ANASTASIIA ONYSHCHENKO 10/20/2020 JOURNAL CLUB

ARTICLES
https://doi.org/10.1038/s41556-019-0439-6

nature cell biology

Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization

Chiara Baccin[1,2,10], Jude Al-Sabah[3,4,10], Lars Velten[1,9,10,11*], Patrick M. Helbling[5], Florian Grünschläger[3,4], Pablo Hernández-Malmierca[3,4], César Nombela-Arrieta[5], Lars M. Steinmetz[1,6,7,11*], Andreas Trumpp[3,4,8,11*] and Simon Haas[3,4,11*]

The bone marrow constitutes the primary site for life-long blood production and skeletal regeneration. However, its cellular and spatial organization remains controversial. Here, we combine single-cell and spatially resolved transcriptomics to systematically map the molecular, cellular and spatial composition of distinct bone marrow niches. This allowed us to transcriptionally profile all major bone-marrow-resident cell types, determine their localization and clarify sources of pro-haematopoietic factors. Our data demonstrate that Cxcl12-abundant-reticular (CAR) cell subsets (Adipo-CAR and Osteo-CAR) differentially localize to sinusoidal and arteriolar surfaces, act locally as 'professional cytokine-secreting cells' and thereby establish perivascular micro-niches. Importantly, the three-dimensional bone-marrow organization can be accurately inferred from single-cell transcriptome data using the RNA-Magnet algorithm described here. Together, our study reveals the cellular and spatial organization of bone marrow niches and offers a systematic approach to dissect the complex organization of whole organs.

Bone marrow (BM) niches are specialized microenvironments where distinct mesenchymal cells, the vasculature and differentiated haematopoietic cells interact to regulate the maintenance and differentiation of haematopoietic and mesenchymal stem cells (HSCs and MSCs, respectively)[1–3]. Classically, BM niches have been studied using genetic approaches that involve labelling cell types and ablating candidate niche factors based on the expression of individual reporter genes. These studies have generated important insights into the functional roles and cellular sources of key cytokines, such as Cxcl12 or Scf (*Kitl*)[4–10]. However, single markers are used to define cell types in such approaches, probably resulting in the labelling of heterogeneous populations. These limitations have resulted in a controversial debate about the importance of distinct cell types and factors, and their localization to sinusoidal, arteriolar or endosteal niches[4–10]. To gain a global understanding of the cell types and niches in the BM, we have generated a molecular map of all major BM populations based on single-cell RNA sequencing (scRNAseq). We then used spatially resolved transcriptomics in combination with computational tools to allocate cell types to different BM niches, determine the molecular mediators of intercellular interactions, and identify the cellular and spatial sources of niche factors.

Results

Identification and characterization of BM-resident cell types by scRNAseq. The frequencies of the various BM cell types differ by several orders of magnitude[11], imposing a challenge to scRNAseq approaches. To capture both highly abundant and extremely rare BM-resident cells, we performed droplet-based scRNAseq[12] of cells from total mouse BM, followed by progressive depletion of the abundant cell types or enrichment of rare populations from undigested BM or enzymatically digested bones (Fig. 1a). In total, our dataset comprised 7,497 cells, which formed 32 clusters corresponding to distinct cell types or stages of differentiation (Fig. 1b and Extended Data Fig. 1). Importantly, this map is not quantitative with regard to the relative size of the different cell populations, as the dissociation rates differ greatly between cell types[11]. As detailed below, the cell-type annotation was performed based on the expression of marker genes (Supplementary Table 1 and Extended Data Figs. 2,3), gene-ontology analyses (Supplementary Table 1) and quantification of the enrichment of cluster gene signatures in the transcriptomic data of previously described bulk populations using the CIBERSORT algorithm[13,14] (Extended Data Fig. 4a,b and Supplementary Note 1). We used SOUP[15] to confirm that the mesenchymal populations described here are primarily distinct cell types, while transition states between clusters exist (Extended Data Fig. 4c–e).

As expected, scRNAseq of the total BM identified the major haematopoietic cell types, including dendritic cells, neutrophils, monocytes, T cells and distinct developmental stages of B cells (Fig. 1b and Extended Data Figs. 2, 4a). Following the depletion of these major immune populations, scRNAseq primarily yielded erythroid progenitors that exhibited low expression of the pan-haematopoietic marker CD45 and displayed erythroid markers such as CD71 (ref. [16]; Extended Data Fig. 2). An additional 2% of these

---

[1]Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. [2]Faculty of Biosciences, Joint PhD degree from EMBL and Heidelberg University, Heidelberg, Germany. [3]Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH), Heidelberg, Germany. [4]Division of Stem Cells and Cancer, Deutsches Krebsforschungszentrum (DKFZ) and DKFZ–ZMBH Alliance, Heidelberg, Germany. [5]Department of Medical Oncology and Hematology, University Hospital and University of Zürich, Zürich, Switzerland. [6]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [7]Stanford Genome Technology Center, Palo Alto, CA, USA. [8]German Cancer Consortium (DKTK), Heidelberg, Germany. [9]Present address: Center for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain. [10]These authors contributed equally: Chiara Baccin, Jude Al-Sabah, Lars Velten. [11]These authors jointly supervised this work: Lars Velten, Lars Steinmetz, Andreas Trumpp, Simon Haas. *e-mail: lars.velten@embl.de; larsms@embl.de; a.trumpp@dkfz-heidelberg.de; s.haas@dkfz-heidelberg.de

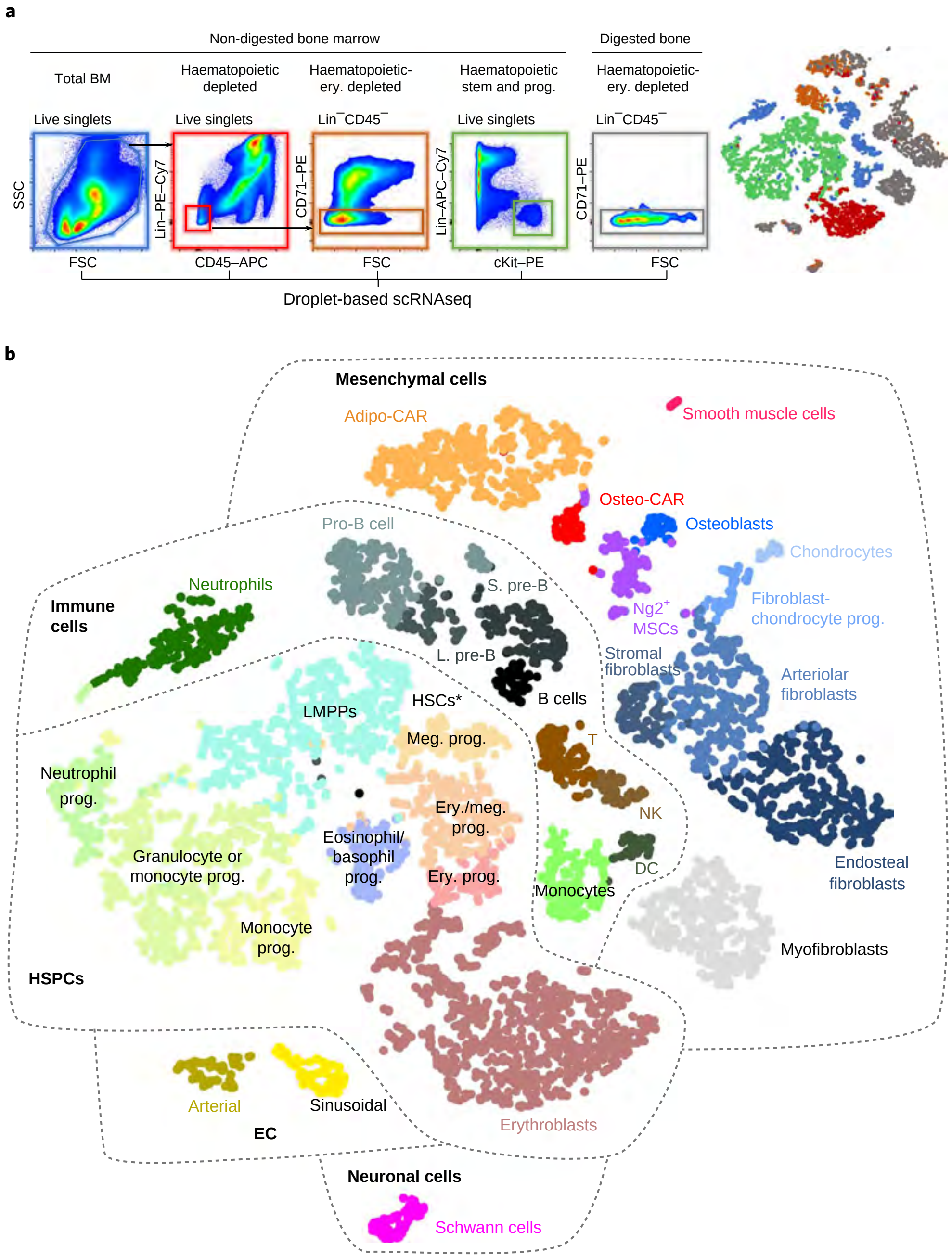


Fig. 1 | Identification of BM-resident cell types by scRNAseq. a, Overview of the FACS strategy (left). Five consecutive scRNAseq runs were performed in total. T-distributed stochastic neighbour embedding (t-SNE) of all cells, colour-coded according to the respective experiment on the left (right). **b**, t-SNE projection of all cells with the clusters colour-coded. *Lin⁻Kit⁺Sca1^{high} HSCs reside at the interface of lymphoid-primed multipotent progenitors (LMPPs) and megakaryocyte progenitors, see also Extended Data Fig. 3g. T, T cells; NK, natural killer cells; S. pre-B, small pre-B cells; L. pre-B, large pre-B cells; DC, dendritic cells; ery., erythroid; meg., megakaryocyte; prog., progenitor; $n = 7,497$ cells.

cells were non-haematopoietic (Fig. 1a and Extended Data Fig. 1). To capture the non-haematopoietic cells in depth, we subsequently depleted cells expressing the lineage markers, CD45 or CD71 from non-digested BM and enzymatically digested bone chips (Fig. 1a). We thereby identified rare populations, including Schwann cells (*Mog*, *Mag*), smooth muscle cells (*Tagln*, *Acta2*), putative myofibroblasts, nine different *Pdgfra*-positive mesenchymal populations and two endothelial cell (EC) clusters (*Cdh5*, *Pecam*; Extended Data Fig. 3 and Supplementary Table 1). The endothelial populations comprised Sca1 (*Ly6a*)-expressing arterial ECs and *Emcn*-expressing sinusoidal ECs (Fig. 2a,b)[17]. The mesenchymal lineage contained chondrocytes (*Acan*, *Sox9*), osteoblasts (*Osteocalcin*/*Bglap*, *Col1a1*) and several less well-described cell types. These included three distinct fibroblast-like populations, which we annotated based on their localization (see below), and two additional populations that showed a high transcriptomic similarity to previously described SCF–green fluorescent protein (GFP)⁺ (ref. [5]) and Cxcl12–GFP⁺ (ref. [9]) Cxcl12-abundant reticular (CAR) cells (Fig. 2c)[4,18]. Remarkably, the latter two populations expressed adipocyte and osteo-lineage genes differentially (Fig. 2d), despite similar overall transcriptomic profiles. We therefore termed these subpopulations 'Adipo-CAR' and 'Osteo-CAR' cells, respectively. The Adipo-CAR cells expressed high levels

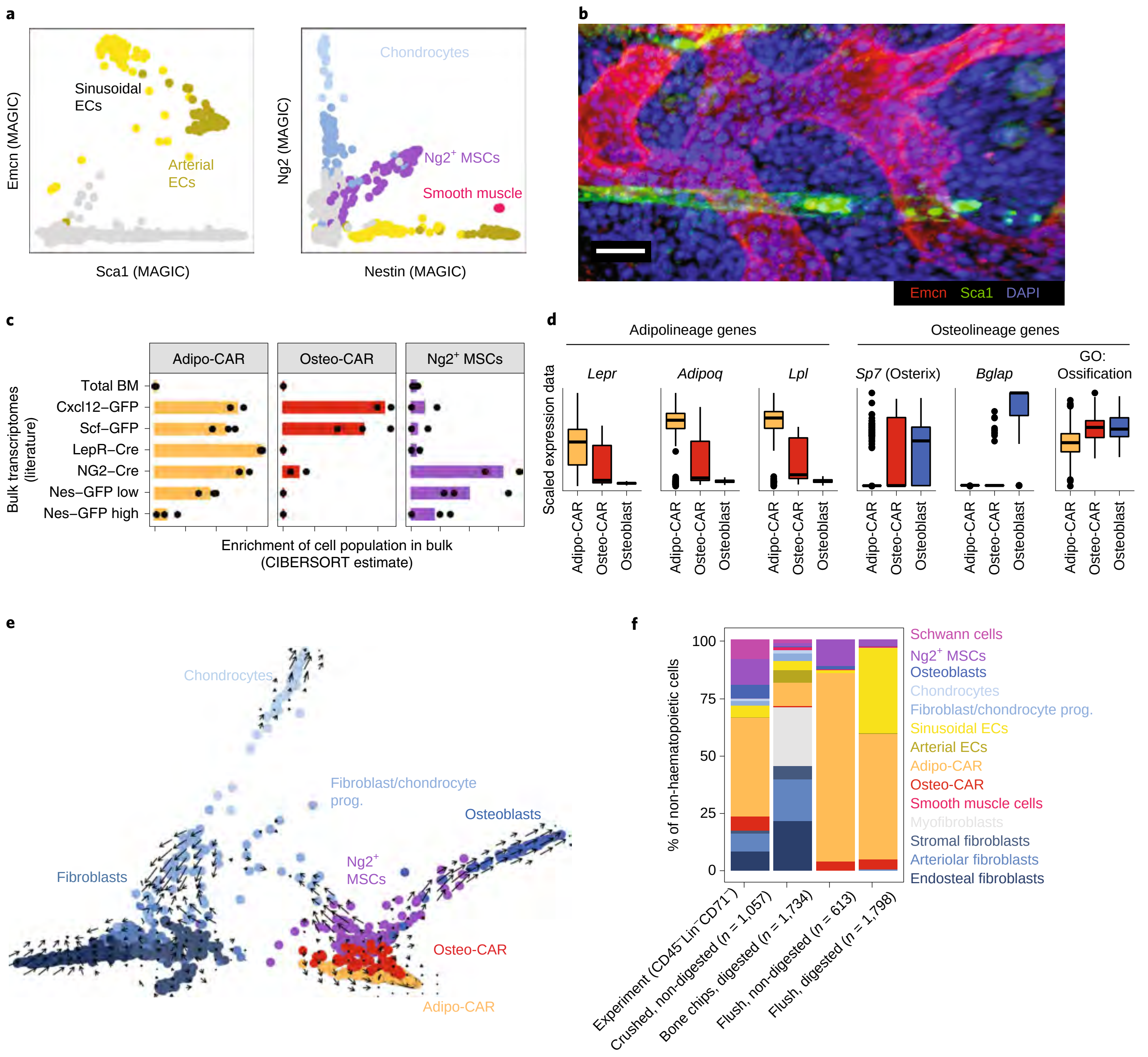


Fig. 2 | Characterization of BM-resident cell types. **a**, Gene expression levels of Sca1 (*Ly6a*), endomucin (*Emcn*), nestin (*Nes*) and Ng2 (*Cspg4*), with the relevant populations colour-coded. MAGIC[44] was used for the imputation of drop-out values. **b**, Deep imaging of a BM section immunostained with antibodies against Sca1 and Emcn. Scale bar, 20 µm. The experiment was repeated independently ten times with similar results. **c**, Enrichment of the gene expression signatures of Adipo-CAR cells (left), Osteo-CAR cells (middle) and Ng2+ MSCs (right) in previously published transcriptomes of relevant genetically labelled populations[5,7–9]. See Extended Data Fig. 4b for further populations and Supplementary Note 1 for a detailed evaluation of the algorithm used (CIBERSORT). Individual data points represent $n = 3$ replicates; the bars denote the mean. **d**, Boxplots of the scaled expression levels of selected genes in single cells from the Adipo-CAR ($n = 528$), Osteo-CAR ($n = 57$) and osteoblast ($n = 53$) populations. The mean expression of all genes annotated with the gene ontology (GO) term 'ossification' was computed for each cell (right). The boxplot elements are defined in Methods ('Data visualization' section). **e**, Projection of all mesenchymal cell types using PHATE[45] with the time derivatives of the gene expression state, as determined by RNA velocity[20], highlighted as arrows; $n = 1{,}784$ cells. **f**, Comparison of the cell-type frequencies obtained through distinct cell isolation methods used for scRNAseq.

of leptin receptor (*Lepr*) and showed a high transcriptomic similarity to LepR–Cre cells[5,8,19] (Fig. 2c,d). In contrast, Osteo-CAR cells expressed higher osterix (*Sp7*; Fig. 2d and Extended Data Fig. 3n) and lower *Lepr* levels. Finally, we identified a cluster of Ng2- and nestin-expressing mesenchymal cells that show transcriptomic similarity to previously described Ng2^{+}Nestin^{+} MSCs, which we therefore termed 'Ng2^{+} MSCs' (Fig. 2a,c). Pseudo-time analysis using RNA-Velocity[20] placed Ng2^{+} MSCs at the apex of a differentiation hierarchy with osteoblasts, CAR cells, chondrocytes and fibroblasts downstream (Fig. 2e).

