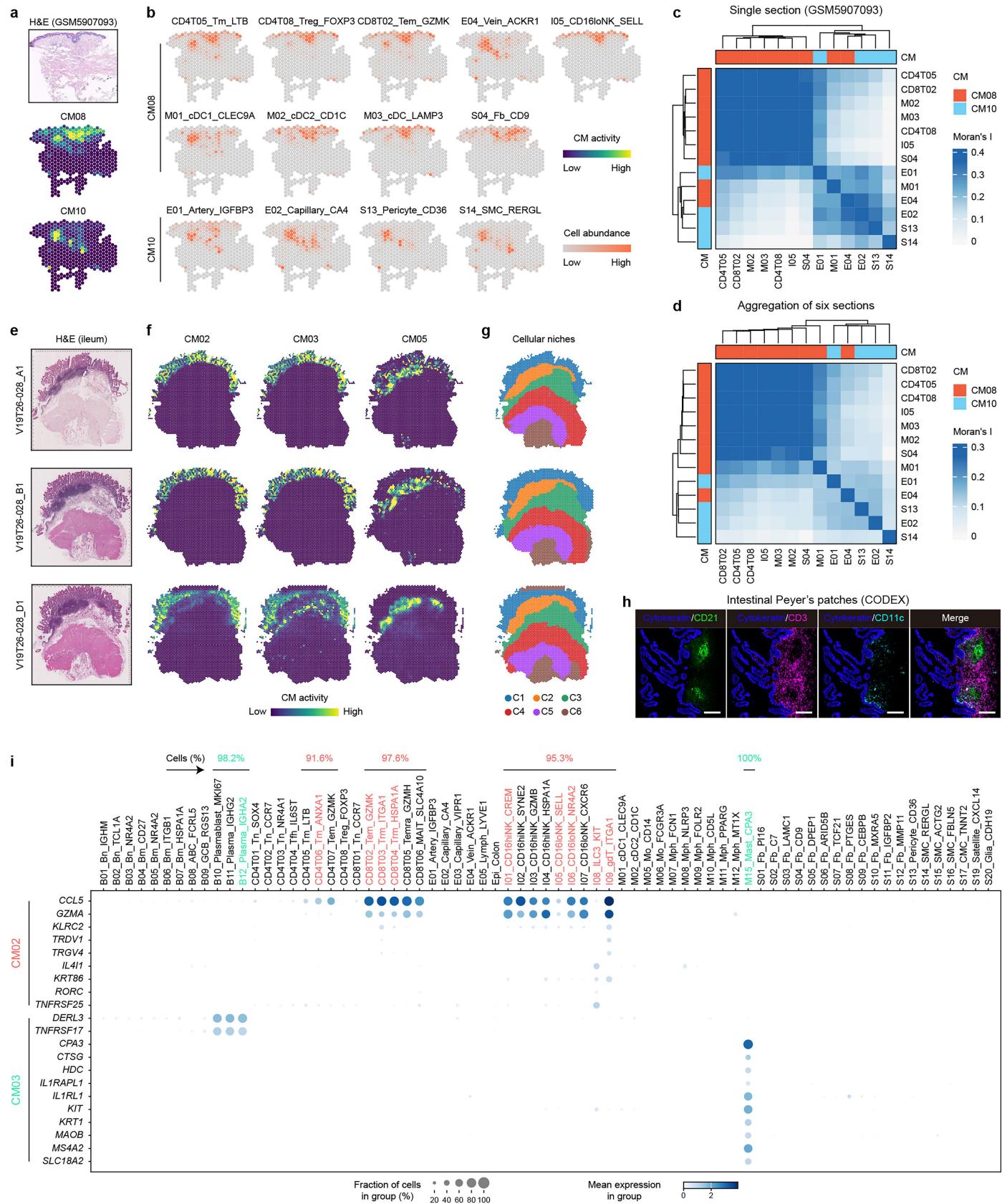


#### Extended Data Fig. 4 | Validation of CMs using GTEx RNA-seq data.

(a) Overview of the GTEx data analysis pipeline. (b) Venn diagram showing intersecting tissues in GTEx RNA-seq and curated scRNA-seq cohorts. (c) CM activities in RNA-seq samples across 23 tissues. Each sample is assigned a CMT based on its dominant CM, with samples ordered by CMT identities and tissues. (d) Heat maps showing the average expression of CMT signature genes (rows) across scRNA-seq (left) and GTEx RNA-seq (right) samples grouped by CMT identities (columns). (e) Gene expression of selected cell subset markers across

samples grouped by CMT identities. Data from scRNA-seq (top) and GTEx RNA-seq (bottom) are shown separately. Sample sizes are noted. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , one-tailed unpaired Wilcoxon tests. (f) Tissue prevalence of CMs measured by  $R_{o/e}$  for tissues profiled using both single-cell (SC) and bulk RNA-seq. For box plots in e, the center line represents the median, the box limits delineate the upper and lower quartiles, and the whiskers extend to the highest and lowest values within  $1.5 \times$  the interquartile range.

