Myocardial infarction is a leading cause of death worldwide. Although advances have been made in acute treatment, an incomplete understanding of remodelling processes has limited the effectiveness of therapies to reduce late-stage mortality. Here we generate an integrative high-resolution map of human cardiac remodelling after myocardial infarction using single-cell gene expression, chromatin accessibility and spatial transcriptomic profiling of multiple physiological zones at distinct time points in myocardium from patients with myocardial infarction and controls. Multi-modal data integration enabled us to evaluate cardiac cell-type compositions at increased resolution, yielding insights into changes of the cardiac transcriptome and epigenome through the identification of distinct tissue structures of injury, repair and remodelling. We identified and validated disease-specific cardiac cell states of major cell types and analysed them in their spatial context, evaluating their dependency on other cell types. Our data elucidate the molecular principles of human myocardial tissue organization, recapitulating a gradual cardiomyocyte and myeloid continuum following ischaemic injury. In sum, our study provides an integrative molecular map of human myocardial infarction, represents an essential reference for the field and paves the way for advanced mechanistic and therapeutic studies of cardiac disease.
Here we used a combination of single-cell gene expression, chromatin accessibility and spatially resolved transcriptomics to study the events of cardiac tissue reorganization and to characterize the cell-type-specific changes in gene regulation, providing an integrated spatial multi-omic map of cardiac remodelling after myocardial infarction. Our multi-omic data-driven approach, including spatial context, enables us to understand how a given cell state changes based on the cells’ neighbourhood and how this relates to transcriptional and regulatory variations. By deconvoluting spatial transcriptomics spots into cell-type abundances, we characterized cell niches occurring in different stages following acute myocardial infarction. We identified different cell states of cardiomyocytes, endothelial cells, myeloid cells and fibroblasts that are associated with disease progression on the basis of the integrated single-cell multi-omics data. Moreover, we inferred the gene-regulatory networks differentiating these cell states and projected this information onto specific tissue locations, thus mapping putative regulators controlling gene expression on specific myocardial tissue zones and disease stages. This enabled us to gain novel insights into the gene-regulatory programmes driving injury of cardiomyocytes, activated phagocytic macrophages and their relation to myofibroblast differentiation in cardiac tissue remodelling. Our results provide a comprehensive spatially resolved characterization of gene regulation of the human heart in homeostasis and after myocardial infarction. We have released our spatial multi-omics data through publicly available platforms to enable users to interactively explore the dataset. We anticipate that this data will be a reference map for future studies and ultimately for the development of novel therapeutics.

**Multi-omic map of myocardial infarction**

We applied an integrative single-cell genomics strategy with single-nucleus RNA sequencing (snRNA-seq) and single-nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) together with spatial transcriptomics from the same tissue mapping human cardiac cells in homeostasis and after myocardial infarction at unprecedented spatial and molecular resolution (Fig. 1a–c and Supplementary Table 1). We profiled a total of 31 samples from 23 individuals, including four non-transplanted donor hearts as controls, and samples from tissues with necrotic areas (ischaemic zone and border zone) and the unaffected left ventricular myocardium (remote zone) of patients with acute myocardial infarction (Fig. 1a). These acute myocardial infarction specimens were collected from heart tissues obtained at different time points after the onset of clinical symptoms (chest pain), before the patients received an artificial heart or a left-ventricular assist device because of cardiogenic shock and as a bridge to transplantation (Supplementary Fig. 1a–c). We also analysed nine human heart specimens at later stages after myocardial infarction (fibrotic zone; Fig. 1b) that exhibited ischaemic heart disease and were available from heart transplantation recipients at the time of orthotopic heart transplantation.

For each cardiac sample, we obtained 10-μm cryo-sections and isolated nuclei from the remaining tissue directly adjacent to the cryo-section with subsequent fluorescence-activated nuclei sorting (FANS) for snRNA-seq and snATAC-seq (Fig. 1c). After filtering out low-quality nuclei, we obtained a total of 191,795 nuclei from all samples for snRNA-seq, with an average of 2,020 genes per nucleus, together with chromatin accessibility data from 46,086 nuclei overall with an average of 28,066 fragments per nucleus (Supplementary Fig. 2a,b and Supplementary Tables 2–5). After controlling for data quality, the spatial transcriptomics datasets contained a total of 91,517 spots (average of 3,389 spots per specimen and 2,001 genes per spot) (Supplementary Figs. 2c,e–g and 3a,b). Quantification based on histology revealed an average of four nuclei per spatial transcriptomic spot from all slides (Supplementary Fig. 2c and Supplementary Table 6). Samples from the ischaemic zone had the lowest abundance of nuclei and an enriched expression of genes associated with cell death and the regulated necrosis pathway, suggesting increased necrotic cell death (Supplementary Fig. 2d). This integrated dataset represents, to our knowledge, the largest and most comprehensive multi-modal profiling of human myocardial infarction tissue including spatial information and samples at distinct disease progression stages. We devised an integrative data analysis approach spanning all three modalities of our single-cell experiments to study cardiac cell-specific information and cell-specific interactions in their spatial and disease progression context (Extended Data Fig. 1a).