In addition, we performed scRNAseq of Lin^{-}cKit^{+} cells to comprehensively cover the haematopoietic stem and progenitor cell (HSPC) populations. This revealed a differentiation continuum[21–24] spanning the megakaryocyte–erythrocyte and lympho–myeloid branches as well as a separate cluster of eosinophil/basophils progenitors (Fig. 1b and Extended Data Fig. 2).

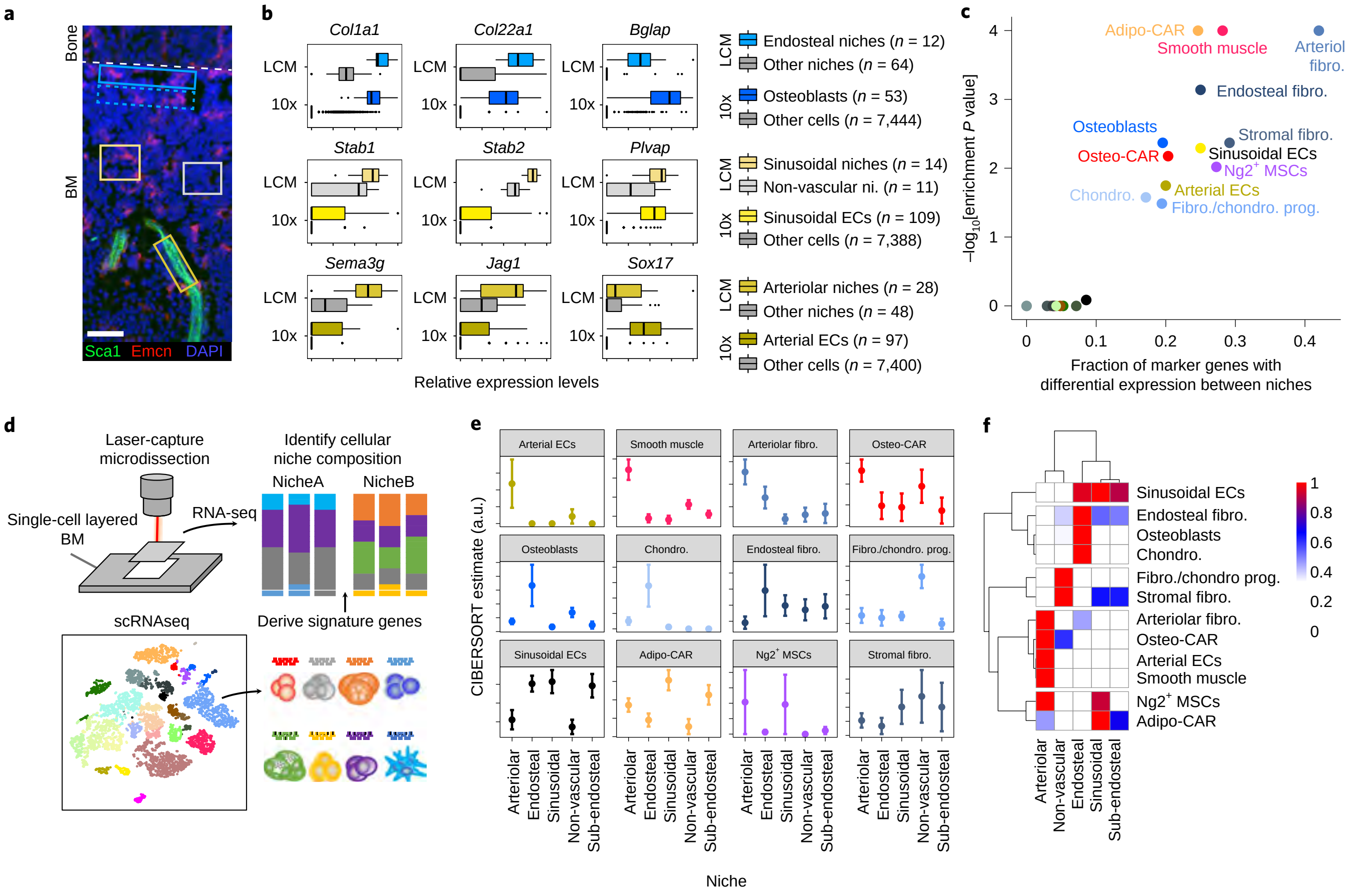


Fig. 3 | Spatial allocation of BM-resident cell types by integrated single-cell and spatial transcriptomics. a, Schematic of the experimental design. Bone sections (12 μm) were stained for arterioles (Sca1) and sinusoids (Emcn). Areas of approximately 14,500 μm^2 surrounding arteries (dark yellow box), sinusoids (yellow box) and the endosteum (blue box), as well as areas with no vessels (grey box) and sub-endosteal areas (dotted blue box) were collected using laser-capture dissection and subjected to RNA-seq. A confocal image is shown for illustrative purposes. Scale bar, 100 μm. For images acquired under the laser-capture dissection microscope and selected areas, see Extended Data Fig. 6b. **b**, Expression of osteoblast-, sinusoid- and arteriole-specific genes in the scRNAseq data (10x) and spatial transcriptomics from different niches (LCM, LCM-seq data); *n*, number of cells or niches as indicated on the graph. The boxplot elements are defined in Methods ('Data visualization' section). **c**, Enrichment of population marker genes (Supplementary Table 1) among genes with differential expression between niches (Supplementary Table 2). The *P* values were derived using a re-sampling-based test; *n* = 25 cell types are shown; all HSPC subtypes were summarized as a single population. **d**, Schematic outline of the computational data-analysis strategy used. **e**, Estimated abundance of different cell types in microscopically distinct niches. The means of the CIBERSORT estimates are shown; the error bars indicate the standard error of the mean. Arteriolar niches, *n* = 28; endosteal niches, *n* = 12; sinusoidal niches, *n* = 14; non-vascular niches, *n* = 11; and sub-endosteal niches, *n* = 11. **f**, Heatmap summarizing the data from **e**. The means of the CIBERSORT estimates are colour-coded according to the legend. Fibro., fibroblast: chondro.; chondrocyte; a.u., arbitrary units.

To assess the impact of isolation strategies on cell-type recovery in the scRNAseq experiments, we compared flushing of BM versus crushing of whole bones and isolation with or without enzymatic digestion. This demonstrated that the majority of the cell populations are found both in flushed and crushed BM but many populations are only released efficiently following intense physical treatment or enzymatic digestion (Fig. 2f and Extended Data Fig. 5a,b). The fibroblast populations, myofibroblasts and Schwann cells were found to be more abundant in crushed bones when compared with flushed bones. Although we confirmed the presence of fibroblast subpopulations in the diaphyseal BM using imaging and spatial transcriptomics (see Figs. 3 and 5), similar cell types deriving from the cortical bone, epiphysis or periosteum might also be present in the scRNAseq datasets from whole bones.

In summary, our dataset spans almost all known BM-resident cell types as well as several previously unknown populations. Osteoclasts, neurons and mature megakaryocytes are not covered by our dataset, probably due to limitations in cell size. The full dataset can be interactively browsed at http://nicheview.shiny.embl.de. A comparison with two recently published scRNAseq studies[25,26] can be found in the supplement (Extended Data Fig. 5c–e).

Spatial allocation of BM-resident cell types by combined single-cell and spatial transcriptomics. Although single-cell transcriptional profiling provides a powerful tool to characterize the identity and molecular make-up of BM cell types, information regarding their spatial distribution is lost. We therefore sought to integrate our scRNAseq dataset with spatially resolved transcriptomics data of the BM. Recently described spatial transcriptomic approaches either require a high image and RNA quality[27–29] or rely on unfixed tissue material[30,31], and have not been successfully adapted to the adult BM. Here, we developed a robust and significantly improved version of laser-capture microdissection coupled with sequencing (LCM-seq)[32]. Our approach makes use of random priming to cope with low quality of input RNA, as encountered in fixed BM sections (Extended Data Fig. 6a; see Methods). Using this strategy,

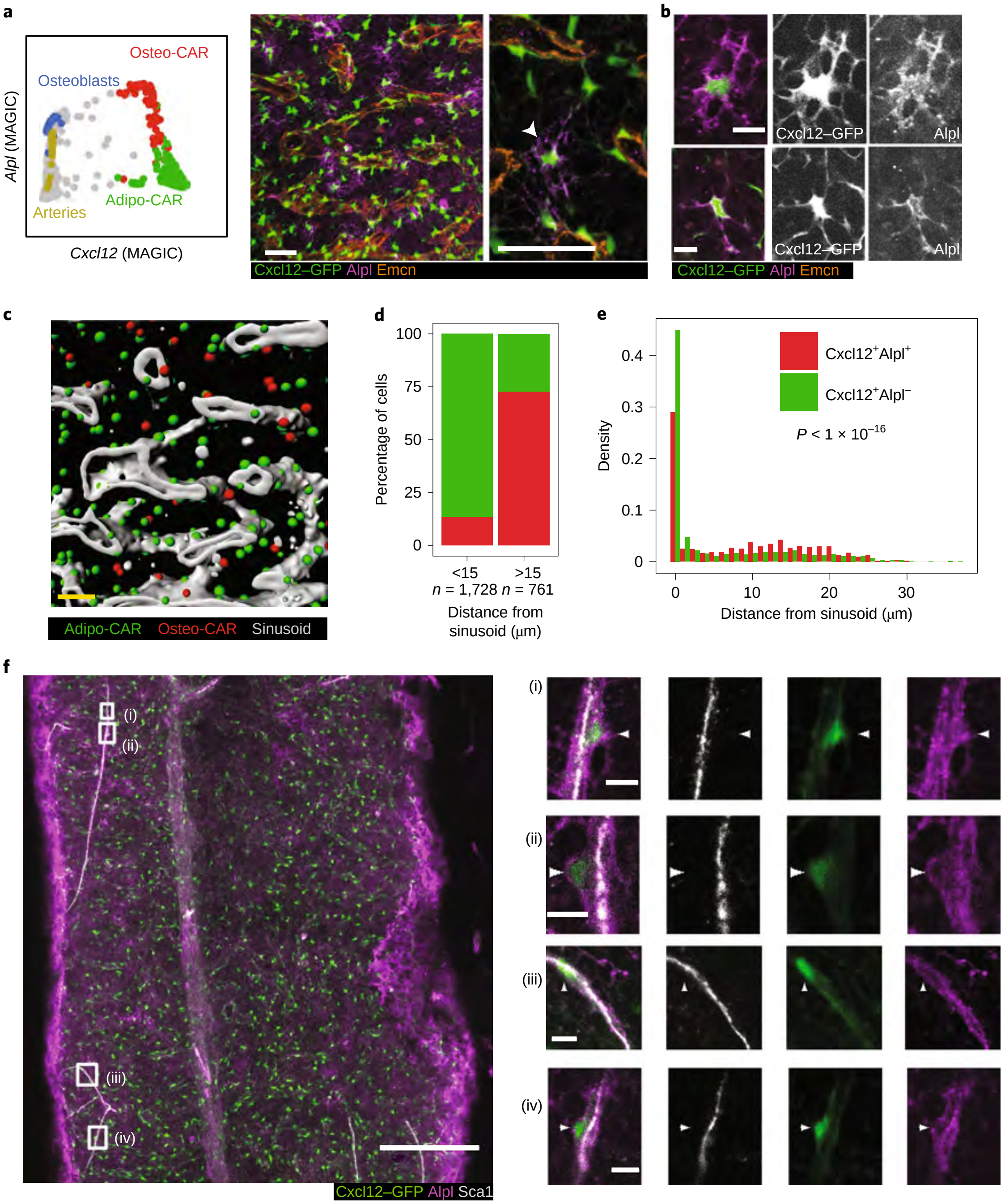


Fig. 4 | Localization of CAR-cell subtypes determined using immunofluorescence. a, Single-cell gene expression levels of *Cxcl12* and *Alpl*, with the relevant cluster identity colour-coded (left). MAGIC[44] was used to impute the drop-out values. Sample regions of interest from a whole-mount image of a femur from a Cxcl12–GFP mouse stained for Alpl and the sinusoidal marker Emcn (right). An Alpl$^+$Cxcl12–GFP$^+$ cell distant from the sinusoids is indicated by an arrowhead. Scale bars, 50 μm. **b**, Additional high-resolution regions of interest, with the Cxcl12–GFP and Alpl signals displayed separately. Scale bars, 10 μm. **c**, Three-dimensional segmentation of the regions of interest from panel **a**. Cxcl12–GFP$^+$ cells were classified as Osteo- or Adipo-CAR cells based on the Alpl signal. See also Extended Data Fig. 8a. **d**, Quantitative distribution of Alpl$^+$ Osteo-CAR cells (red) and Alpl$^-$ Adipo-CAR cells (green) in the sinusoidal and non-sinusoidal niches in the central BM. Data are displayed in summarized form; *n*, number of cells analysed. **e**, Data from **d** displayed as a histogram; $n = 2{,}489$ cells analysed by microscopy. The *P* value was computed using a two-sided Kolmogorov–Smirnov test. **f**, Whole-mount imaging of a femur from a Cxcl12–GFP mouse stained for Alpl and the arteriolar marker Sca1. The arrowheads point to Alpl$^+$Cxcl12$^+$ cells near, but not overlapping with, Sca1$^+$ arteriolar ECs. Scale bars, 200 μm (left) and 10 μm (magnified views of (i)–(iv); right). All imaging experiments were repeated independently two times with similar results. See also Extended Data Fig. 8b.

we generated full-length, high-quality transcriptomic data from LCM-dissected areas of fixed BM sections containing 200–300 cells in a single-cell layer. We applied LCM-seq to 76 microdissected regions collected from the diaphyseal BM, based on the presence or absence of sinusoidal and arteriolar blood vessels or based on the distance from the endosteum, to characterize the endosteal, sub-endosteal, arteriolar, sinusoidal and non-vascular niche composition (Fig. 3a and Extended Data Fig. 6b).

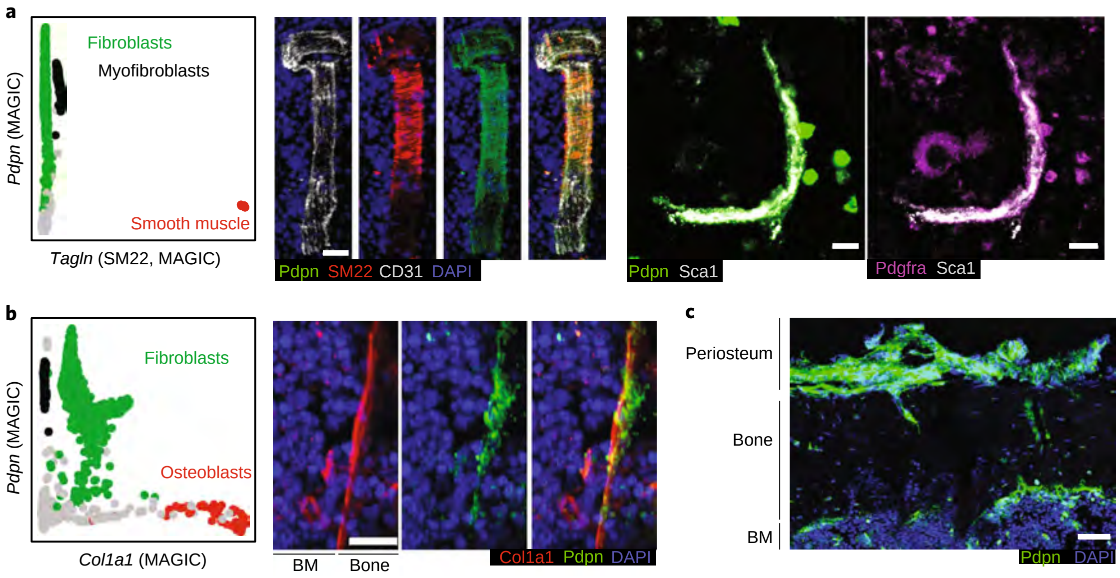


Fig. 5 | Localization of additional mesenchymal cell types determined using immunofluorescence. a, Single-cell gene expression levels of SM22 (*Tagln*) and *Pdpn* with the relevant cluster identity colour-coded (left). Immunofluorescence staining of SM22, Pdpn and CD31 in a BM arteriole (centre). Immunofluorescent staining of Pdpn, Sca1 and Pdgfra in a BM arteriole (right). Scale bars, 20 µm. **b**, Single-cell gene expression levels of *Col1a1* and *Pdpn* with the relevant cluster identity colour-coded (left). Immunofluorescence staining of Col1a1, Pdpn and DAPI at the endosteal surface (right). Scale bar, 20 µm. **c**, Immunofluorescence staining of Pdpn at the endosteum, cortical bone and periosteum. Scale bar, 50 µm. All imaging experiments were repeated independently two times with similar results.