# Article

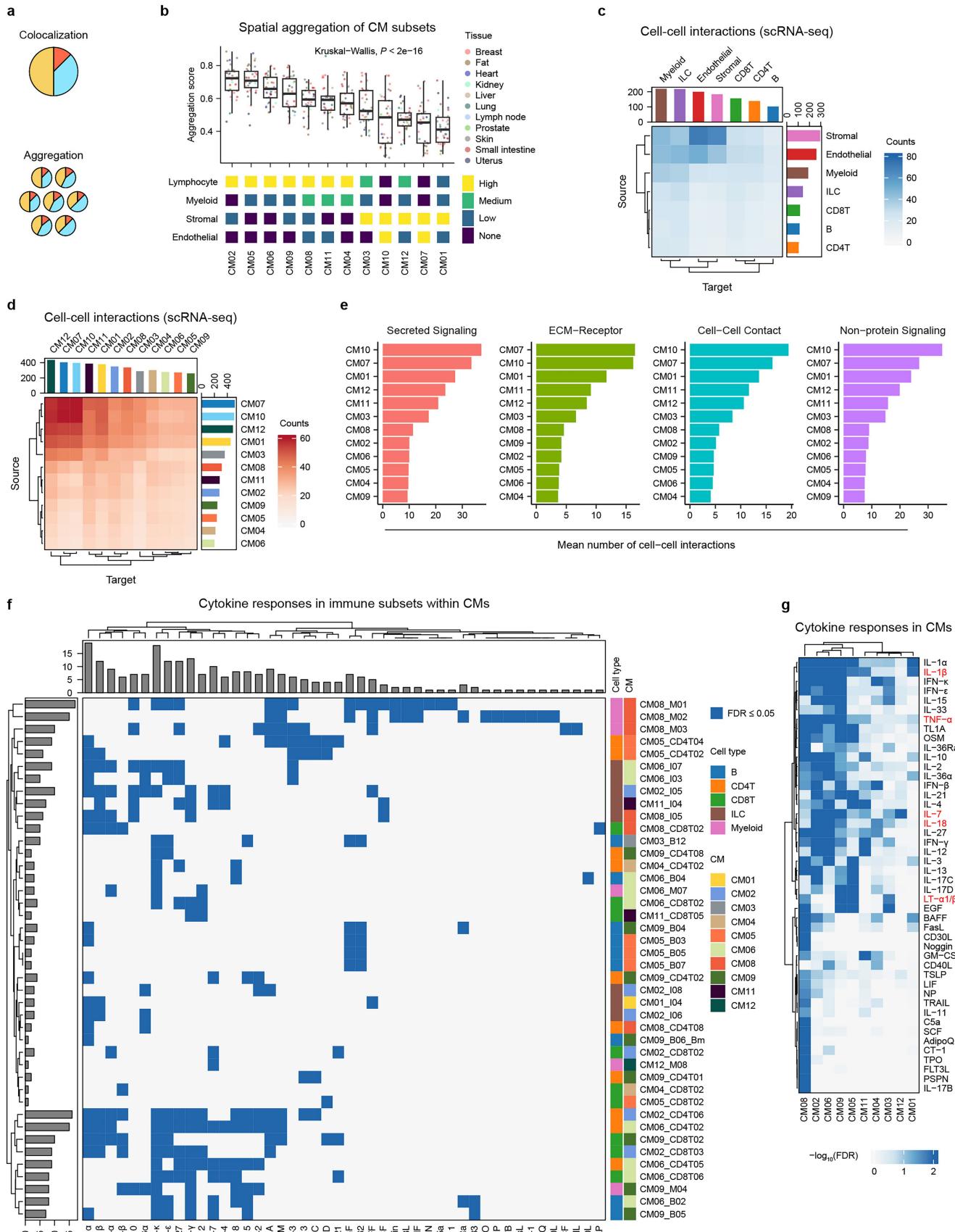


Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | Spatial characterization of CMs.** (a, b) Spatial distribution of CM08 and CM10 (a) and their component cell subsets (b) in a Visium section of the skin. (c) Spatial aggregation (based on Moran's I) among cell subsets within CM08 or CM10 in the Visium section shown in a, b. (d) Averaged spatial aggregation (based on Moran's I) among cell subsets within CM08 or CM10 across six Visium sections from six available donors. (e–g) H&E staining (e), spatial distribution of CMs (CM02, CM03, and CM05) (f), and cellular niches identified by CellCharter (g) in three adjacent Visium sections of the ileum from one donor.

(h) CODEX imaging of intestinal Peyer's patches showing B cells (CD21, green), T cells (CD3, pink), and DCs (CD11c, cyan). Scale bars, 200  $\mu$ m. Data from the HuBMAP portal (dataset ID: HBM785.FJVT.469). (i) Dot plot illustrating scRNA-seq expression of a gene panel differentiating CM02 and CM03 among various cell subsets in the colon. The percentages of colored subsets among similar subsets are indicated at the top. For example, 98.2% of the plasma cells (B10, B11, and B12) are B12.