We established a map of major human heart cell types using the snRNA-seq and snATAC-seq datasets independently. First, we clustered cells on the basis of the integrated snRNA-seq data from all samples after batch correction (Extended Data Fig. 1b). Clusters were annotated with curated marker genes from the literature4–6 and ten major cardiac cell types were identified (Fig. 1d,e). We also found an additional cluster with enriched expression of the cell-cycle marker gene MKI67, which showed a high score of cell-cycle G2/M and S phases and was mainly recovered in ischaemic zone samples (Extended Data Fig. 1c,d). To validate the annotations, we compared the data with a recent study on healthy human hearts4 and an independent novel dataset of ischaemic heart samples (n = 3, generated during this study) and observed a high agreement and correlation in terms of molecular profiles and cellular composition (Extended Data Fig. 1e–g). Of note, the cycling cells were also captured in the independent ischaemic dataset (Extended Data Fig. 1f).

We next integrated and clustered the snATAC-seq data from all samples (Extended Data Fig. 2a). These clusters were annotated on the basis of gene chromatin accessibility with the same markers as for snRNA-seq. This approach identified eight major cell types, matching all cell types from snRNA-seq data with the exception of two rare cell types (that is, mast cells and adipocytes) (Fig. 1f,g). Label transfer from snRNA-seq to snATAC-seq indicated that the annotations between these two modalities were consistent (Extended Data Fig. 2b,c). This was further supported by a high correlation of cellular composition between snRNA-seq and snATAC-seq and the presence of the same eight cell types in the majority of samples (Extended Data Fig. 2d,e). To explore regulatory information provided by the snATAC-seq, we performed transcription factor footprinting analysis using cell-type-specific pseudo-bulk ATAC-seq profiles. This revealed footprinting-based binding activity of known transcription factors such as MEF2C (ref. 7) in cardiomyocytes, CEBPD8 in myeloid cells, FOS–JUNB9 in fibroblasts and SRF10 in vascular smooth muscle cells (vSMCs), which correlated with the expression of their predicted target genes in snRNA-seq data (Extended Data Fig. 2f).

Together, our integrative analysis of transcriptomic and chromatin accessibility data defined a robust catalogue of cell types in the adult human heart across multiple modalities and samples.

**Molecular mapping of cell types in space**

Using these data, we first identified overrepresented biological processes for each major histomorphological region (control, remote zone, border zone, ischaemic zone and fibrotic zone) using spatially variable genes (Supplementary Table 7). We identified cardiac muscle contraction in remote zones and controls, with adaptive immune system in the border and ischaemic zones and with matrisome processes in the fibrotic zones (Extended Data Fig. 2g). Overall, this analysis confirmed that the spatial data clearly reflect typical zones of biological processes following acute human myocardial infarction.

Since each spatial transcriptomics spot captured a group of cells, we increased its resolution by estimating the cell-type compositions of each spot. To this end, we deconvoluted each spot on the basis of the annotated snRNA-seq data from the same sample (Fig. 1h, Supplementary Figs. 2e–g and 3a,b, Supplementary Tables 8 and 9 and Methods). The estimated cell-type compositions from spatial transcriptomics of
each patient generally agreed with their respective observed compositions in the snRNA-seq and snATAC-seq data (Extended Data Fig. 2h). We then estimated signalling pathway activities with PROGENy (Methods) for each spot from the spatial gene expression data. The comparison of spatially localized pathway activities with the estimated cellular abundance per spot enabled us to link the information on spatial cell composition to cellular function for each slide. For example, in areas with an abundance of fibroblasts, we detected increased TGFβ signalling activity, and in ischaemic regions, increased myeloid cell abundance occurred in areas of higher NFκB signalling activity (Fig. 1h, i).