To evaluate our approach, we selected marker genes of cell types known to be specific to the respective niches and compared their expression in scRNAseq and corresponding spatial transcriptomics data (Fig. 3b and Supplementary Table 2). As expected, osteoblast genes were selectively enriched at the endosteum, genes specific for sinusoidal ECs were enriched at regions with a high abundance of sinusoids and arterial endothelial genes were enriched at the arterioles. Marker gene sets for haematopoietic populations, Schwann cells and myofibroblasts were not significantly associated with any of the defined niches, suggesting that these cell types are either relatively evenly distributed across niches (haematopoietic populations) or insufficiently covered in the LCM-seq data (Schwann cells and myofibroblasts; Fig. 3c and Extended Data Fig. 6d,e). In contrast, marker gene expression of the remaining 12 populations differed significantly across niches (Fig. 3c).

To systematically assess the preferential localization of these BM cell types to candidate niches, we estimated the frequencies of the cell populations defined by scRNAseq in the spatially resolved transcriptomics data using the CIBERSORT algorithm[13,14] (Fig. 3d–f). We validated the ability of CIBERSORT to decompose bulk transcriptomes using a single-cell reference and evaluated its performance on assembled pools of 100 cells with known composition (see Supplementary Note 1 and Extended Data Fig. 7). As expected, osteoblasts and chondrocytes were found exclusively at the endosteal niches (Fig. 3e,f). In addition, a specific fibroblast population localized preferentially to the endosteum and was therefore tentatively termed 'endosteal fibroblasts'. In contrast, arterial ECs, smooth muscle cells and a distinct fibroblast population localized specifically to the arteriolar niches (Fig. 3e,f). We termed this fibroblast population 'arteriolar fibroblasts'. Sinusoidal ECs were found in the sinusoidal niches but were also present in the (sub)-endosteal niches, in accordance with a widespread web of sinusoids spanning the entire BM (Fig. 2b and Extended Data Fig. 8a). Adipo-CAR cells were also found predominantly in areas with high sinusoidal occurrence, in line with their similarity to LepR–Cre cells and the reported peri-sinusoidal localization of these cell types[4,5]. In contrast, the previously undescribed Osteo-CAR cells preferentially localized to arteriolar or non-vascular niches, suggesting that the two CAR cell populations occupy distinct niches. Ng2$^+$ MSCs could not be unanimously assigned to a niche, potentially due to additional heterogeneity within this population (Supplementary Note 1 and Extended Data Fig. 7i). In summary, the systematic integration of spatial transcriptomics with single-cell transcriptomics data allowed us to localize the majority of annotated as well as previously unknown BM-resident populations to distinct endosteal, sinusoidal, arteriolar and non-vascular niches.

Validation of cell-type localization. To confirm the spatial relationships of BM cell types identified by LCM-seq, we determined marker combinations specific to the individual populations and performed immunofluorescence imaging of bone sections.

Gene expression analyses suggested that the expression of alkaline phosphatase (*Alpl*) and *Cxcl12* allows discrimination between Osteo-CAR (Cxcl12$^+$Alpl$^+$), Adipo-CAR cells (Cxcl12$^+$Alpl$^-$), and Cxcl12$^-$Alpl$^+$ cell types, such as osteoblasts, Ng2$^+$ MSCs and arterial ECs (Fig. 4a). To confirm the in situ localization of the Adipo- and Osteo-CAR cells, we performed whole-mount immunofluorescence imaging of the long bones of Cxcl12–GFP mice[4]. Co-staining with the sinusoidal marker Emcn revealed that Cxcl12$^+$Alpl$^-$ Adipo-CAR cells in the central BM predominantly ensheathed sinusoids, in line with the results from LCM-seq (Fig. 4a and Extended Data Fig. 8a). In contrast, Cxcl12$^+$Alpl$^+$ Osteo-CAR cells in the central BM typically showed a non-sinusoidal localization and a highly reticular morphology (Fig. 4a–e and Extended Data Fig. 8a). Importantly, Cxcl12$^+$Alpl$^+$Sca1^{low} Osteo-CAR cells were also frequently observed in the immediate vicinity of Sca1$^+$GFP^{dim} arterioles, with GFP^{high} protrusions of Osteo-CAR intensively covering the arteriolar blood vessels (Fig. 4f and Extended Data Fig. 8b). Together, these observations quantitatively confirm the predominantly non-sinusoidal localization of Osteo-CAR cells and qualitatively confirm that these cells localize to both arterioles and non-vascular regions, as predicted from the LCM-seq. The ubiquitous staining of Alpl in the

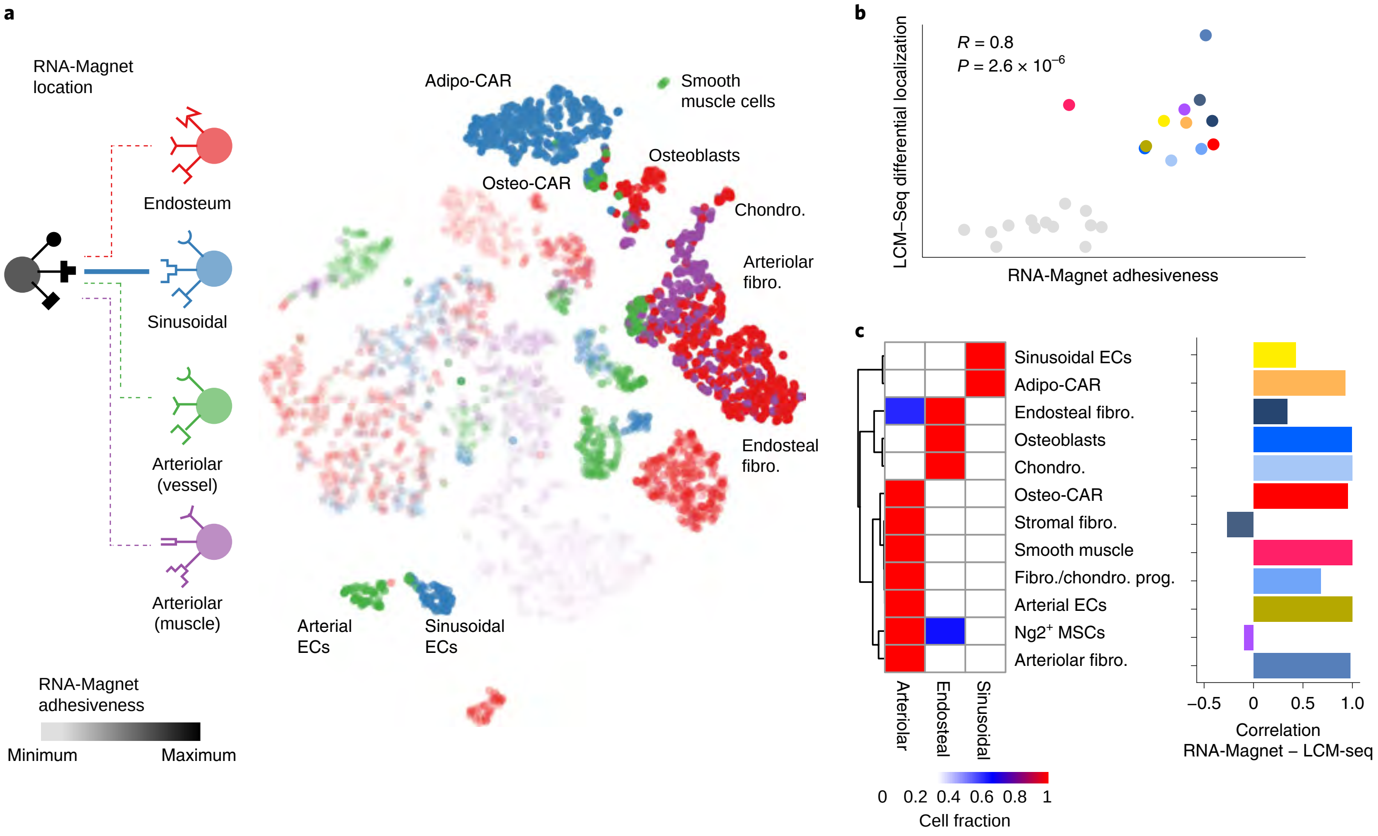


Fig. 6 | Inference of cellular interactions from single-cell gene expression data by RNA-Magnet. a, t-SNE highlighting the cell type each cell is most likely to physically interact with (RNA-Magnet location, indicated by colour) and the estimated strength of adhesion (RNA-Magnet adhesiveness, indicated by opacity). $n = 7{,}497$ cells. **b**, Comparison of the estimated strength of adhesion (RNA-Magnet score) with the degree to which each cell type is differentially localized between niches (spatial transcriptomics; see also Fig. 3). $n = 25$ cell types are shown, all HSPC subtypes were summarized as a single population. The Pearson's correlation (R) and its associated P value are shown. **c**, Heatmap depicting a summary of the inferred localization based on RNA-Magnet (left). The fraction of cells assigned to a certain niche is colour-coded according to the legend. Pearson's correlation between the RNA-Magnet estimate of localization and the LCM-seq estimate of localization across $n = 3$ niches (right). See also Extended Data Fig. 9d.

endosteal regions (probably derived from osteoblasts) prevented the analysis of the CAR cell populations in these areas.

We next used CD31, SM22, Pdpn, Pdgfr, Col1a1 and Sca1 as markers to specifically identify and localize smooth muscle cells (SM22$^+$Pdpn$^-$), the fibroblast populations (Pdpn$^+$Pdgfr$^+$), osteoblasts (Pdpn$^-$Col1a1$^+$) and arterial ECs (Pdpn$^-$CD31$^+$Sca1$^+$; Fig. 5). As suggested by the LCM-seq data, CD31/Sca1-expressing arterioles were enveloped by SM22$^+$Pdpn$^-$ smooth muscle cells and Pdpn$^+$ fibroblasts (Fig. 5a). Arteriolar fibroblasts seem to be the cellular source for the collagen layer of the tunica externa surrounding arterioles (Extended Data Fig. 6c) and probably overlap with previously described peri-arteriolar Pdpn-expressing stromal cells[33]. Moreover, immunofluorescence confirmed the existence of Pdpn$^+$Col1a$^{\text{low}}$ fibroblasts localizing to the bone-facing side of the endosteal lining made up of Pdpn$^-$Col1a$^{\text{high}}$ osteoblasts (Fig. 5b). Notably, Pdpn$^+$ cells were also found at the cortical bone and periosteal regions, which raises the possibility that fibroblast-like cells similar to those in the central marrow also derive from distal regions of the BM (that is, the cortical bone, epiphysis or periosteum; Fig. 5c).

Together, these data validate the ability of our approach to identify previously unknown cell types, such as Adipo- and Osteo-CAR cells, and spatially allocate them in the BM.

The spatial relationships of the BM-resident cell types can be accurately predicted based on scRNAseq data. How the spatial relationships of cell types are established and maintained in complex organs such as the BM remains poorly understood. It has been suggested that the expression of cell adhesion molecules represents an important mechanism that translates basic genetic information into complex three-dimensional patterns of cells in tissues[34]. We compiled a comprehensive list of well-annotated cellular adhesion receptors and their cognate plasma membrane or extracellular matrix bound ligands and developed the RNA-Magnet algorithm to investigate whether the cell-type-specific localizations of BM populations can be predicted by the differential expression of cell adhesion molecules (Extended Data Fig. 9a, Supplementary Table 3 and Supplementary Note 2). RNA-Magnet predicts the potential physical interactions between single cells and selected attractor populations (anchors) based on the expression patterns of the cell-surface receptors and their cognate surface-expressed binding partners. RNA-Magnet provides scores for the strength of attraction (RNA-Magnet adhesiveness) for each cell and a direction indicating the attractor population the cell is most attracted to (RNA-Magnet location). To investigate whether RNA-Magnet is able to recapitulate the spatial relationships of BM cell types, we introduced four anchor populations representing the following niches: osteoblasts for the endosteal niche, sinusoidal ECs for the sinusoidal niche, as well as arterial ECs and smooth muscle cells to represent the arteriolar niches (Fig. 6a). The predicted adhesiveness of BM populations to distinct niches correlated strongly with their degree of differential localization, as measured by spatial transcriptomics (Fig. 6b). Localization was also recapitulated with high accuracy for almost all populations (Fig. 6c and Extended Data Fig. 9b–d), including the differential localization of Adipo- and Osteo-CAR cells to the sinusoidal and arteriolar endothelia, respectively. Interestingly, smooth muscle cells were most attracted to arterial ECs, whereas arteriolar fibroblasts adhered to smooth muscles, recapitulating the consecutive layering observed in blood vessels with the tunica intima (endothelial cells) surrounded by the tunica media (smooth muscle cells) and the tunica externa (extracellular matrix produced by arteriolar fibroblasts)[35]. Together, these observations demonstrate the ability of RNA-Magnet to predict spatial localization from single-cell gene expression data and highlight the importance of cell adhesion proteins for tissue organization in the BM.

Cellular and spatial sources of cytokines and growth factors in the BM. Key biological processes occurring in the BM are thought to be mediated by the coordinated action of a diverse set of cytokines and growth factors. However, the identity of the cytokine-producing cells and their organization into spatial and functional BM niches remain poorly understood. To identify cells serving as the source of HSC maintenance factors, we initially examined our scRNAseq dataset. Our data show that *Cxcl12* and *Kitl* (*Scf*) are indeed expressed by arterial ECs[36] and some mesenchymal cell types[9]. However, their expression was several orders of magnitude higher in the Adipo- and Osteo-CAR populations (Fig. 7a; see also Fig. 4). To confirm Cxcl12 expression from Adipo- and Osteo-CAR cells at the protein level, we developed fluorescence-activated-cell-sorting (FACS) marker strategies to discriminate these cells types and confirmed them using FACS-based index-scRNAseq (Fig. 7b and Extended Data Fig. 10). Comparative analyses demonstrated that Adipo-CAR cells are $CD45^-CD71^-Ter119^-CD41^-CD51^+VCAM1^+CD200^{mid}CD61^{low}$, whereas Osteo-CARs and $NG2^+$ MSCs expressed CD200 and CD61 at high levels (Fig. 7c). Intracellular flow cytometric analyses confirmed that both CAR populations are the main producers of Cxcl12, whereas endothelial cells produced detectable but significantly lower levels (Fig. 7d; see also Fig. 4a,f). The CAR cell populations were also among the main producers of key cytokines required for B-cell and myeloid lineage commitment, such as *Il7* and *m-CSF* (*Csf1*; Fig. 7a). Among all of the BM cell types, the CAR cell populations produced the highest numbers of distinct cytokines and growth factors, and attributed the highest proportion of transcriptional activity to cytokine production, suggesting that they act as 'professional cytokine-producing cells' (Fig. 7e,f). Together, these observations suggest a model in which the differential localization of professional cytokine-producing cells to cellular scaffolds results in the establishment of specific microniches. In line with this, spatial transcriptomics revealed that the five niches investigated displayed unique production patterns of signalling mediators (Fig. 7g,h). Importantly, CAR-cell-derived cytokines were predominantly produced in the arteriolar and sinusoidal niches, and the net production of growth factors and cytokines was significantly higher in the vascular niches compared with non-vascular niches, in line with the preferential localization of Osteo- and Adipo-CAR cells to their respective endothelial scaffolds (Fig. 7g). Together, these observations suggest that the specific localization of professional cytokine-producing BM-resident cells results in the establishment of unique niches with both the arteriolar and sinusoidal niches as key sites for the production of HSC maintenance and differentiation factors.