# Article

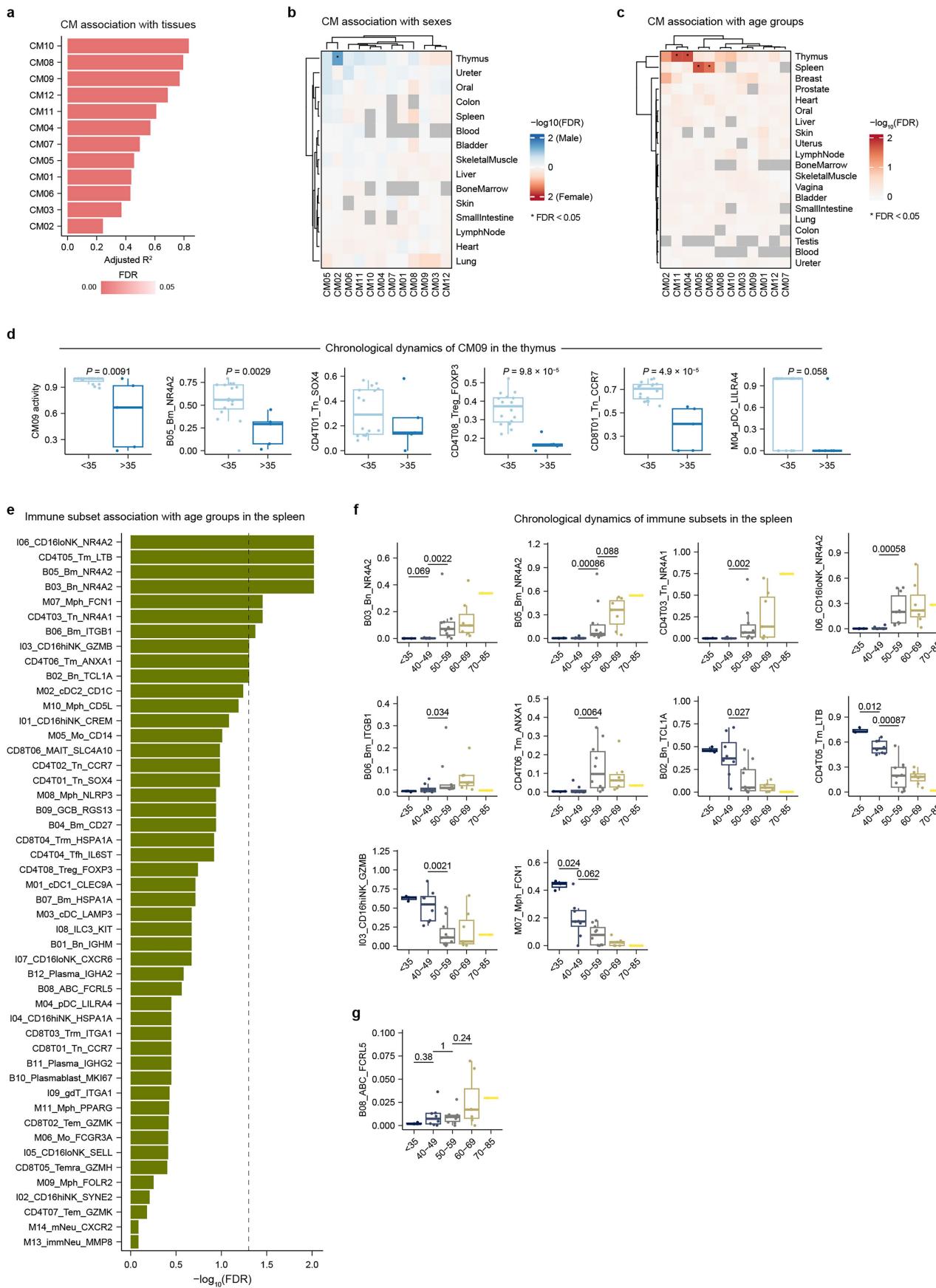


**Extended Data Fig. 6** | See next page for caption.

**Extended Data Fig. 6 | Spatial organization and cell-cell interactions of CMs.** (a) Schematic of two spatial metrics for CMs using 10x Visium data. The colocalization score evaluates the extent to which two cell subsets colocalize within the same spots, while the aggregation score measures the spatial adjacency of two cell subsets, encompassing both colocalization and broader spatial relationships. (b) Spatial aggregation (top) and cellular composition (bottom) of CMs. Top: CM aggregation scores across Visium samples colored by tissues (Methods). CMs are ordered by the median of aggregation scores across samples. Bottom: Proportion of different cell lineages in all cell-subset nodes within individual CM networks shown in Fig. 2b. None = 0 < Low < 0.2 ≤ Medium < 0.4 ≤ High. (c,d) CellPhoneDB cell-cell interaction analysis based on

scRNA-seq data. Heat maps showing of cell-cell interaction counts within and across cell types (c) or CMs (d). Bar plots on the top and right represent the sums of the columns and rows, respectively. (e) CellChat cell-cell interaction counts based on scRNA-seq, stratified by ligand-receptor pair types. (f) Heat map showing the responsive cytokines (columns) among cell subsets within individual CMs (rows) (Methods). Bar plots on the top and left represent the sums of columns and rows, respectively. (g) Cytokine responsiveness in CMs. For each CM, the CM FDR denotes the minimum FDR value among all subsets within the CM (Methods). For box plot in b, the center line represents the median, the box limits delineate the upper and lower quartiles, and the whiskers extend to the highest and lowest values within  $1.5 \times$  the interquartile range.