Mapping the information obtained from the snATAC-seq data to space resulted in spatially resolved footprinting-based transcription factor binding activity, as exemplified by the previously described transcription factors associated with cardiomyocytes (for example, MEF2C; ref. 7), myeloid cells (for example, CEBPD and ATF111), fibroblasts (for example, FOS–JUNB9) and vSMCs (for example, SRF10) (Fig. 1j). To test...
Spatial organization of myocardial tissue

To explore the spatial organization of the myocardial tissue, we leveraged the spatial transcriptomics data. Unsupervised clustering of spots from all samples on the basis of their cell-type compositions identified nine clusters, which we defined as major cell-type niches (Fig. 2a and Extended Data Fig. 3a–d). We hypothesized that these niches represent potential structural building blocks that are shared between different slides and could facilitate comparisons between subjects. Visualization of these niches in space revealed that some niches aligned closely with the underlying sample condition; for example, cell-type niche 6 was equally distributed across a control slide, whereas cell-type niche 5 localized to distinct regions on the ischaemic slide (Fig. 2b). We then tested the overrepresentation of the annotated cell types derived from snRNA-seq in the cell-type niches. We observed 4 myogenic cell-type niches (1, 7, 8 and 9), which were enriched with cardiomyocytes, endothelial cells, and pericytes (Fig. 2c); an inflammatory cell-type niche (niche 5); and a fibrotic cell-type niche (niche 4) containing fibroblasts, myeloid and lymphoid cells. The fibrotic cell-type niche (4) contained a higher proportion of fibroblasts, whereas the inflammatory cell-type niche (5) contained more myeloid and lymphoid cells (Fig. 2c).

Finally, we observed niches associated with rare cell types of the myocardium, such as vSMCs (cell-type niches 3 and 6), adipocytes, lymphoid and cycling cells (cell-type niche 2) (Fig. 2c and Extended Data Fig. 3d).

Our integrated results provide a comprehensive description of cellular colocalization events, enabling downstream molecular comparisons within this atlas across all tissue zonations. We next tested whether the abundances of major cell types within spots could be predicted by their spatial context described by the cell-type compositions of their neighbourhood. We evaluated three different neighbourhood area sizes (colocalization) (Fig. 2d), in the local neighbourhood (radius of 1 spot), and in an extended neighbourhood that expanded to a radius of 15 spots. We observed that endothelial cells were the most predictive of the abundance of vSMCs, pericytes, adipocytes and cardiomyocytes within all spots, probably reflecting dependencies between cell types of the vasculature (Fig. 2d). Lymphoid and myeloid cells showed strong dependencies with each other in line with zones of immune cell infiltration and inflammation—similarly captured by cell-type niche 5 (Fig. 2d). Notably, we observed strong dependencies between myeloid cells and fibroblasts, which were strongly co-enriched in niche 4 (Fig. 2d and Extended Data Fig. 3e), in line with a known key role of macrophages in fibroblast activation and fibroblasts in macrophage attraction.

Between immediate and extended neighbouring spots (Extended Data Fig. 3f–h), we observed stronger dependencies between cells associated with the cardiac vasculature (vSMCs, endothelial cells, pericytes and fibroblasts) indicating that the myocardial vascular network dominates cardiac tissue structural organization.
To link tissue organization to function, we analysed spatial dependencies between signalling pathways and cell types. Modelled importance of colocalized pathways captured relationships between PI3K and P53 signalling (Extended Data Fig. 4a–e), which showed a mutually exclusive spatial distribution (Extended Data Fig. 4c). Both pathways were related to the abundance of cardiomyocytes (Extended Data Fig. 4a). PI3K signalling in cardiomyocytes controls the hypertrophic response to preserve cardiac function, whereas P53 is known to act as a master regulator in cardiac homeostasis. Spatial segregation of these cardiomyocyte-related pathways points towards functional cardiomyocyte heterogeneity. We observed colocalized and extended neighbour-

hoo observations of known key pathways in fibrosis including TGFβ and Nkx2B predicted by fibroblasts, and JAK–STAT and NFκB predicted by immune cells (Extended Data Fig. 4a–e). Overall, cardiomyocytes were the best predictor cell types of the activities of the estimated pathways. Hypoxia and WNT pathways showed a colocalization to card-

myocytes in ischaemic specimens (Extended Data Fig. 4b–e), highly consistent with the cardiomyocyte differentiation events occurring after myocardial infarction. Our results compiled principles of tissue organization of the human heart that relate to coordinated cellular processes and provide a basis for comparative analysis.