Systems-level analysis of intercellular signalling interactions of BM cell types. We applied RNA-Magnet to soluble signalling mediators (for example cytokines, growth factors and so on) and their receptors to gain a systems-level overview of potential intercellular signalling interactions (Fig. 8a and Extended Data Fig. 9e). Unlike previous approaches for the reconstruction of signalling networks from single-cell data[37–39], RNA-Magnet incorporates information on surface receptors with low messenger RNA expression, is based on a highly curated list of ligand–receptor pairs (Supplementary Table 3) and specifically identifies the enrichment of signalling interactions between pairs of cell types, with an improved runtime compared with the packages available at present[37] (see Methods).

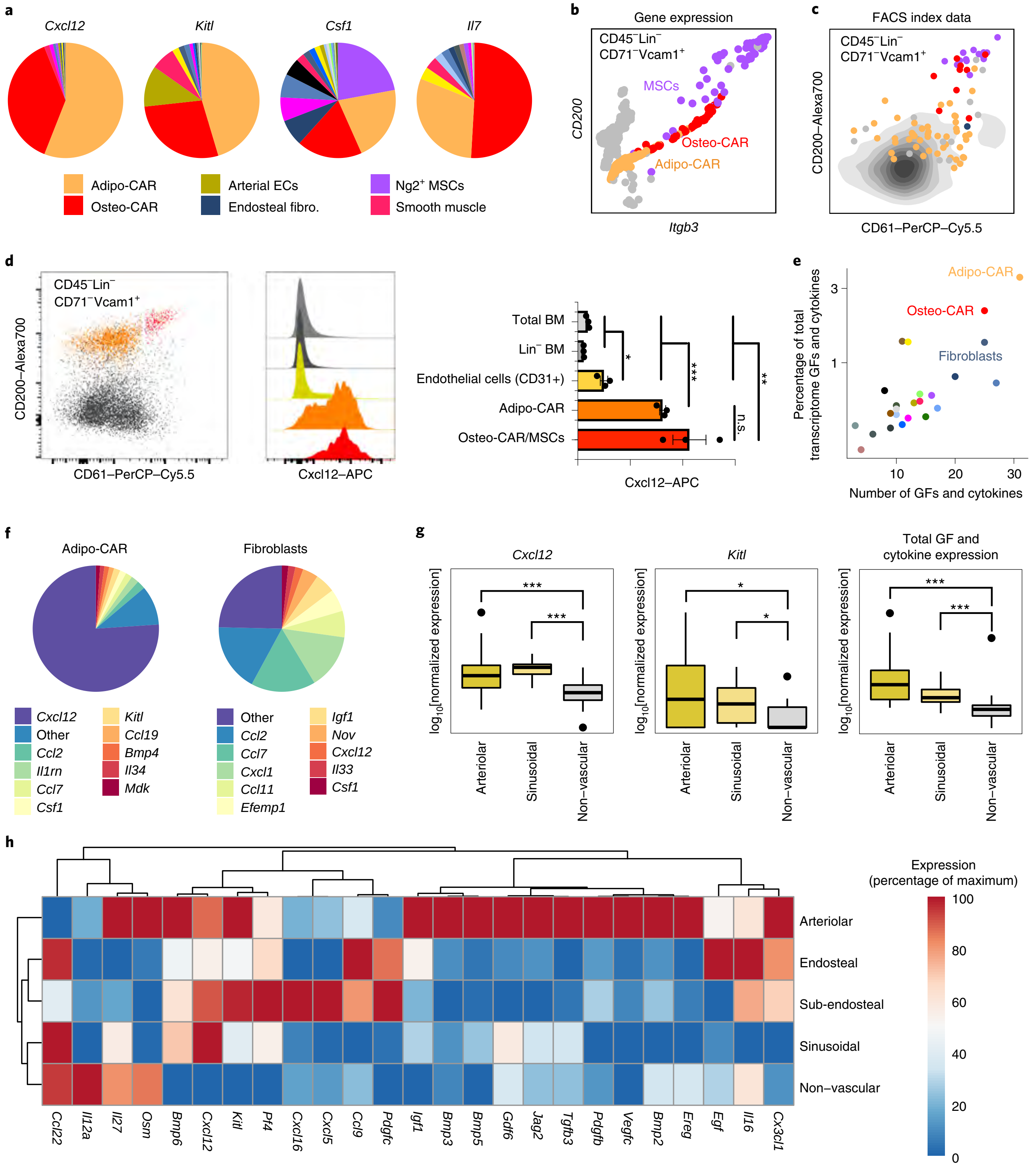
The network obtained from the RNA-Magnet analyses formed two disconnected signalling clusters consisting of either mature immune or non-haematopoietic cells, suggesting that immune and non-immune cells preferentially communicate within their respective groups. For example, many signals potentially sensed by osteoblasts[40] (for example, Bmp-, PTHrP- and FGF-signalling; see Extended Data Fig. 9f) are released by the endosteal fibroblasts described above. In contrast to mature immune cells, the HSPC populations frequently received signals from non-haematopoietic cells, suggesting that HSPCs gradually switch from a mesenchymal to an immune signalling niche following lineage commitment. Importantly, both CAR cell populations were the most important source of the signals sensed by all myeloid and lymphoid progenitors (Fig. 8a). In accordance with the specific localization of the Osteo- and Adipo-CAR populations to arteriolar or sinusoidal scaffolds, an analysis of the net signalling output of distinct local niches implied that lymphoid and myeloid progenitors receive strong input from cytokines produced in the vascular and especially sinusoidal niches (Fig. 8b,c). Together, these analyses support a concept where distinct biological BM processes are mediated by specific combinatorial signalling input from different local niches[41,42] and provide a systems-level view of the signalling interactions in the BM.

Discussion

In this study we combined single-cell and spatially resolved transcriptomics to transcriptionally map all major BM-resident cell types and spatially allocate them to distinct BM niches. Our data clarify the cellular and spatial origins of key cytokines regulating BM haematopoiesis. Key HSC factors, such as Cxcl12 and Scf, are

Fig. 7 | Cellular and spatial sources of key cytokines in the BM. a, Contribution of different cell types to distinct cytokine pools. The mean gene expression across all cells constituting each cell type is compared. **b**, Single-cell gene expression levels of *Cd200* and *Itgb3* (*CD61*) from scRNAseq data in $CD45^-Lin^-CD71^-Vcam1^+$ cells. MAGIC[44] was used to impute the drop-out values. **c**, Surface marker levels of CD200 and CD61 from index-scRNAseq data ($n = 91$) in $CD45^-Lin^-CD71^-Vcam1^+$ cells (see Methods). The colour indicates the most similar cell type from the main dataset, as identified using scmap[46]. **d**, Intracellular FACS analyses of Cxcl12 expression in total BM, lineage-negative BM, Lin^-CD31^+ ECs and CD61/CD200 subpopulations of $CD45^-Lin^-CD71^-CD51^+Vcam1^+$ CAR cells. Data are depicted as the mean and s.e.m. with $n = 3$ per group. Statistics were performed using an unpaired two-tailed *t*-test. The experiment was performed two times independently with a similar outcome. **e**, Quantification of the number of growth factors (GFs) and cytokines expressed by each cell type and the fraction of total mRNA devoted to producing GFs and cytokines. **f**, Relative expression of cytokines and GFs in Adipo-CAR cells and fibroblasts. **g**, Expression of *Cxcl12* (left), *Kitl* (middle) and summed expression of all cytokines and GFs (right) in the indicated niches from spatial transcriptomics data. The *P* values for differential expression of *Cxcl12* and *Kitl* (left and middle) relative to non-vascular niches were computed using the moderated *t*-test used in limma/voom[47] and corrected for multiple testing using the Benjamini–Hochberg method. Right, a two-sided Wilcoxon rank-sum test was applied. Arteriolar niches, $n = 28$; sinusoidal niches, $n = 14$; and non-vascular niches, $n = 11$. The boxplot elements are defined in Methods ('Data visualization' section). **h**, Expression of cytokines, chemokines and GFs in the different niches measured by LCM-seq. Only factors with significant differences between niches are included. $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; and n.s., not significant.

mainly produced by two distinct and previously unappreciated subpopulations, which we have termed Osteo- and Adipo-CAR cells according to their gene expression profile. Besides stem-cell-maintenance factors, the CAR cell subsets produce the highest quantities of cytokines among all BM cell types, including the main cytokines mediating myeloid and lymphoid differentiation, in line with a recent study demonstrating that IL7 and Cxcl12 are produced by the same BM cell type[43]. This suggests that CAR cells act as 'professional cytokine-producing cells' and constitute central niche cells orchestrating different aspects of haematopoiesis. Whereas the more abundant Adipo-CAR cells localize to sinusoidal endothelia, Osteo-CAR cells localize to the non-vascular regions or cover arteriolar endothelia. Accordingly, a high production of stem-cell-maintenance and differentiation factors could be observed in both sinusoidal and arteriolar niches when compared with the non-vascular and endosteal niches. Together, this suggests that both the



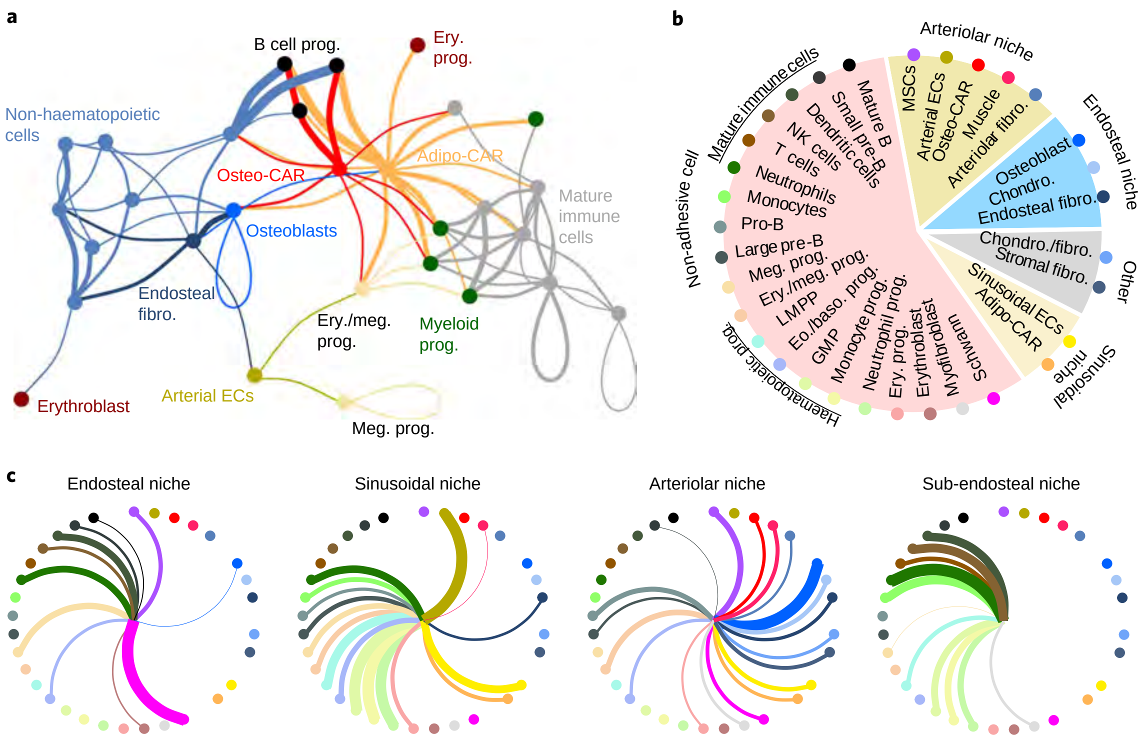


Fig. 8 | Systems-level analysis of signalling potential in the BM. a, Inference of signalling interactions between cell types by RNA-Magnet. If a cell type is enriched in the expression of ligands for receptors expressed by a second cell type, a line is drawn between these cell types, with the colour indicating the ligand-producing cell type and the line width indicating the strength of the enrichment. See Methods for details and Extended Data Fig. 9e for a fully labelled version of the figure. **b**, Summary of the niche composition, as estimated from LCM-seq (see also Fig. 3f). **c**, Inference of the signalling interactions between niches and cell types. The line width indicates the strength of the enrichment for expression of the ligand–receptors pairs. The cell types are arranged as in **b**. Meg., megakaryocyte; ery., erythroid; eo., eosinophil; baso., basophil; prog., progenitor.

arteriolar and sinusoidal vascular scaffolds represent key sites for the production of factors required for HSC maintenance and differentiation, with the Adipo- and Osteo-CAR cell subsets constituting central cellular hubs.

Conceptually, our data support a model where the establishment of unique niches is mediated by the differential localization of professional cytokine-producing cells to cellular scaffolds. Distinct niche milieus might differentially regulate haematopoietic activities, in line with recent data from genetic studies[41]. In the future, it will be of interest to investigate whether such extrinsic, niche-driven variations determine early fate decisions of haematopoietic stem cells[21].

Overall, our approach combining scRNAseq and spatially resolved transcriptomics reveals the molecular, cellular and spatial organization of BM niches and offers a broadly applicable strategy to systemically map the organization of whole organs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-019-0439-6.

Received: 15 July 2019; Accepted: 20 November 2019;
Published online: 23 December 2019

References

1. Ramasamy, S. K. et al. Regulation of hematopoiesis and osteogenesis by blood vessel-derived signals. *Annu. Rev. Cell Dev. Biol.* **32**, 649–675 (2016).
2. Morrison, S. J. & Scadden, D. T. The bone marrow niche for haematopoietic stem cells. *Nature* **505**, 327–334 (2014).
3. Wei, Q. & Frenette, P. S. Niches for hematopoietic stem cells and their progeny. *Immunity* **48**, 632–648 (2018).
4. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12–CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977–988 (2006).
5. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
6. Ding, L. & Morrison, S. J. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* **495**, 231–235 (2013).
7. Kunisaki, Y. et al. Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature* **502**, 637–643 (2013).
8. Asada, N. et al. Differential cytokine contributions of perivascular haematopoietic stem cell niches. *Nat. Cell Biol.* **19**, 214–223 (2017).
9. Greenbaum, A. et al. CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* **495**, 227–230 (2013).
10. Méndez-Ferrer, S. et al. Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).
11. Gomariz, A. et al. Quantitative spatial analysis of haematopoiesis-regulating stromal cells in the bone marrow microenvironment by 3D microscopy. *Nat. Commun.* **9**, 2532 (2018).
12. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
13. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
14. Baron, M. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* **3**, 346–360 (2016).
15. Zhu, L., Lei, J., Klei, L., Devlin, B. & Roeder, K. Semisoft clustering of single-cell data. *Proc. Natl Acad. Sci. USA* **116**, 466–471 (2018).
16. Boulais, P. E. et al. The majority of CD45$^-$ Ter119$^-$ CD31$^-$ bone marrow cell fraction is of hematopoietic origin and contains erythroid and lymphoid progenitors. *Immunity* **49**, 627–639 (2018).
17. Coutu, D. L., Kokkaliaris, K. D., Kunz, L. & Schroeder, T. Three-dimensional map of nonhematopoietic bone and bone-marrow cells and molecules. *Nat. Biotechnol.* **35**, 1202–1210 (2017).

18. Omatsu, Y. et al. The essential functions of adipo-osteogenic progenitors as the hematopoietic stem and progenitor cell niche. *Immunity* **33**, 387–399 (2010).
19. Zhou, B. O. et al. Bone marrow adipocytes promote the regeneration of stem cells and haematopoiesis by secreting SCF. *Nat. Cell Biol.* **19**, 891–903 (2017).
20. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
21. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
22. Karamitros, D. et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat. Immunol*. **19**, 85–97 (2018).
23. Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–e31 (2016).
24. Tusi, B. K. et al. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54–60 (2018).
25. Baryawno, N. et al. A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. *Cell* **177**, 1915–1932 (2019).
26. Tikhonova, A. N. et al. The bone marrow microenvironment at single-cell resolution. *Nature* **569**, 222–228 (2019).
27. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
28. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
29. Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
30. Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
31. Silberstein, L. et al. Proximity-based differential single-cell analysis of the niche to identify stem/progenitor cell regulators. *Cell Stem Cell* **19**, 530–543 (2016).
32. Nichterwitz, S. et al. Laser capture microscopy coupled with Smart-seq2 (LCM-seq) for robust and efficient transcriptomic profiling of mouse and human cells. *Nat. Commun.* **7**, 12139 (2016).
33. Tamura, S. et al. Podoplanin-positive periarteriolar stromal cells promote megakaryocyte growth and proplatelet formation in mice by CLEC-2. *Blood* **127**, 1701–1710 (2016).
34. Gumbiner, B. M. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* **84**, 345–357 (1996).
35. Potente, M. & Mäkinen, T. Vascular heterogeneity and specialization in development and disease. *Nat. Rev. Mol. Cell Biol.* **18**, 477–494 (2017).
36. Xu, C. et al. Stem cell factor is selectively secreted by arterial endothelial cells in bone marrow. *Nat. Commun.* **9**, 2449 (2018).
37. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).
38. Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533–538 (2017).
39. Cohen, M. et al. Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting. *Cell* **175**, 1031–1044 (2018).
40. Siddiqui, J. A. & Partridge, N. C. Physiological bone remodeling: systemic regulation and growth factor involvement. *Physiology* **31**, 233–245 (2016).
41. Pinho, S. et al. Lineage-biased hematopoietic stem cells are regulated by distinct niches. *Dev. Cell* **44**, 634–641.e4 (2018).
42. Itkin, T. et al. Distinct bone marrow blood vessels differentially regulate haematopoiesis. *Nature* **532**, 323–328 (2016).
43. Cordeiro Gomes, A. et al. Hematopoietic stem cell niches produce lineage-instructive signals to control multipotent progenitor differentiation. *Immunity* **45**, 1219–1231 (2016).
44. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716-729 (2018).
45. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
46. Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15**, 359–362 (2018).
47. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

Methods

Mouse experiments. The mice were purchased from the distributors Janvier and Envigo, and housed under specific pathogen-free conditions at the central animal facility of the German Cancer Research Center. All of the animals used were 8- to 12-week-old C56Bl/6J females. All of the animal experiments were performed according to protocols approved by the German authorities (Regierungspräsidium Karlsruhe).