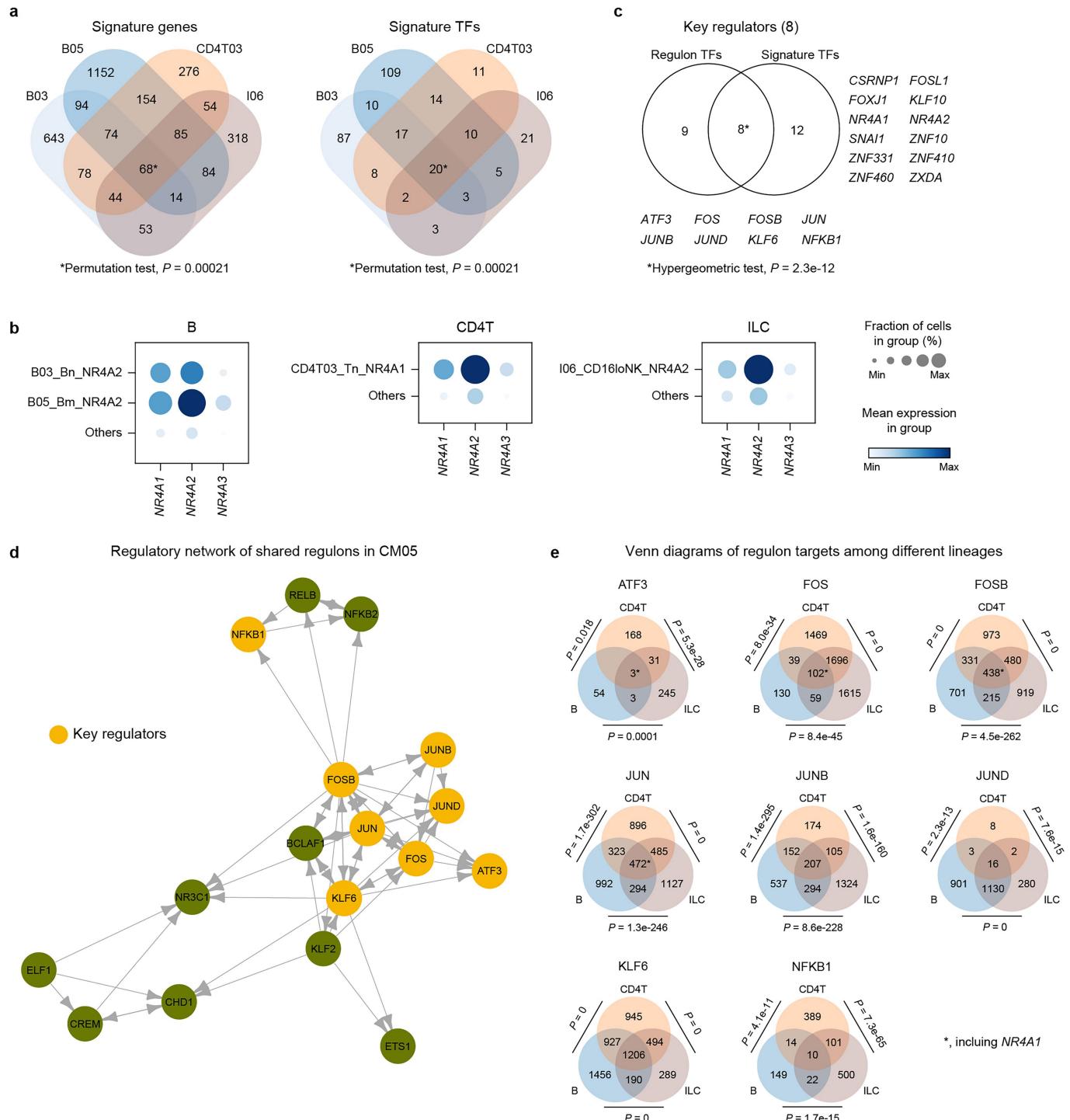
# Article



Extended Data Fig. 7 | See next page for caption.

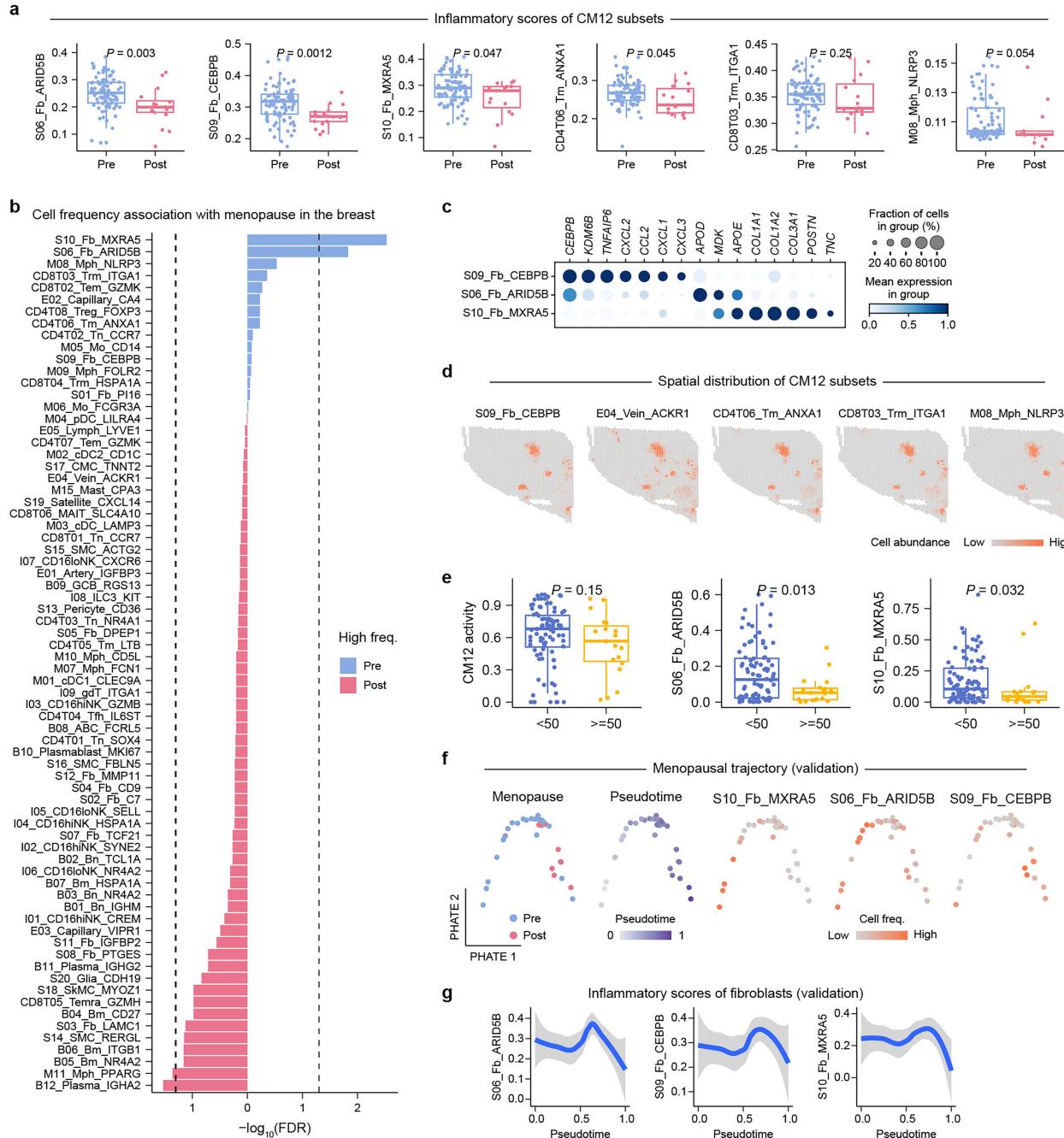
**Extended Data Fig. 7 | Associations between CM activity and tissue, sex, and age group.** (a) Bar plot showing the proportion of variance explained of CM activities by tissues, measured as adjusted R<sup>2</sup> from linear models. Significance of linear models is indicated. (b) Heat map showing associations of CM activity with sexes in non-reproductive tissues. Two-tailed unpaired Wilcoxon tests. (c) Heat map showing associations of CM activity with age groups. Kruskal–Wallis tests. (d) CM09 activity and frequencies of its subsets in individual thymus samples stratified by age groups. Two-tailed unpaired Wilcoxon tests. (e) Bar plot showing associations between age groups and frequencies of immune cell subsets in the spleen. Kruskal–Wallis tests. (f) Frequencies of selected immune subsets in individual spleen samples stratified by age groups. Two-tailed unpaired Wilcoxon tests. (g) Frequencies of the B08 subset in individual spleen samples stratified by age groups. Two-tailed unpaired Wilcoxon tests. For box plots in d,f,g, the center line represents the median, the box limits delineate the upper and lower quartiles, and the whiskers extend to the highest and lowest values within 1.5× the interquartile range.