Molecular variation following infarction

To study the molecular differences between similar tissue structures in an unbiased manner across samples, we generated a set of molecular niches by clustering of spots on the basis of their gene expression (Fig. 3a,b and Extended Data Fig. 5a–d). We identified molecular niches associated with inflammatory and fibrotic processes (molecular niches 3, 6 and 9), vSMCs (molecular niche 11) and myogenic-enriched regions (molecular niches 1, 2, 4, 5 and 12) (Fig. 3c). The molecular niches enriched in cardiomyocytes were depleted in ischaemic-enriched samples, whereas the fibro- and inflammatory-enriched molecular niches were depleted in myogenic-enriched samples (Fig. 3d and Extended Data Fig. 5e). The VSMC-enriched molecular niche 11 had a more distinct cell-type marker gene expression of vSMCs (MYH11) compared with the cell-type defined niche 6 (Fig. 3b versus Extended Data Fig. 3d).

Of note, we observed molecular niches that enabled us to differentiate border zone, remote zone and control samples (Extended Data Fig. 5g), which were indistinguishable using the major cell-type niches (Extended Data Fig. 4m). Molecular niche 3, enriched in fibroblasts and immune cells, was more present in remote zones and border zones compared with control samples. Moreover, we observed differences in the proportions of the molecular niches 1, 2 and 4 among border zone, remote zone and controls (Extended Data Fig. 5g). These three molecular niches were enriched mainly in cardiomyocytes (Fig. 3c), but with a distinct molecular profile: among the top 5 upregulated genes of niche 2 was AIRI, which encodes an intercalated disc ion-channel-interacting protein and RRAD, which encodes a GTPase known to regulate L-type Ca2+ channels and contractile functions of the heart, which are depleted in myogenic-enriched samples (Extended Data Fig. 5h). Overall, molecular niche 1 was enriched in control and remote zone samples and niche 2 was enriched in the damaged tissue areas in border zone samples (Fig. 3e,f and Extended Data Fig. 5g). We observed slight changes in enrichment of molecular niches 2 and 4, and a depletion of niche 1 in border zones compared with controls (Extended Data Fig. 5g, i), suggesting that differences in cardiomyocyte phenotypes might also be present between these groups. In summary, the comparison of molecular niches pointed towards subtle changes between the remote myocardium and controls.
and expected differences between border zone and both controls and remote zone that were not detectable in the cell-type niche comparison. Overall, this suggested the existence of functional differences between cardiomyocyte states in our data.

**Disease-specific cardiomyocyte states**

To further investigate distinct cardiomyocyte states, we aimed to understand the molecular heterogeneity of cardiomyocytes after myocardial infarction. We co-embedded the snRNA-seq and snATAC-seq data from cardiomyocytes into a common low-dimensional space and clustered the cells (Extended Data Fig. 6a). This uncovered five cell states of ventricular cardiomyocytes (vCMi−5), spanning multiple samples and modalities (Fig. 4a and Supplementary Table 10). Differential gene expression analysis revealed a significant upregulation of **ANKRD1** in both vCM2 and vCM3, whereas **NPPB** showed a distinct upregulation and increased chromatin accessibility in vCM3 (Fig. 4b and Extended Data Fig. 6b,c). We validated this upregulation by single-molecule fluorescence in situ hybridization (smFISH) in an independent patient cohort (Fig. 4c and Extended Data Fig. 6d). Both **NPPB** and **ANKRD1** have been reported to be upregulated in the border zone after myocardial infarction in mice.22 vCM2 additionally showed enhanced expression of **MYH7** (Extended Data Fig. 6b), a cardiomyocyte-associated stress gene that encodes the β-myosin heavy chain.22 Thus, we annotated the vCM2-state as ‘pre-stressed’. In addition, we observed a higher correlation between ion-channel-related genes and vCM1 marker genes compared with ‘stressed’ vCM marker genes in spatial transcriptomics, which further highlights the functional differences between these two cardiomyocyte states (Extended Data Fig. 6e). Accordingly, we annotated the vCM3 state as stressed. Moreover, when comparing the differential expression of individual genes belonging to these ion-channel-related gene sets in snRNA-seq data, we observed mostly upregulations in vCM1 compared with vCM3 (Extended Data Fig. 6f,g). Cellular composition comparison between sample groups revealed that vCM1 was associated with myogenic-enriched samples and vCM3 was significantly associated with ischaemic-enriched samples. This was validated in an independent cohort using in situ hybridization, suggesting that these cardiomyocyte states represent distinct cellular stress states within the acute myocardial infarction phase (that is, vCM1, ‘non-stressed’; vCM2, ‘pre-stressed’, and vCM3, ‘stressed’) (Fig. 4c,d and Extended Data Fig. 6h,i).