Tissue harvesting and processing. Femurs, tibiae, hips and spines were dissected and cleaned of surrounding tissue. For all cell sorting and flow cytometry analyses, the bones were crushed in cell suspension medium (RPMI 1640 (Sigma) containing 2% fetal bovine serum) using a mortar and pestle. Dissociated cells were filtered through a 40-µm filter, spun down at 1,500 r.p.m. for 5 min and then incubated with 5 ml ACK lysis buffer (Thermo Fisher Scientific) for 5 min at room temperature for red-blood-cell lysis. Neutralization was achieved with 20 ml cell suspension medium. The cells were then lineage-depleted using the Dynabeads untouched mouse CD4 cells kit (Thermo Fisher Scientific) according to the manufacturer's recommendations, using a homemade lineage cocktail.

For cell extraction from bones, crushed bone chips were washed four times each with 10 ml cell suspension medium and then incubated with 10 ml digestion medium (1 mg $\mathrm{ml^{-1}}$ of each collagenase II and dispase in Hanks' balanced salt solution; all from Gibco) for 30 min at 37 °C in a water bath. The cell suspension was then removed and filtered through a 40-µm filter and the digestion reaction was stopped by adding 40 ml cell suspension medium. From this point on, the cells were treated exactly the same as BM cells above.

Flow cytometry. Lineage-depleted bone and BM cells obtained following crushing and digestion were stained with FACS antibodies (Supplementary Table 4) for 30 min on ice and then washed with cell suspension medium. For intracellular CXCL12 staining, the cells were fixed/permeabilized using BD Cytofix/Cytoperm for 30 min on ice, washed with BD Perm/Wash, incubated overnight with anti-CXCL12 antibody (1:10) and then washed with BD Perm/Wash. All flow cytometric analyses were performed using BD Fortessa flow cytometers. Cell sorting was done using BD Aria I, Aria II and Aria Fusion sorters.

scRNAseq—10x Genomics. For scRNAseq using 10x Genomics, bone and BM cells were processed as described earlier. In addition to FACS markers, the cells were stained with a DNA dye (Vybrant DyeCycle Violet, Thermo Fisher Scientific) to exclude debris and ensure that only cells were sorted for droplet-based scRNAseq. For this purpose, 2.5 µl $\mathrm{ml^{-1}}$ Vybrant dye in cell suspension medium was incubated with 3×10^6 cells at 37 °C for 30 min in a water bath. Following the incubation, the cells were placed on ice and a total of $1.0–1.5\times10^4$ events were sorted immediately for each experiment into 15 µl PBS containing 2% fetal bovine serum. The cell numbers were confirmed using a LUNA automated cell counter (Logos Biosystems). A volume of 33.8 µl of the cell suspension was used as input without further dilution or processing, with final concentrations around 100–200 cells $\mathrm{\mu l^{-1}}$. Reverse transcription and library construction were carried out following the Chromium single cell 3′ reagent v2 protocol (10x Genomics) according to the manufacturer's recommendations. Total complementary-DNA synthesis was performed using 14 amplification cycles, with final cDNA yields ranging from approximately 2 ng $\mathrm{\mu l^{-1}}$ to 10 ng $\mathrm{\mu l^{-1}}$. The 10x Genomics sequencing libraries were constructed as described and sequenced on an Illumina NextSeq 500, with read lengths of 26 + 58 or 26 + 98.

FACS-indexed scRNAseq. Lineage-depleted BM cells were obtained by crushing and stained with the following antibodies on ice for 30 min: CD41, CD45, CD51, CD61, CD71, CD200, Ter119 and VCAM1 (Supplementary Table 4). Indexed single-cells were sorted into 5 µl Smart-Seq2 lysis buffer (2 µM Oligo-dT30VN primer, 2 mM dNTP mix (10 mM each; NEB), 1:50 RNAse inhibitor (Promega) and 1:125 10% Triton X-100 (Sigma-Aldrich)) and immediately snap frozen in an ethanol and dry-ice bath. The plates were kept at −80 °C until processing. Following 3 min at 72 °C, cDNA amplification was performed using a modified Smart-Seq2 protocol by adding 5 µl of reverse transcription mix containing 1×SMART first strand buffer (Clontech), 2 mM dithiothreitol (Clontech), 2 µM template-switching oligo (Exiqon), 10 U $\mathrm{\mu l^{-1}}$ SMARTScribe (Clontech) and 10 U $\mathrm{\mu l^{-1}}$ RNAsin plus (Promega). Transcriptome amplification was performed using 1×KAPA HiFi HS MM and 0.1 µM ISPCR primer, with 21 PCR enrichment cycles. Libraries were constructed using Tn5 produced in house[48] at a 1:100 dilution and sequenced on an Illumina NextSeq 500 sequencer, with 75 cycles of single-end sequencing.

Bone preparation and sectioning for immunofluorescence staining. Femurs were dissected, cleaned of muscle and immediately placed in 4% paraformaldehyde at 4 °C for 30 min. Subsequently, the femurs were washed three times in 1×PBS before incubating with 15% sucrose for 2 h, followed by 30% sucrose for another 2 h. All incubations were performed at 4 °C. The femurs were placed in OCT compound and frozen at −80 °C until sectioning.

For sectioning, a cryotome (Thermo Fisher) was used to generate 12-µm sections at −20 to −22 °C, which were then transferred to slides using the CryoJane Tape-Transfer System (Leica Biosystems). The slides were post-fixed with 4% paraformaldehyde for 1 min to enhance the adherence of sections to the slide, then washed for 2 min by dipping in PBS. The sections were stained with antibodies in blocking buffer (PBS containing 10% goat serum and 0.2% Triton X-100; see Supplementary Table 4 for the antibodies used) at 4 °C overnight and then washed by dipping into PBS for 1 min. Secondary antibody staining was done in blocking buffer at room temperature for 2 h. The sections were imaged using an LSM710 microscope (Zeiss) equipped with a polychromatic META detector (458, 488, 514, 561, 594 and 633 nm lasers) and an Olympus FV3000 confocal laser scanning microscope with four GaAsP spectral detectors, fluorescence recovery after photobleaching and fluorescence resonance energy transfer (405, 445, 488, 514, 561, 594 and 640 nm lasers). Imaris software (v8.41) was used for the data analysis and representation.

Whole-mount imaging of immunostained mouse femurs. The employed whole-mount staining protocol has been previously described in detail[11]. Briefly, femurs were isolated—the surrounding connective and muscle tissue were thoroughly removed using precision wipes (Kimberly-Clark Professional)—and fixed in 2% paraformaldehyde in PBS (6 h, 4 °C). After dehydration in 30% sucrose in PBS (72 h, 4 °C) the femurs were embedded in Tissue-Tek OCT compound (Sakura), snap frozen in liquid nitrogen and bi-sectioned using a cryotome (Leica) to expose the BM. The thick BM slice obtained was washed in PBS (3 × 5 min, room temperature), incubated in blocking solution overnight (0.2% Triton X-100 (Sigma), 1% bovine serum albumin (Sigma) and 10% donkey serum (Jackson ImmunoResearch) in PBS) at 4 °C. Antibody stainings (see Supplementary Table 5) were performed in 1.5 ml Eppendorf tubes with gentle agitation on a laboratory rocker at 4 °C. The thick BM slices were incubated with primary antibodies (diluted in 400 µl blocking solution) for 3 d, followed by three 1-h washing steps with PBS. Secondary antibody staining was performed analogously for 3 d, followed by three 1-h washing steps with PBS. The immunostained thick BM slices were subsequently immersed into 400 µl RapiClear 1.52 (SUNJIN LAB) for a minimum of 6 h at 4 °C to increase the optical transparency. Images were acquired on an SP8 Leica confocal microscope system using a white light laser source for optimal excitation and hybrid detectors for light detection. Imaris software (v8.41) was used for the data analysis and representation.

Bone sectioning for LCM-seq. Femurs for LCM-seq were carefully processed following the guidelines of good RNA work practice. The femurs were harvested and cleaned as quickly as possible, placed immediately in ice-cold 4% paraformaldehyde, fixed for 30 min and dehydrated in 15 and 30% sucrose solutions, prepared using RNase-free sucrose powder (Acros Organics), for 2 h each. All incubation steps were carried out on ice. The femurs were then flash frozen in a 2-methylbutane and dry-ice bath and stored overnight at −80 °C. Bone sectioning for RNA retrieval was performed in a sterilized cryotome, the blades were wiped with RNaseZAP (Sigma) and slides were stored on dry ice until staining (on the same day). For the staining, we adapted a shortened immunostaining protocol[49] starting with thawing the slide quickly at room temperature, then incubating with the primary antibody (1:20 to 1:40, depending on the antibody) on an aluminium rack on ice for 10 min, washing in ice-cold PBS for 30 s and adding the secondary antibody for 5 min on ice. Higher antibody concentrations may be needed for antibodies of lower quality. The final wash and dehydration were performed by dipping the slides for 30 s in each of the following: ice-cold PBS, RNase-free H_2O, and 70, 95 and 100% ethanol, in this order.

LCM-seq. Bone sections were processed using the Zeiss PALM MicroBeam Axio Observer Z1 (Zeiss) with a monochromatic Axiocam 506 mono camera, shooting laser at 355 nm and filter sets FS18 C adv (DAPI), FS44 C adv (FITC) and FS45 C adv (mCherry). Image acquisition and sample isolation were carried out using a LD Plan-NEOFLUAR ×20/0.4 objective with the adjustment ring set to one (Zeiss). The cutting energy was set to 45 and focus to 67, while laser pulse catapult (LPC) energy was used with a delta of 12. We used the 'CloseCut+AutoLPC' option for cutting and shooting. No more than four sections (two per slide) were processed and scanned in parallel, keeping collection time after staining under 30 min. A three-channel colour image was acquired for each sample, as well as before/after LPC images and metadata including the day of collection, slide number and distance of the area of interest from the bone lining. On average, areas of 14,500 $\mathrm{\mu m^2}$ (corresponding to around 200–300 cells) were isolated from different BM districts and collected separately in 200 µl AdhesiveCap opaque Eppendorf tubes. After LPC, each collection lid was covered with 15 µl of a 1:16 dilution of proteinase K in PKD buffer (Qiagen), incubated at room temperature for 5 min and subsequently snap frozen in dry ice and stored at −80 °C overnight. For reverse crosslinking, the samples were thawed for 5 min at room temperature, centrifuged for 30 s to collect all the liquid at the bottom of the tubes and incubated for 1 h at 56 °C on a PCR block[50]. After incubation, the samples were resuspended in 100 µl TRI reagent (Sigma-Aldrich) under a laminar flow hood and stored in eight-PCR-tube strips at −80 °C until RNA extraction. The RNA extractions and library construction were performed in batches of 16. In brief, after thawing and spinning, each sample was transferred to a 1.5 ml Eppendorf tube and 20 µl chloroform was added. Phase separation was achieved at room temperature after vigorous shaking and centrifugation at 12,500 r.p.m. for 5 min, 40 µl of the aqueous phase was collected from each sample and added to 75.5 µl isopropanol and glycoblue (Invitrogen) diluted 1:150 as a co-precipitant. The RNA was then dehydrated at −80 °C for at least 24–36 h and precipitated by centrifugation at 4 °C and maximum speed. The supernatant was removed and the pellet was washed once with 70% ethanol, air dried and resuspended in 8 µl nuclease-free water. Library preparation followed the SMARTER stranded total RNA-Seq kit v2—Pico input mammalian (Takara Bio). Due to the degraded nature of the input material, we omitted the fragmentation step and 16 enrichment cycles were used in the final RNA-seq amplification. The libraries were then eluted in a minimal volume of 12 µl and sequenced on an Illumina NextSeq 500 with 75 bp single-end reads.

Bioinformatics data analysis of single-cell data. Raw sequencing data were processed using the CellRanger pipeline (10x Genomics) or kallisto[51] (for indexed scRNAseq). Count tables were loaded into R and further processed using the Seurat R package[52]. We removed all cells with fewer than 500 distinct genes observed or cells with more than 5% of unique molecular identifiers stemming from mitochondrial genes. Principal component analysis was then performed on significantly variable genes and the first 16 principal components were selected as input for clustering and t-SNE based on manual inspection of a principal component variance plot ('PC elbow plot'). Clustering was performed using the default method from the Seurat package, with the resolution parameter set to five. Whereas lower-resolution parameters caused biologically distinct groups with a low number of cells to be merged into single clusters (for example, sinusoids and arterioles were merged into a single cluster), this relatively large parameter resulted in groups with a high number of cells to be split into an undesirable number of subgroups. We therefore computed the mean scaled gene expression values for each cluster and performed hierarchical clustering of the means using a correlation distance. Clusters with correlations greater than 0.8 were then merged, resulting in the final clusters displayed in Fig. 1b. The marker genes for each population were identified using the FindMarkersAll function and ROC-based test statistics. The Seurat label transfer routines[53] were used to map cells to a reference (Extended Data Fig. 5), except in the context of single-cell index-sorting (Fig. 7c and Extended Data Fig. 10) where the cell number was insufficient for this method and scmap[46] was used instead. All mapping results were confirmed by the analysis of the marker-gene expression (Extended Data Fig. 10b). For Fig. 2e, RNA velocity was run with a neighbourhood size of 50 and a linear velocity scale. The result was insensitive to the choice of parameters, except that the relative arrow lengths varied.

Bioinformatics data analysis of LCM-seq and bulk RNA-seq data. Reads were aligned to version 38.73 of the mouse genome using STAR[54], and reads falling on exons of genes were counted using htseq-count[55]. Samples containing fewer than one million reads on feature as well as two outlier samples identified by principal component analysis were removed from the dataset. Differential expression between niches was determined using the limma/voom workflow[47] while accounting for batch (that is, slide number). The cell-type proportions in the different samples were then estimated using a custom R implementation of the CIBERSORT algorithm[13] run on raw count data. CIBERSORT requires the specification of a population-specific gene expression 'signature matrix'; for that purpose, average gene expression profiles were computed for each cell type and 1,571 population-specific genes (Extended Data Fig. 7a; genes were defined by specificity to a given population of 0.8 or greater, as quantified from areas under the ROC curve). To simplify the analyses, we merged the highly similar HSPC subtypes into one population for CIBERSORT. Beyond the selection of genes, the CIBERSORT algorithm does not have any free parameters. Further considerations underlying the analyses using CIBERSORT, as well as a discussion on the impact of selecting different sets of genes, are detailed in the Supplementary Note 1.

For the reanalysis of published RNA-seq data (Fig. 2c and Extended Data Fig. 4a,b), count tables were created from raw sequencing data available in the Gene Expression Omnibus (GEO; GSE89811 and GSE48764) or count tables were downloaded from the GEO (GSE109125). For published microarray data, raw expression matrices were downloaded from the GEO (GSE33158, GSE43613 and GSE57729). It has previously been shown[56] that CIBERSORT can be reasonably applied to decompose microarray data using a RNA-seq reference.