# Article



**Extended Data Fig. 8 | CM05 analysis in the spleen.** (a) Venn diagrams showing the overlap of four CM05 subsets in all signature genes (left) and signature TFs (right). (b) Expression of *NR4A1*, *NR4A2*, and *NR4A3* in CM05 subsets and other cells within the same cell types. (c) Venn diagram showing the overlap of regulon TFs identified by SCENIC analysis and signature TFs.

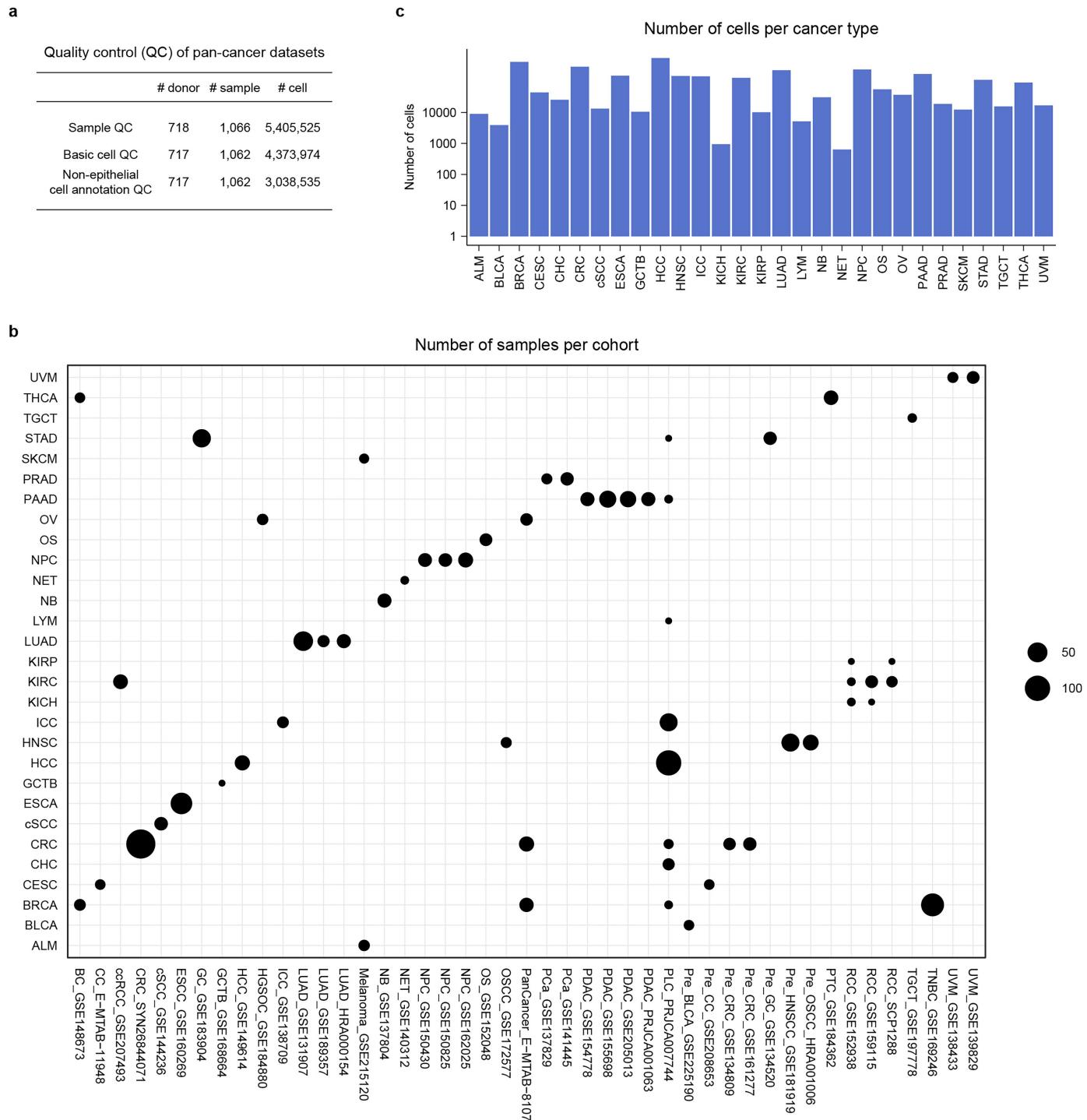
The eight shared TFs (key regulators) and other signature TFs are indicated. (d) Regulatory network of 17 convergent regulons in CM05. The key regulators are highlighted. Regulatory relationships are determined by SCENIC analysis. (e) Venn diagrams showing the overlap in regulon targets among different cell types. Hypergeometric tests. TF, transcription factor.