Data visualization. All plots were generated using the ggplot2 (v. 3.1.0) and pheatmap (v. 1.0.10) packages in R 3.4.1. Boxplots are defined as follows: the middle line corresponds to the median; the lower and upper hinges correspond to first and third quartiles, respectively; the upper whisker extends from the hinge to the largest value no further than 1.5× the inter-quartile range (or the distance between the first and third quartiles) from the hinge and the lower whisker extends from the hinge to the smallest value at most 1.5× the inter-quartile range of the hinge. Data beyond the end of the whiskers are called 'outlying' points and are plotted individually[57].

RNA-Magnet. RNA-Magnet is documented in Supplementary Note 2.

Statistics and reproducibility. Statistical analyses were performed using R. Summarized views on data that underlie the statistical tests are available in the Source Data files. The statistical details for each experiment are also provided in the figure legends. All microscopy-based experiments were performed independently at least four times.

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

Data availability

The LCM-seq and scRNAseq that support the findings of this study have been deposited in the GEO under accession code GSE122467. Data are further available in processed form for download and interactive browsing at https://nicheview.shiny.embl.de. Previously published sequencing data that were reanalysed here are available under the accession codes GSE89811, GSE48764, GSE109125, GSE33158, GSE43613 and GSE57729 (bulk transcriptome studies reanalysed in Fig. 2c and Extended Data Fig. 4a,b), or GSE128423 and GSE108892 (single-cell transcriptome studies reanalysed in Extended Data Fig. 5). Source Data for Figs. 1–6 and 8 and Extended Data Figs. 1–5, 7 and 10 are available online. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Code availability

Our implementation of RNA-Magnet and CIBERSORT, as well as vignettes for re-creating key analysis steps are available at https://github.com/veltenlab/rnamagnet.

References

48. Hennig, B. P. et al. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3* **8**, 79–89 (2018).
49. Nichterwitz, S., Benitez, J. A., Hoogstraaten, R., Deng, Q. & Hedlund, E. In *RNA Detection. Methods in Molecular Biology, vol 1649* (ed Gaspar, I.) (Humana Press, New York, NY, 2018).
50. Thomsen, E. R. et al. Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat. Methods* **13**, 87–93 (2016).
51. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
52. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
53. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
54. Dobin, A et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
55. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
56. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. in *Cancer Systems Biology, vol. 1711* (ed. von Stechow, L.) 243–259 (Humana Press, New York, 2018).
57. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer New York, 2009).
58. Shay, T. & Kang, J. Immunological genome project and systems immunology. *Trends Immunol.* **34**, 602–609 (2013).
59. Worthley, D. L. et al. Gremlin 1 identifies a skeletal stem cell with bone, cartilage, and reticular stromal potential. *Cell* **160**, 269–284 (2015).

Acknowledgements

We would like to thank M. Milsom, A. Grozhik, J. Velten, I. Lohmann, B. Velten and the members of the Steinmetz, Haas and Trumpp labs for their helpful discussions and critical proofreading of the manuscript. The Cxcl12–GFP mice were originally from T. Nagasawa. We thank K. Bauer and J. Mallm from the DKFZ Single-cell open lab, D. Krunic from the DKFZ microscopy core, M. Paulsen from the EMBL flow cytometry core facility, M. Eich from the DKFZ flow cytometry, the EMBL genomics core facility, S. Terjung from the EMBL Advanced Light Microscopy Facility (ALMF), the Carl Zeiss AG, J. Schnell, L. Becker, S. Renders, P. Werner and S. Sood for technical support. This work was supported by the SFB873, FOR2674 and FOR2033 funded by the Deutsche Forschungsgemeinschaft (DFG), the SyTASC consortium (Deutsche Krebshilfe) and the Dietmar Hopp Foundation (all to A.T.); the US National Institutes of Health (grant no. P01HG00020527) and the ERC (grant no. 742804) to L.M.S; and the José Carreras Foundation for Leukemia Research (grant no. DCJLS 20R/2017 to L.V., A.T., L.M.S. and S.H.).

Author contributions

C.B. and J.A.-S. developed the experimental methods and performed the majority of the experiments, with conceptual input from S.H., L.V., A.T. and L.M.S. L.V. analysed the data with conceptual input from S.H. and the other authors. L.V., C.B.

and L.M.S. developed RNA-Magnet. S.H. and L.V. conceived the study, supervised the experimental work and wrote the manuscript, with contributions from C.B., J.A.-S., A.T. and L.M.S. F.G. and P.H.-M. provided experimental support for this work. P.M.H. and C.N.-A. performed the whole-mount imaging. All authors have carefully read the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

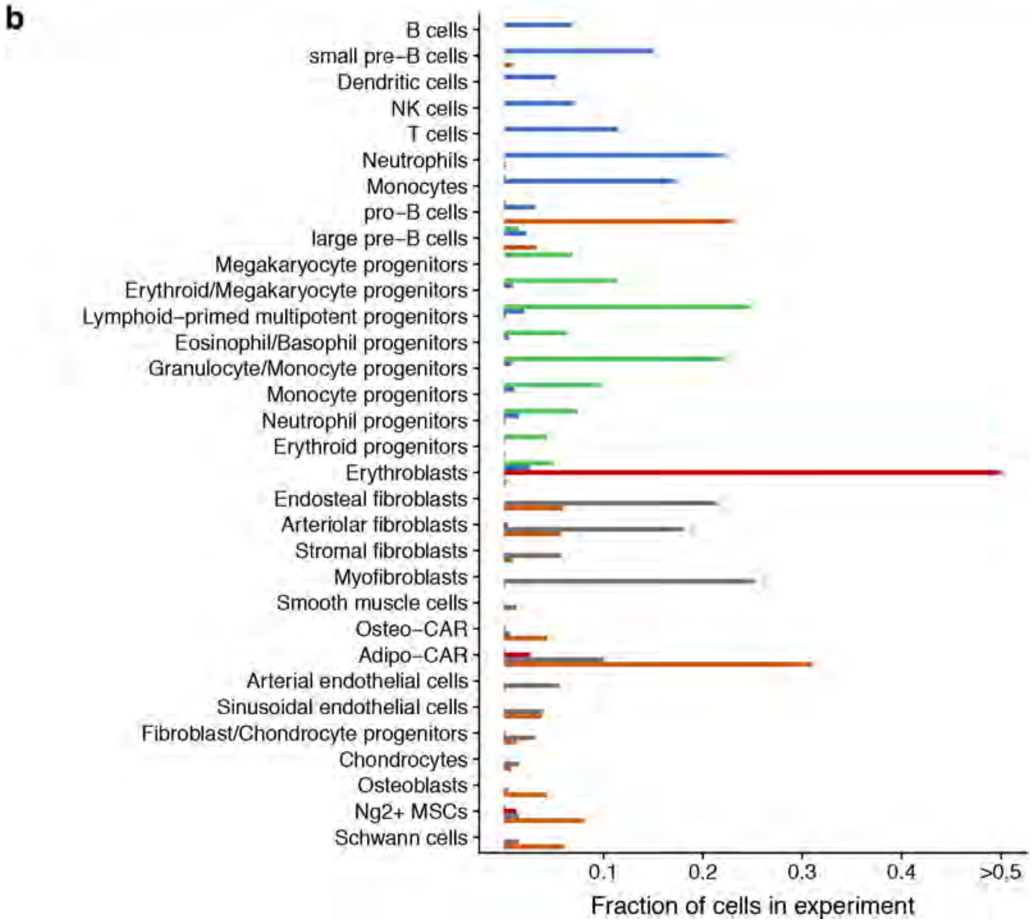
Extended data is available for this paper at https://doi.org/10.1038/s41556-019-0439-6.

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-019-0439-6.

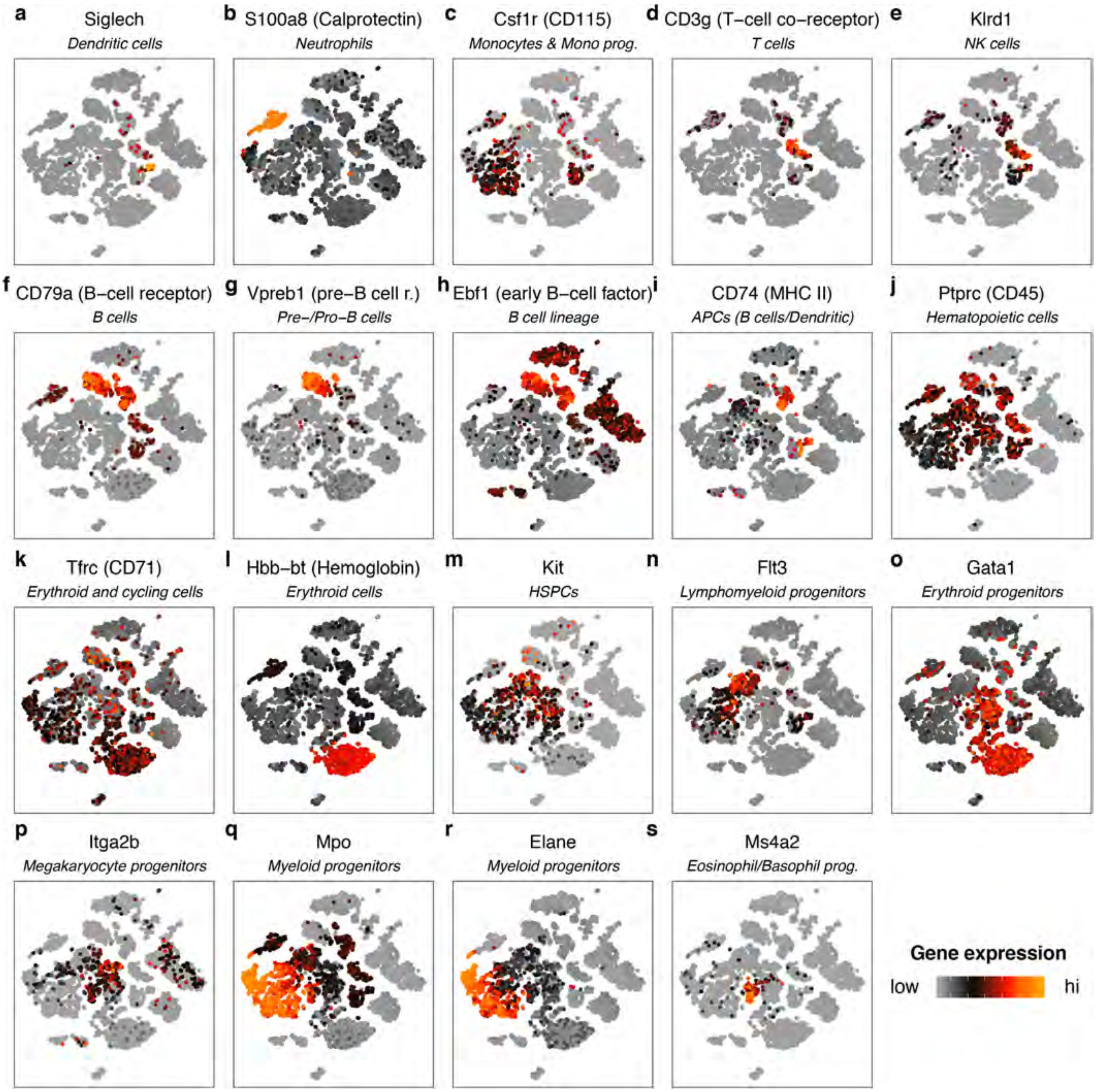
Correspondence and requests for materials should be addressed to L.V., L.M.S., A.T. or S.H.

a

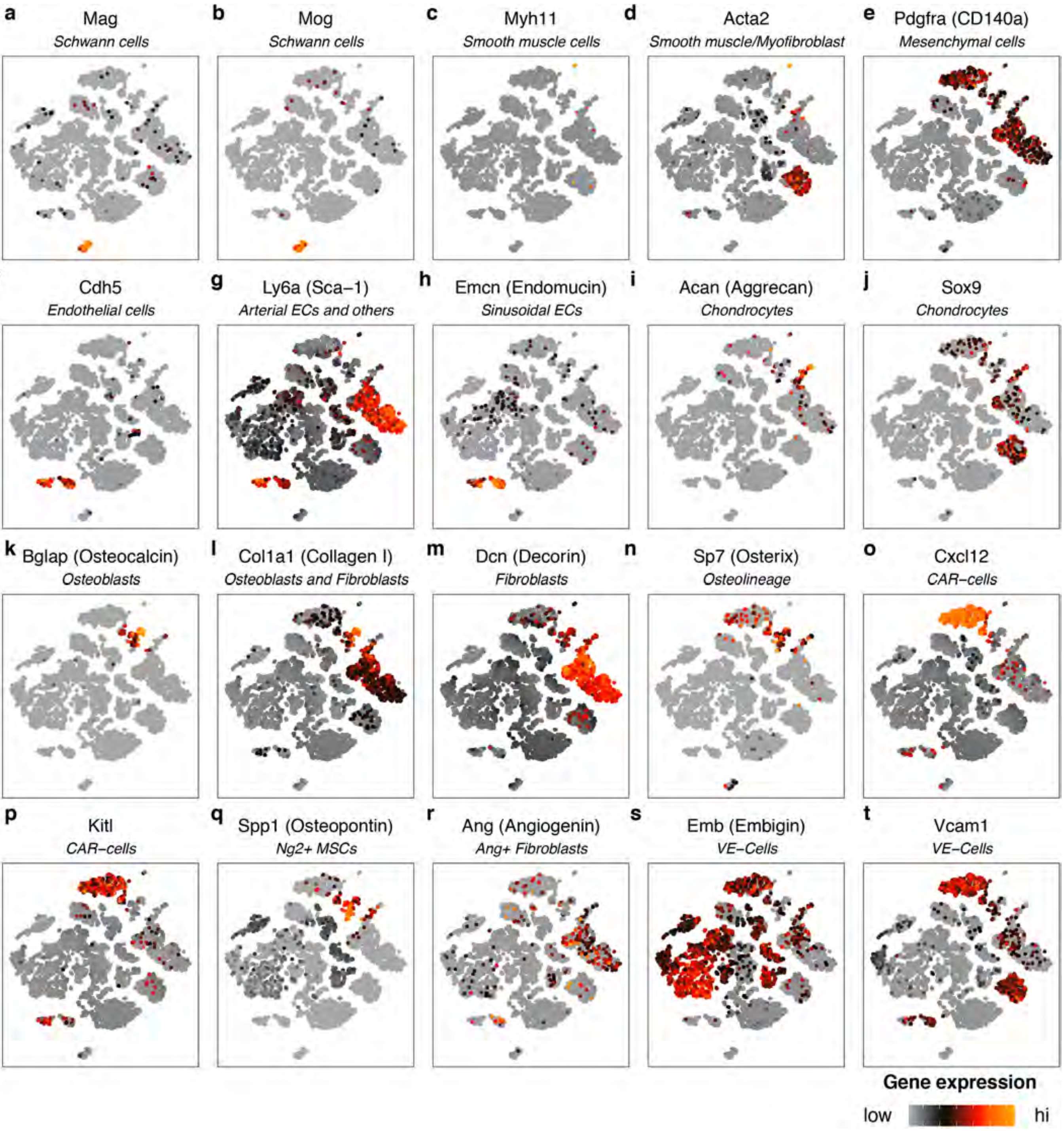
Population	Color	% of total
Total Bone Marrow	🟦	24.1%
cKit+ BM	🟩	1.1%
CD45– BM	🟥	2.8%
CD45–CD71– BM	🟧	0.1%
CD45–CD71– digested bone	⬛	0.7%



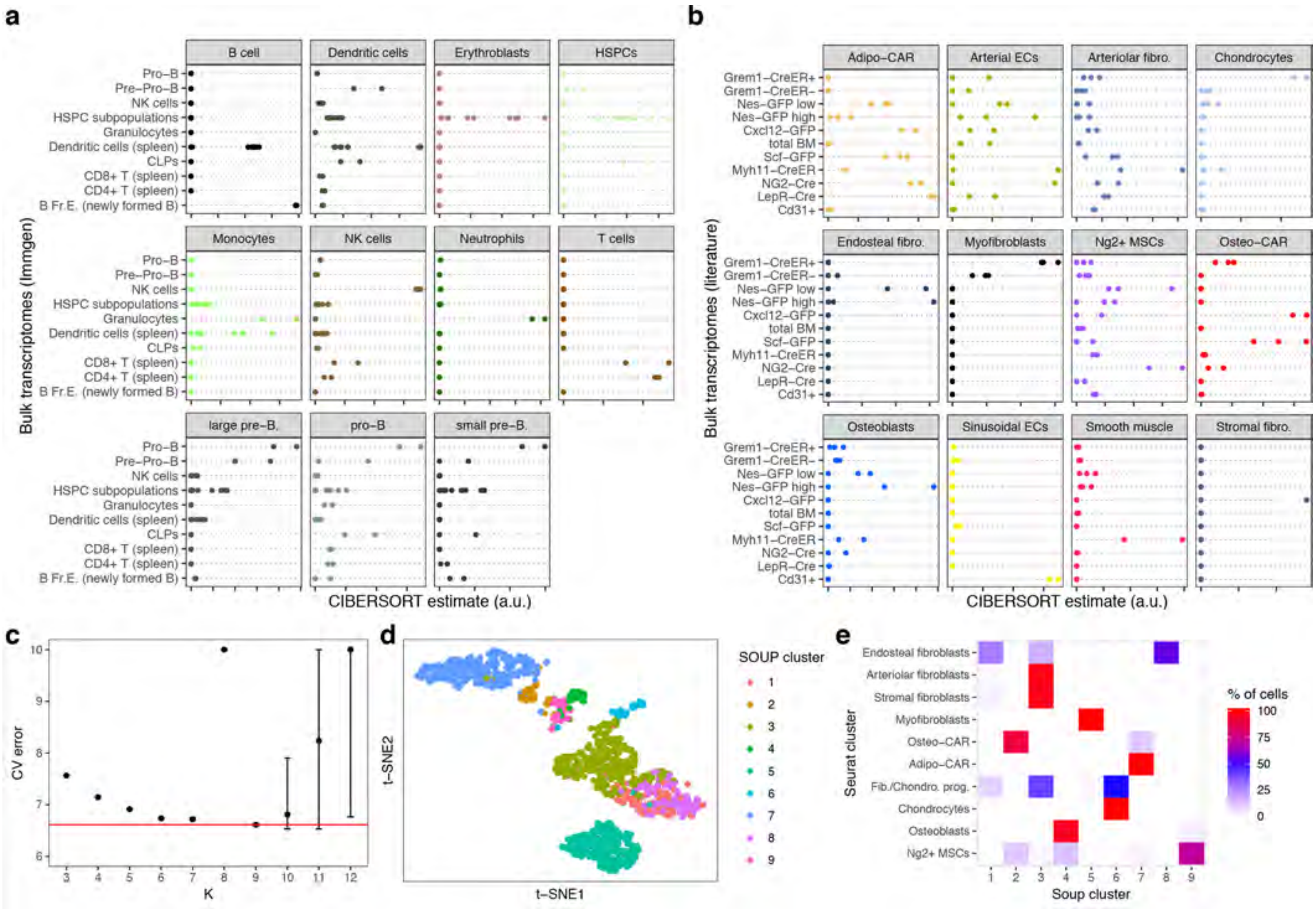
Extended Data Fig. 1 | Cellular composition of populations defined by flow cytometry. a, Abundance of different gates as fraction of total BM. **b**, Quantification of cell type composition for each FACS gate shown in main Fig. 1a.



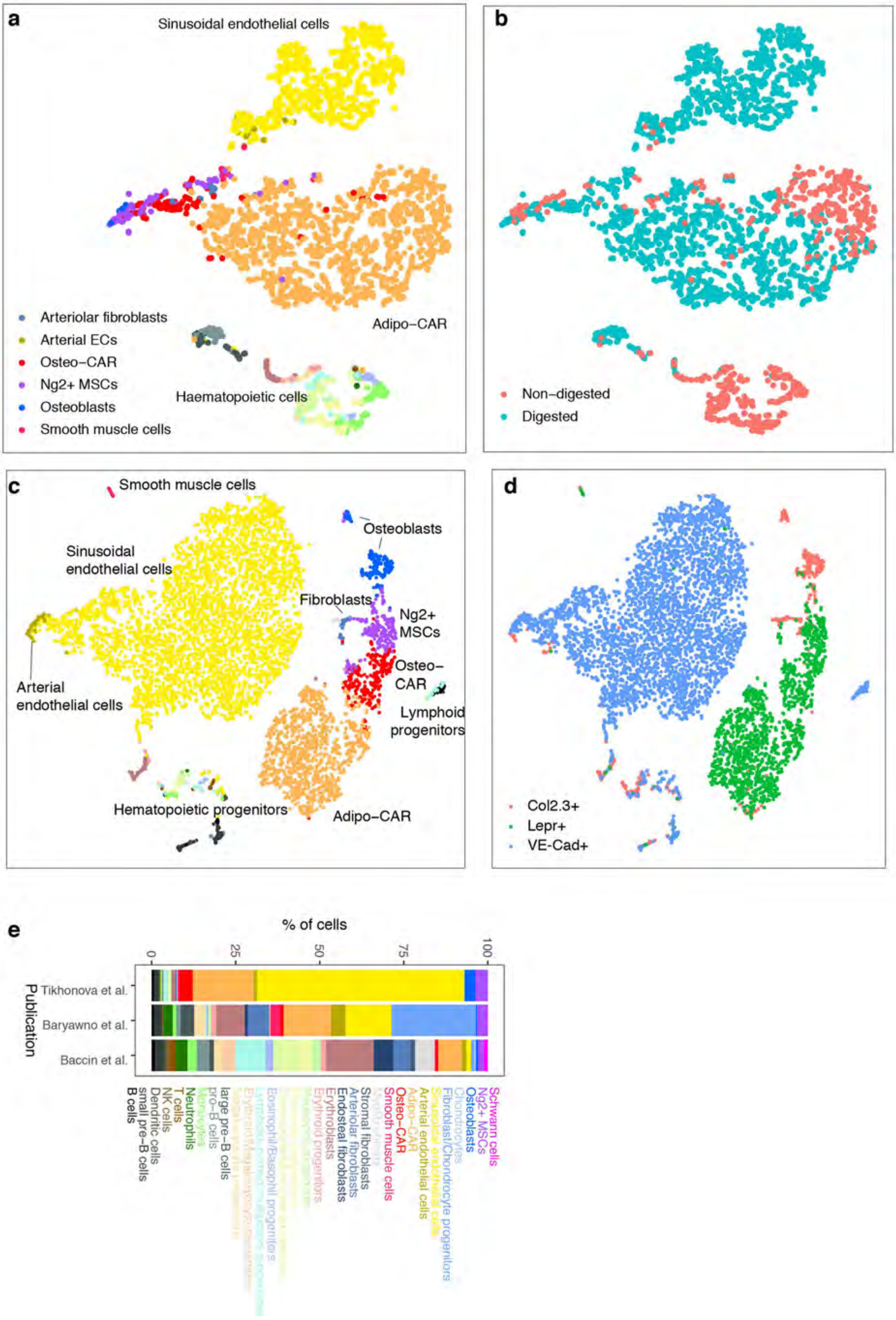
Extended Data Fig. 2 | Expression of marker genes for haematopoietic populations highlighted on t-SNE. For full lists of marker genes, see Supplementary Table 1. n=7497 cells are shown.



Extended Data Fig. 3 | Expression of marker genes for non-haematopoietic populations highlighted on t-SNE. For full lists of marker genes, see Supplementary Table 1. n=7497 cells are shown.

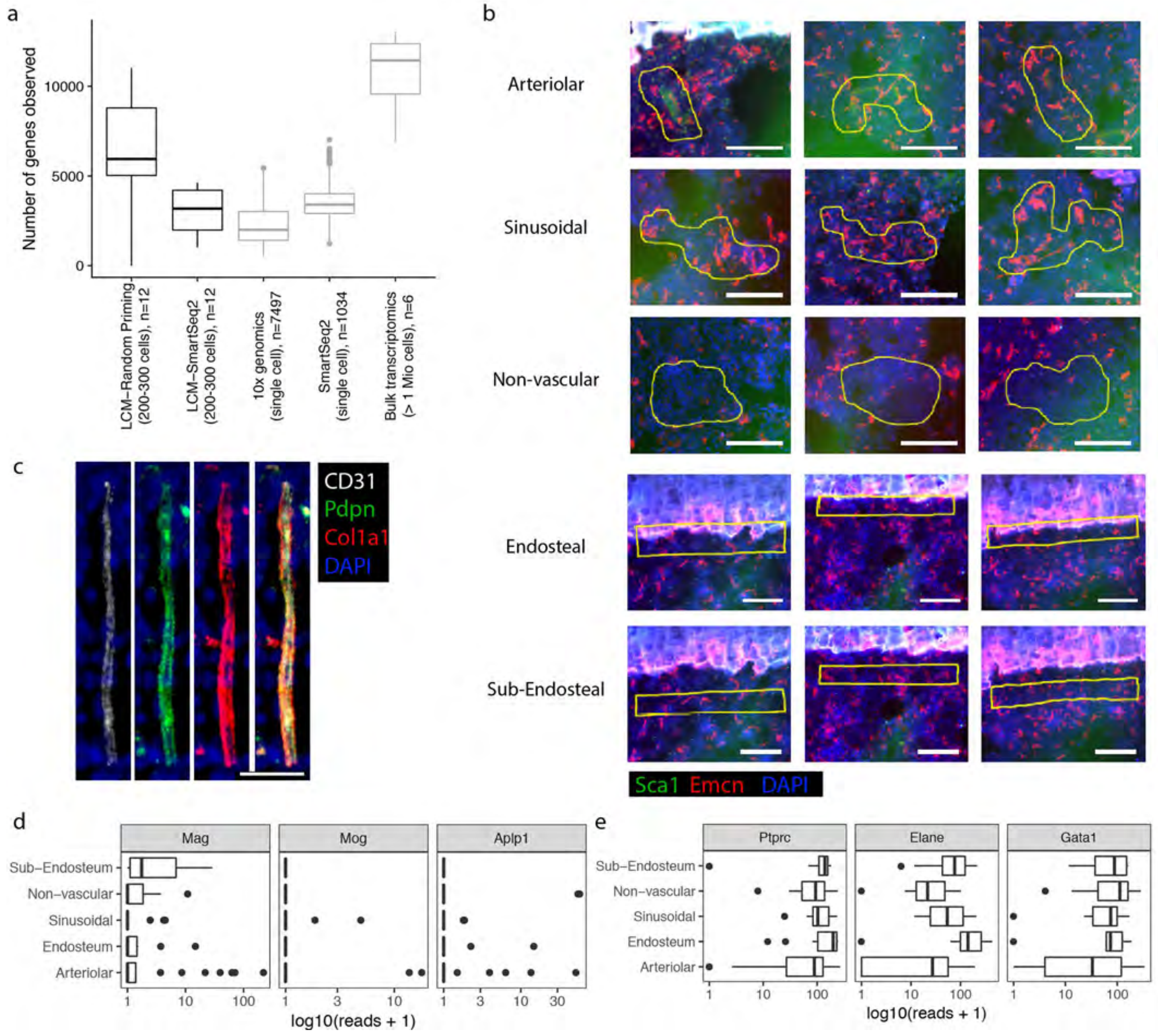


Extended Data Fig. 4 | Comparison of cell-type transcriptomes determined by scRNAseq to data from bulk populations described in literature. **a**, Enrichment of gene expression signatures of haematopoietic populations in immune cell transcriptomes published by the ImmGen consortium (data source: GEO GSE109125)[58]. Data points in the same row correspond to experimental replicates. **b**, Enrichment of gene expression signatures of non-haematopoietic populations in published transcriptomes of populations defined by genetic markers[5,7–9,59]; see methods for specification of data sources, and see the Supplementary Note 1 for a detailed evaluation of the algorithm used. **c**, The optimal number of mesenchymal cell clusters was determined using the SOUP method[15], a semi-soft clustering algorithm designed to distinguish between distinct cell types and transition states between cell types. Error bars indicate 90% confidence intervals across 10 independent SOUP runs, central values correspond to medians. **d**, Main cluster identity from SOUP highlighted on the t-SNE from Fig. 1b (mesenchymal cell types only, n=2223 cells). **e**, Comparison of clusters identified by Seurat (Fig. 1b) to clusters identified by SOUP (panel d) demonstrates strong overlap between both methods.

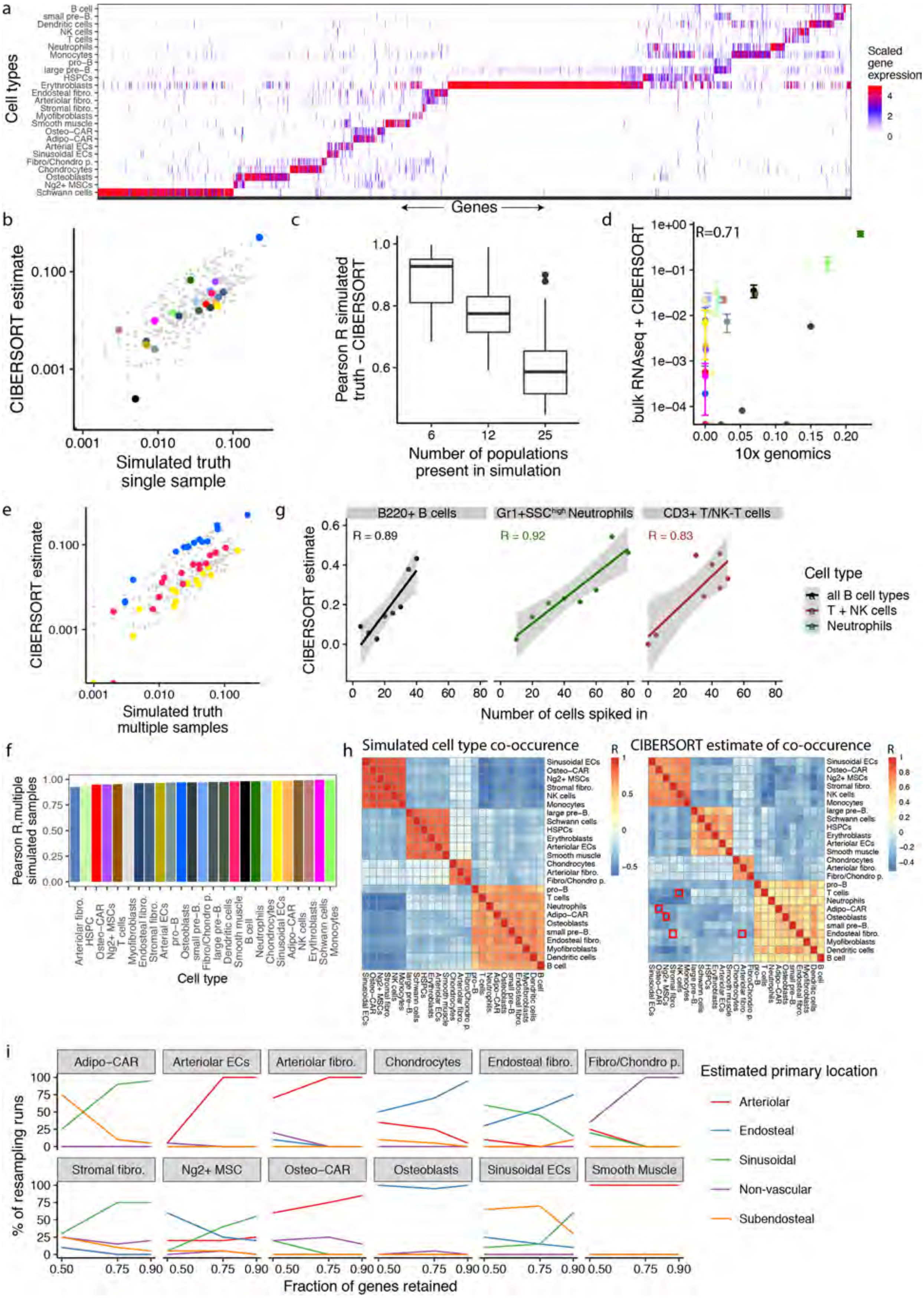


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Comparison of cell isolation methods and reference datasets. a,**b**, Additional scRNAseq data was generated as described, except that BM was derived by flushing bones and subjected or not subjected to enzymatic digestion. Data was projected to two dimensions using t-SNE and cell type labels were assigned using the anchoring approach implemented in seurat3[53]. N=2411 cells are shown. **c**,**d**, Single-cell RNA-seq data from a recent study of different genetically labelled populations from flushed BM[26] was projected to two dimensions using t-SNE and cell type labels were assigned using the anchoring approach implemented in seurat3[53]. N=6617 cells are shown. **e**, Comparison of cell type frequencies between various published datasets. For Tikhonova et al.[26], n=6617 cells, Baryawno et al.[25], n=25933 cells, Baccin et al., n=7497 cells.