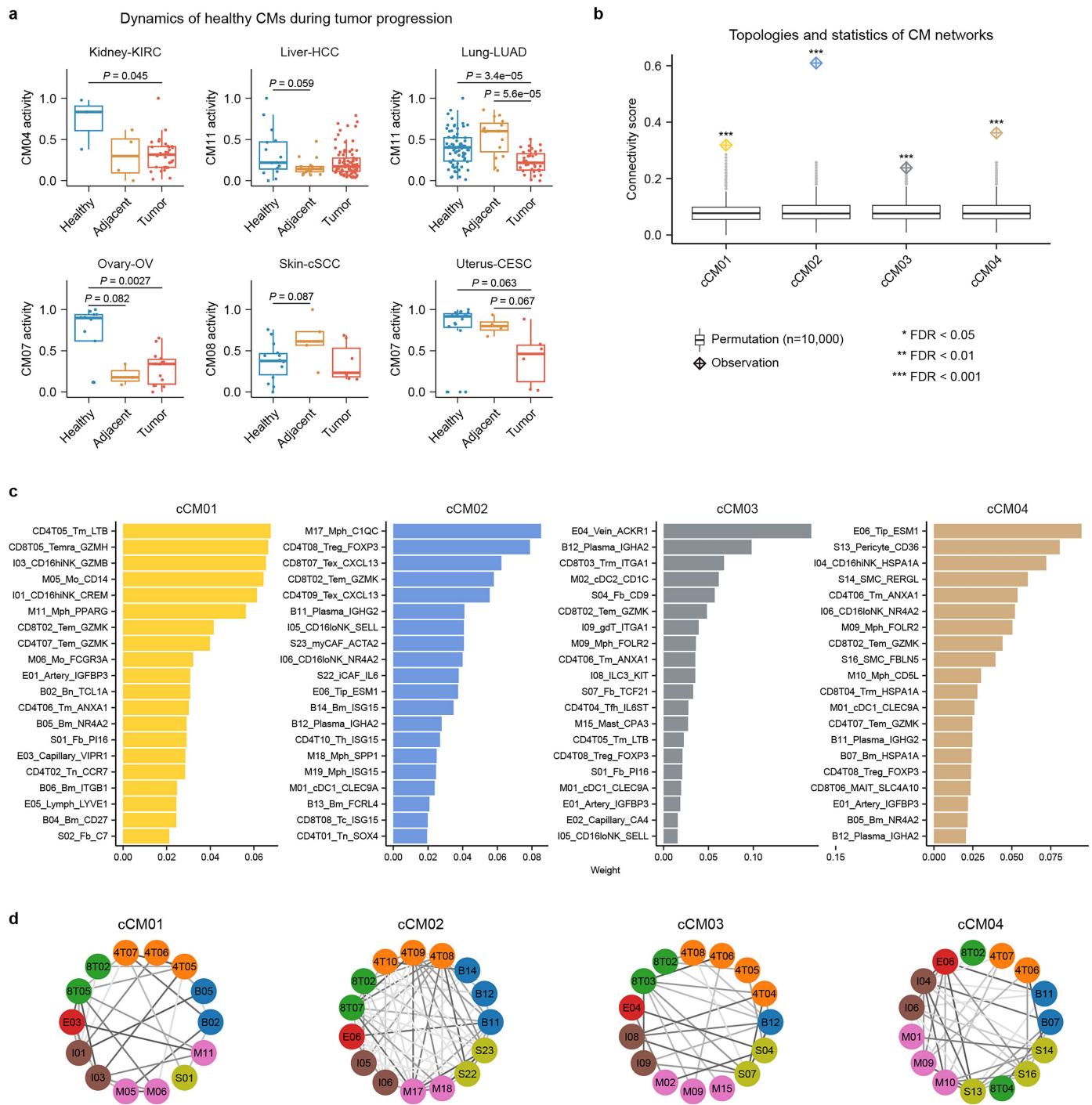
**Extended Data Fig. 9 | CM12 analysis in the breast.** (a) Inflammatory scores of CM12 subsets in pre- and post-menopausal samples. Two-tailed unpaired Wilcoxon tests. (b) Bar plot showing the association between menopausal status and frequencies of all nonepithelial subsets in the breast. The dotted line indicates FDR at 0.05. Two-tailed unpaired Wilcoxon tests. (c) Dot plot showing the scaled expression of selected signature genes of CM12-related specialized fibroblast subsets in the breast. (d) Spatial distribution of CM12 cell subsets in a spatial transcriptomics slide of the breast. (e) CM12 activity and frequencies of S10 and S06 fibroblasts in samples under 50 years old and over 50 years old.