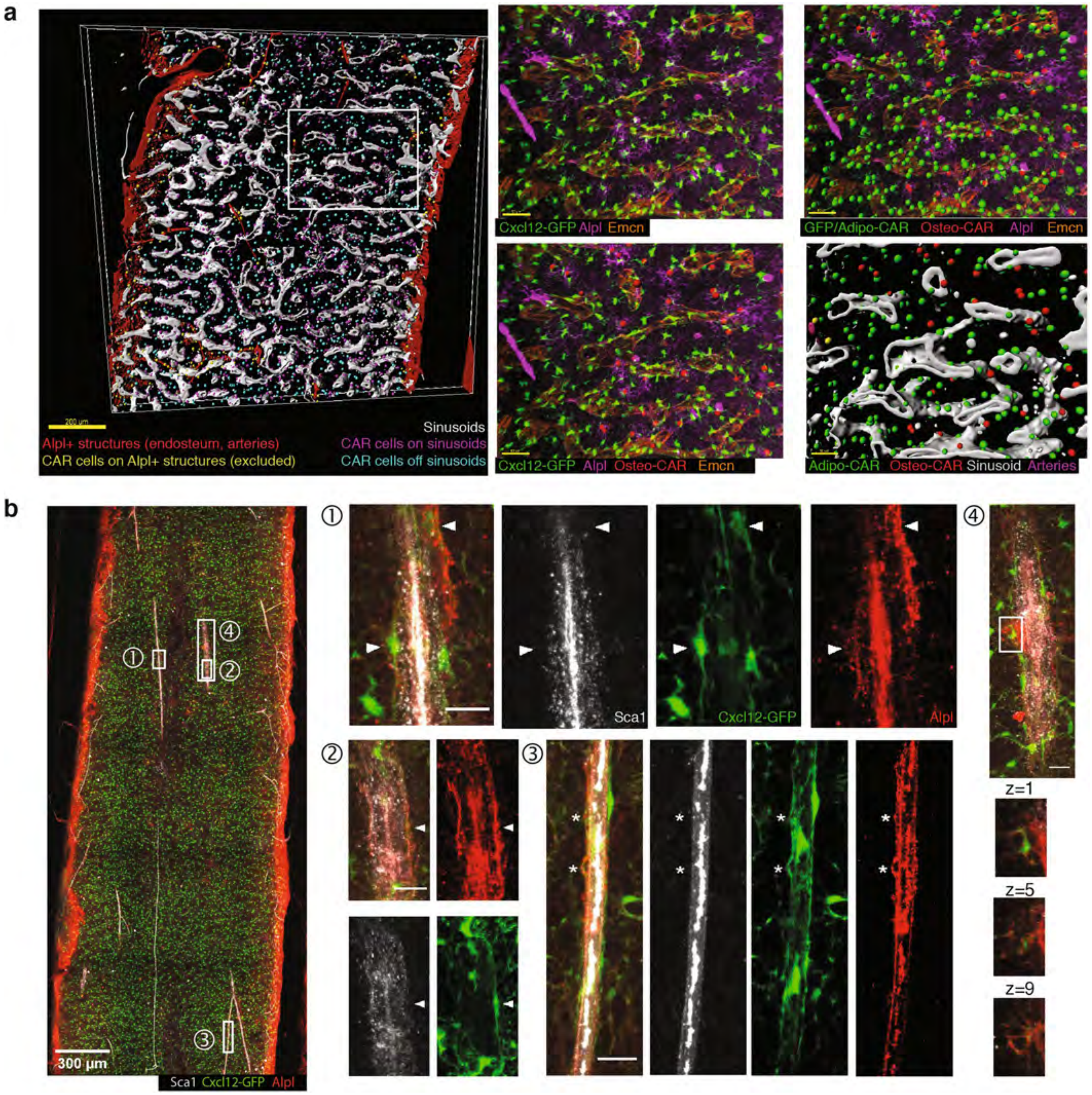


Extended Data Fig. 6 | Technical properties of the LCM-seq dataset. a, Boxplots comparing the number of genes observed per sample in different protocols. All samples were down-sampled to 1 million reads for comparison. For the dataset presented in main Fig. 3, the protocol relying on random priming was used. N represent numbers of samples or cells as indicated on the graph. Boxplot elements are defined in the Methods section (section on data visualization). Smart-Seq2 data is from ref. 21, bulk transcriptome data is from ref. 7. **b**, Representative images of samples collected for LCM-seq; scale bar: 100µm. **c**, Immunofluorescence staining of a BM arteriole stained for Col1a1, Pdpn and CD31. Scale bar: 20 µm. The experiment was repeated independently 4 times with similar results. **d**, Schwann cell markers were lowly expressed across all niches **e**, haematopoietic markers were highly expressed across all niches. In panel d and e, sample sizes are as follows: Arteriolar niches, n=28, endosteal niches, n=12, sinusoidal niches, n=14, non-vascular niches, n=11, sub-endosteal niche, n=11. Boxplot elements are defined in the Methods section (section on data visualization).

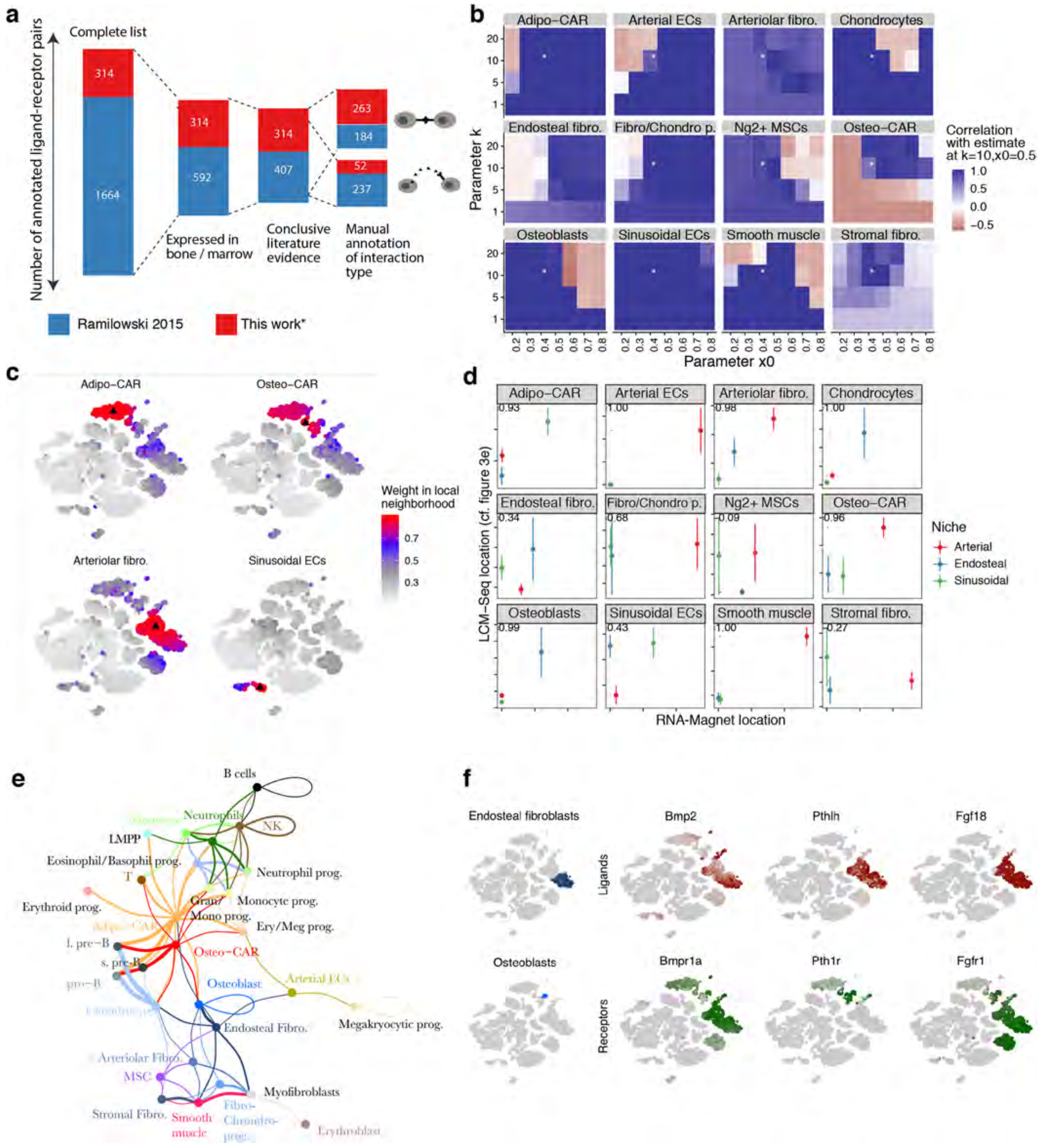


Extended Data Fig. 7 | See next page for caption.

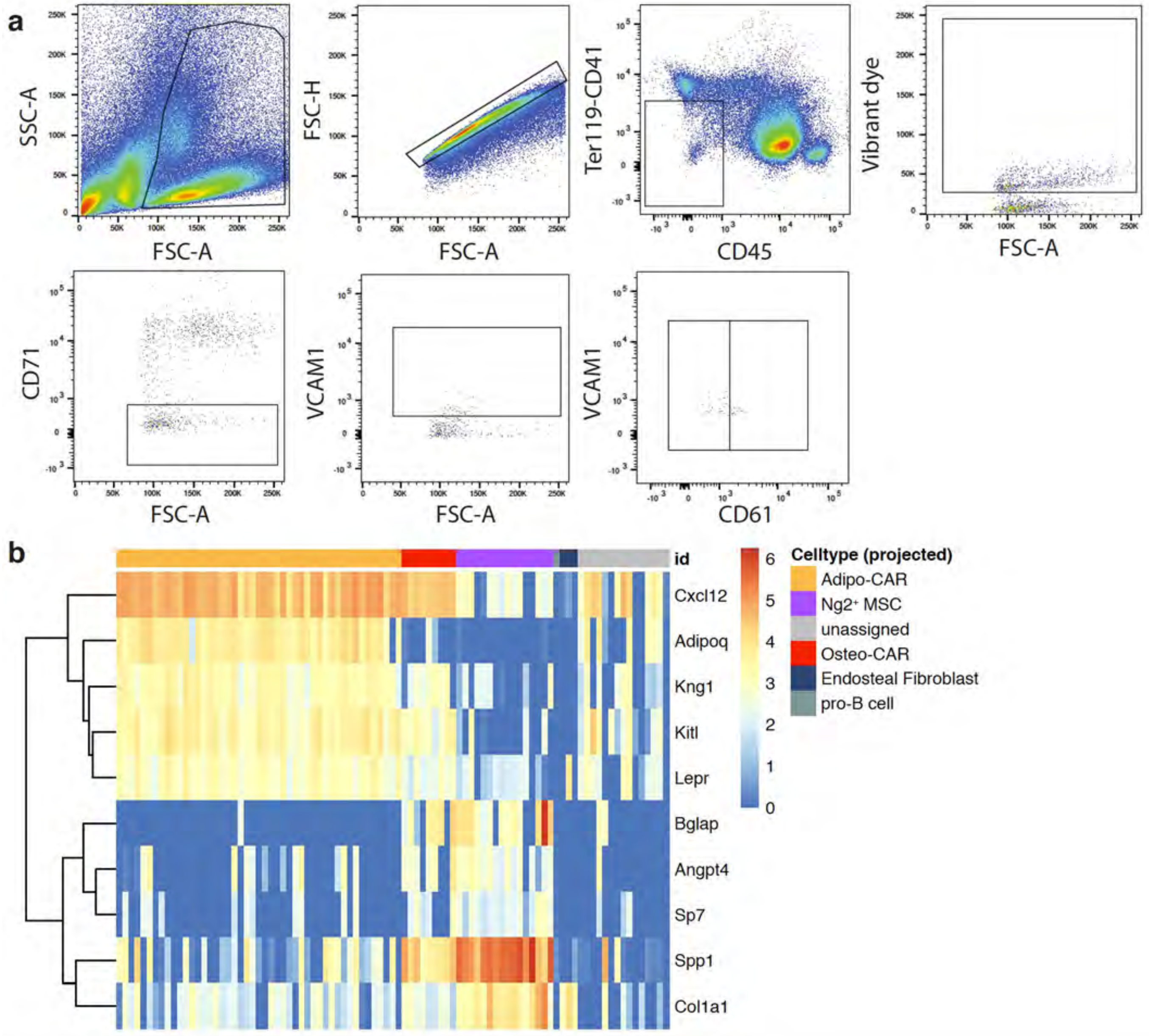
Extended Data Fig. 7 | Evaluation of the CIBERSORT algorithm, see also Supplementary Note 1. a, Heatmap of population-specific marker genes used for the algorithm. **b**,**c**, Simulations to assess the ability of CIBERSORT to decompose individual samples; see Supplementary Note 1 for detail. n=15 sampling runs were performed. Boxplot elements are defined in the Methods section (section on data visualization). **d**, CIBERSORT estimates of cell type composition of total BM, compared to the cell type composition estimate from 10x genomics (see Extended Data Fig. 1b). Central values indicate means, error bars correspond to standard error of the mean of n=3 samples. **e**,**f**, Simulations to assess the ability of CIBERSORT to identify changes in population frequencies across multiple samples; see Supplementary Note 1 for detail. n=15 samples runs were performed. **g**, FACS was used to assemble N=8 different pools of B220$^+$ B-cells, CD3$^+$ T/NK-T cells and Gr1$^+$SSC^{high} neutrophils. Each pool contained a total of 100 cells at predefined ratios of B cells, T cells and neutrophils. Pools were then fixed and processed using the LCM-seq protocol, and CIBERSORT was used to decompose their composition. Estimates for T and NK cells, as well as different B-cell subpopulations, were summed for the display. Pearson correlation coefficients are shown. **h**, Simulations to assess the ability of CIBERSORT to discriminate between similar cell types; see Supplementary Note 1 for detail. Red squares highlight pairs of similar cell types. Pearson correlation coefficients across N=100 simulated samples are shown. **i**, Stability of the CIBERSORT estimates from main Fig. 3e with regard to re-sampling of the marker gene lists used; see Supplementary Note 1 for detail.



Extended Data Fig. 8 | Whole-mount imaging and data analysis. **a**, Whole-mount imaging data of a Cxcl12–GFP bone section stained for Alpl and Emcn was segmented in 3D using the imaris software. Large Alpl$^+$ surfaces (red, corresponding to endosteum and arteries) were identified and any GFP+ spots with <20µm proximity to these structures were excluded from further analysis (yellow spots). The strong Alpl signal in these regions prevents a quantitative assessment. Remaining GFP$^+$ spots were classified as within 15µm of sinusoidal vessels (purple dots), of away from sinusoidal vessels (cyan dots). GFP$^+$ spots were further classified as Alpl$^+$ (right panels, red spots) or Alpl$^-$ (right panels, green spots). Scale bar left panel: 200µm; right panels: 50µm. **b**, Additional whole-mount imaging data of a femur from a Cxcl12–GFP mouse, stained for Alpl and the arteriolar marker Sca1. Arrowheads point to Alpl$^+$Cxcl12$^+$ cells near, but not overlapping with, Sca1$^+$ arteriolar endothelial cells. See also Fig. 4f. In ROI 3, asterisk correspond to GFP$^+$Alpl$^+$ protrusions on, but clearly distinct from, Sca1$^+$ arteriolar endothelial cells. In ROI 4, various z-sections of a highly reticulate Cxcl12–GFP$^+$Alpl$^+$ cell are shown. Scale bars in ROIs: 20µm; scale bar in main panel: 300µm.



Extended Data Fig. 9 | Analyses using RNA-Magnet, see also Supplementary Note 2. a, Overview of the receptor–ligand list used. See methods for data sources, and Supplementary Table 3 for the complete list. **b**, Stability of the RNA-Magnet location estimate for different choices of the fuzzification parameters k and x_0. For each parameter set, RNA-Magnet location estimates were summarized per cell type, and compared to the summarized location estimate displayed in Fig. 6c. The asterisk indicates the parameter set used in Fig. 6c. **c**, Choice of local neighbourhoods. As detailed in the methods section, RNA-Magnet works by identifying interactions specific to a single cell compared to similar cells. The figure displays the size of local neighbourhoods for four representative cells demarked by a black triangle. **d**, Detailed comparison of location estimates obtained from LCM-seq and RNA-Magnet. See also main Fig. 6c. Sample sizes are as follows: Arteriolar niches, n=28, endosteal niches, n=12, sinusoidal niches. Mean values and standard error of the mean are shown. **e**, Fully labelled display of the network from main Fig. 8a. **f**, Expression of selected cytokines and growth factors involved in bone remodelling.



Extended Data Fig. 10 | Index-sorting analysis of Lin⁻Vcam1⁺ cells. a, Sorting scheme used. **b**, Expression of key marker genes confirm the cell type assignment obtained by scmap, cf. main Fig. 7c.