Two-tailed unpaired Wilcoxon tests. (f) Validation of menopausal trajectory using another breast dataset (Methods). The PHATE embedding showing breast samples (individual dots) colored by menopausal status, pseudotime and frequencies of fibroblasts. (g) Trend lines show LOESS-smoothed fibroblast inflammatory scores along pseudotime in breast samples. The error bands show the 95% confidence intervals. For box plots in a,e, the center line represents the median, the box limits delineate the upper and lower quartiles, and the whiskers extend to the highest and lowest values within 1.5× the interquartile ranges.

# Article



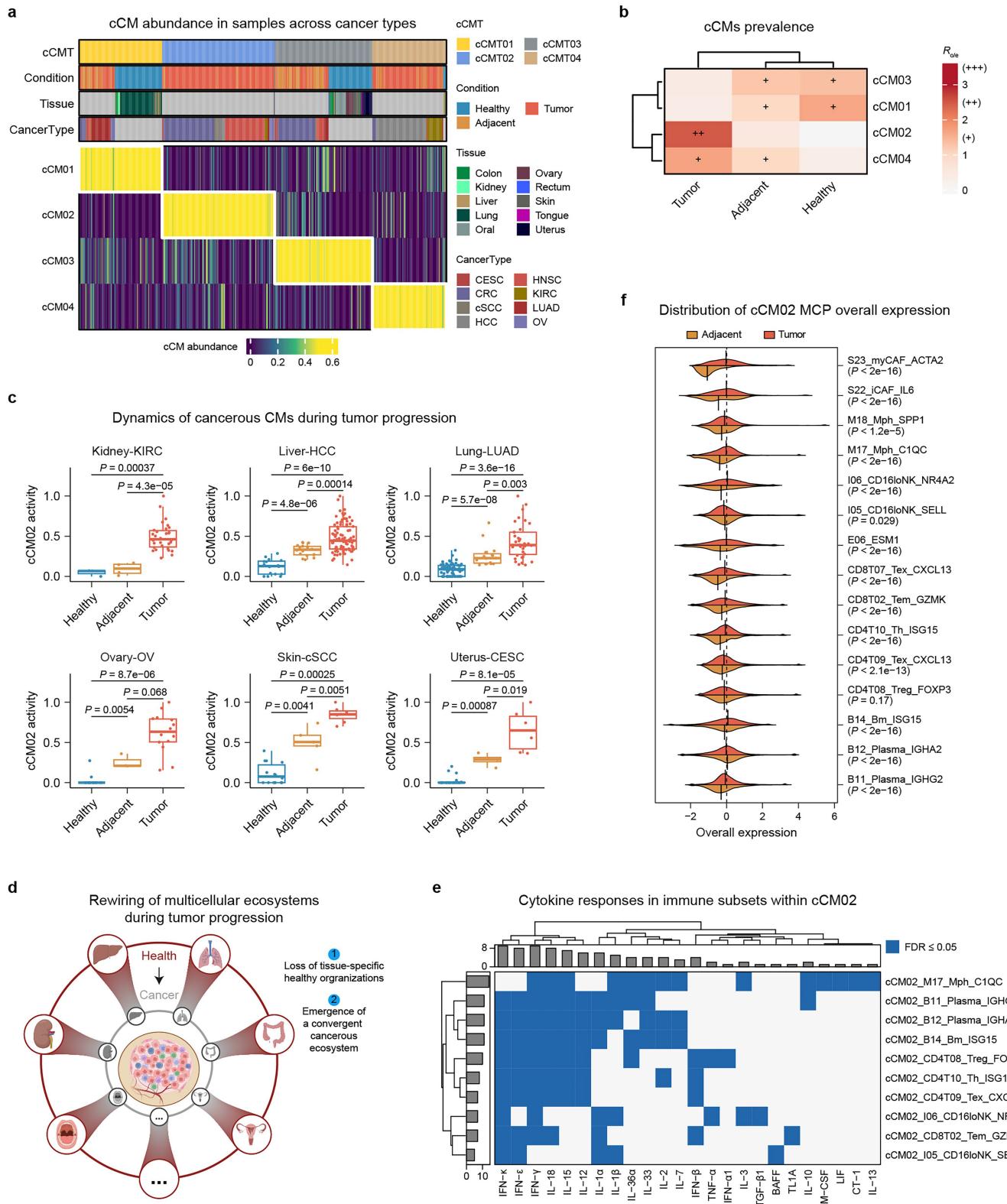
**Extended Data Fig. 10 | Overview of pan-cancer single-cell datasets.** (a) Statistics of donor, sample and cell numbers before and after QC for scRNA-seq datasets. (b) Dot plot showing the number of samples per cohort. (c) Bar plot showing the number of cells per cancer type.



**Extended Data Fig. 11 | Analysis of CMs and cCMs.** (a) Healthy CM activities across healthy, adjacent non-tumor, and tumor samples. Dots represent individual samples. Two-tailed unpaired Wilcoxon tests. (b) Topological and statistical evaluations of cCMs, with the connectivity score defined as the ratio of observed edges to the total possible edges among all nodes within the cCM network. (c) Bar plots showing the top 20 cell subsets ranked by NMF weights

for each factor (cCM). (d) Network plots of the four CMs identified from the pan-cancer analysis. Nodes represent cell subsets labeled by short names and colored by cell types. Edge color indicates correlation specificity. For box plots in a,b, the center line represents the median, the box limits delineate the upper and lower quartiles, and the whiskers extend to the highest and lowest values within  $1.5 \times$  the interquartile range.

# Article



**Extended Data Fig. 12 | Pan-cancer analysis of cCMs.** (a) cCM activities across cancer types. Each sample is assigned a cCMT by its dominant cCM, and all samples are ordered by their cCMT identities and cancer types. (b) Tissue prevalence of cCMs measured by  $R_{\text{pre}}$ . (c) cCM02 activities in healthy, adjacent non-tumor, and tumor samples across cancer types. Dots represent individual samples. Two-tailed unpaired Wilcoxon tests. (d) Schematic representation of rewiring of multicellular ecosystems during tumor progression. Created in BioRender; Shi, Q. (2025), <https://BioRender.com/atp19sm>. (e) Heat map

showing the responsive cytokines among cell-subset components of cCM02. Bar plots on the top and left represent sums of columns and rows, respectively. (f) Distribution of cCM02 MCP overall expression across different cell subsets for cells from tumor and adjacent non-tumor samples. Vertical black line indicates the median of the distribution. Two-tailed unpaired Wilcoxon tests. For box plots in c, the center line represents the median, the box limits delineate the upper and lower quartiles, and the whiskers extend to the highest and lowest values within  $1.5 \times$  the interquartile range. cCMT, cCM type.

## Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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	scRNA-seq and spatial transcriptomics data were downloaded from external studies and curated using custom code written in Python (v3.11.6) or R (v4.3.0). TCGA datasets were downloaded by TCGAbiolinks (v2.30.4). Custom code is available on Zenodo ( <a href="https://zenodo.org/records/14175262">https://zenodo.org/records/14175262</a> )
Data analysis	General: Python (v3.11.6), R (v4.3.0) with publicly available packages scRNA-seq analysis: Scanpy toolkit (v1.9.3) (which includes Scrublet ( <a href="https://github.com/swolock/scrublet">https://github.com/swolock/scrublet</a> ), BBKNN (v1.5.1), Harmony (v0.1.5), Scanorama (v1.7.3)), scIB (v1.0.3), scvi-tools/scVI/scANVI (v0.20.1), scalex (v1.0.0), CellTypist (v1.6.0), TOSICA ( <a href="https://github.com/JackieHanLab/TOSICA">https://github.com/JackieHanLab/TOSICA</a> ) Cell module analysis: CoVarNet developed by this study ( <a href="https://github.com/QiangShiPKU/CoVarNet">https://github.com/QiangShiPKU/CoVarNet</a> ), nmf (v0.26), Seurat (version 5.0.0), pySCENIC (0.12.1) Spatial transcriptomics analysis: cell2location (version 0.1.3), CellCharter (v0.2.3), spdep (v1.3.6), Xenium Explorer 3 (version 3.1.0) Cell-cell interaction analysis: CellPhoneDB (version 4.0.0), CellChat (version 2.1.0) Multicellular program analysis: DIALOGUE ( <a href="https://github.com/livnatje/DIALOGUE">https://github.com/livnatje/DIALOGUE</a> ), AUCell (v1.24.0) Cytokine response analysis: clusterProfiler (version 4.10.1) Trajectory analysis: PHATE (v0.3.0), Palantir (v1.3.3) Visualization: ComplexHeatmap (v2.14.0), igraph (v2.0.3), dendextend (v1.17.1)

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

This study relied on the curation and integrative analysis of external studies and did not generate new primary data. Curated pan-tissue and pan-cancer scRNA-seq datasets can be downloaded on Zenodo (<https://doi.org/10.5281/zenodo.15169362>) and can be interactively explored at <http://cm.cancer-pku.cn>. Spatially resolved transcriptomics datasets used are listed in Supplementary Table 3. Other public data used in this study include GTEx (<https://www.gtexportal.org/home/>) and TCGA (<https://portal.gdc.cancer.gov/>). Source data are provided with this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender The study was conducted using curated public data, and sex was not taken into account during data collection.

Reporting on race, ethnicity, or other socially relevant groupings No socially relevant information was used.

Population characteristics All samples were from healthy human tissues or treatment-naive cancer patients

Recruitment This study relied on the curation of public data without additional recruitment.

Ethics oversight This study relied on the curation of public data.

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

## 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](https://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	No formal sample size calculation was performed for this study. We aimed to include as many samples as possible to ensure comprehensive coverage and robustness of our analysis. We curated a total of 706 normal samples across 33 datasets and 1062 cancer samples across 48 datasets. After applying stringent quality control measures, we utilized all available samples in our analysis. We believe these sample sizes are sufficient to provide a comprehensive view of cross-tissue cellular coordination and its rewiring in cancer, given the extensive datasets we compiled from various sources.
Data exclusions	According to the concept of cellular modules, we excluded samples where certain cell types were absent due to experimental cell type enrichment operations.
Replication	All our findings have been well validated to ensure high reproducibility. The scRNA-seq-derived cellular modules were cross-verified using independent datasets from the GTEx project and spatial transcriptomics data. This multi-faceted approach enhances the credibility of our results. Furthermore, to facilitate the reproduction of our findings, we have made the datasets used in our analysis publicly available along with the detailed analysis code. This transparency allows other researchers to replicate our study and build upon our findings, further confirming their robustness.
Randomization	Our study utilized publicly available datasets, compiling data from various sources to ensure comprehensive coverage. The samples were not allocated into experimental groups by us, as we relied on the existing data; thus, random allocation was not relevant.
Blinding	Blinding was not relevant in our study as we conducted an analysis of publicly available datasets that had already been collected and processed by other researchers. Since we did not have direct involvement in the data collection process, nor did we perform any experimental interventions, blinding to group allocation was not relevant. Our investigation was focused on computational analysis and integration of existing data to identify and characterize cross-tissue cellular modules and their changes in cancer.

# 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
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	Plants

## Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

## Plants

### Seed stocks

*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

### Novel plant genotypes

*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

### Authentication

*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.*