Next, we checked vCM marker genes in spatial transcriptomics in border zone samples, since spatial remodelling of this area is inextricably linked to the recovery of cardiac function. Interestingly, despite homogenous H&E staining and unique molecular identifier (UMI) distribution across spots (Supplementary Fig. 2g), we observed extensively heterogeneous spatial gene expression patterns of **ANKRD1** and **NPPB** (Fig. 4e). Pathway analysis of the spatial gene expression data indicated an increased TGFB signalling activity within the injured area (lower right), but a homogeneous distribution of hypoxia pathway activity (Fig. 4f). Mapping of cell states to space in a border zone sample revealed that vCM1 were solely located in the top left uninjured corner, vCM2 were located in the middle-top area, serving as a transition zone from injured towards remote myocardium, and vCM3 were primarily located below the transition zone within the injured area (Fig. 4g). Of note, such a spatially distributed pattern was also observed in another border zone sample, indicative of a similar remodelling process (Extended Data Fig. 6j).

**Variability of cardiomyocyte states**

To infer an enhancer-based gene-regulatory network (eGRN), we leveraged our multi-omics data to further investigate molecular mechanisms differentiating the relevant cardiomyocyte states (that is, vCM1–vCM3) (Methods and Supplementary Table 11). To this end, we paired the cells between snATAC-seq and snRNA-seq data and studied gene-regulatory changes along the cellular continuum from vCM1 to vCM3 (Extended Data Fig. 7a). Next, we estimated an enhancer-mediated transcription factor–target network by considering transcription factor activity (from snATAC-seq), expression of transcription factor and target genes (from snRNA-seq), and motif-supported peak-to-gene links (Extended Data Fig. 7b–d). Clustering of these transcription factors to the target network revealed three major modules, with each corresponding to a distinct cardiomyocyte state (Extended Data Fig. 7e).

We next used network analysis to visualize and detect major transcription factors (Fig. 4h). We identified the mineralocorticoid receptor (NR3C2), a major target of therapy for common heart failure, as a major regulator of the vCM1 state (Fig. 4h). Decreased **NR3C2** expression has been associated with the development of severe heart failure and cardiac fibrosis,14 and we observed decreased transcription factor binding activity and gene expression along the pseudotime of vCM1...
to vCM3 differentiation (Fig 4i). Target genes of NR3C2 include several ion channel genes (such as $\text{SLC}8A1$), which also showed decreased gene expression along the pseudotime axis (Fig 4i). Notably, these target genes were also differentially expressed in cardiomyocyte-enriched molecular niches (Fig 3e,f and Extended Data Fig. 5h) and aligned spatially in the border zone with the vCM1 state (Fig 4j). Notably, we also

**Fig. 4** | Sub-clustering of cardiomyocytes. a. Sub-clustering of cardiomyocytes. b. Gene expression of $\text{ANKRD1}$ and $\text{NPPB}$. c. Left, smFISH staining of vCM3 marker genes. Scale bar, 50 µm. Right, quantification of $\text{NPPB}$ and $\text{ANKRD1}$ gene expression over pseudotime. Norm., normalized.

d. Expression of transcription factor target genes in the border zone sample, as in e, k. l, Mean importance of the abundance of major cell types within a spot (k) and the local neighbourhood (within a 5-spot radius) (l) in the prediction of vCM3 in spatial transcriptomics. m, Importance of vSMC abundance predict vCM3 in myogenic, ischaemic and fibrotic groups within a spot (adj. $P$-value, two-sided Wilcoxon rank-sum test; myogenic: $n = 9$, ischaemic: $n = 7$, fibrotic: $n = 4$).

e, Expression of $\text{ANKRD1}$ and $\text{NPPB}$ (e), TGFβ and hypoxia-signalling activities (f) and expression of vCM-state marker genes (g) in a border zone sample. h, eGRN analysis including vCM1, vCM2 and vCM3. Each node represents a transcription factor (regulator) or a gene (target). i, Transcription factor activity and expression over pseudotime. Norm., normalized. j, Expression of transcription factor target genes in the border zone sample, as in e, k. l, Mean importance of the abundance of major cell types within a spot (k) and the local neighbourhood (within a 5-spot radius) (l) in the prediction of vCM3 in spatial transcriptomics. m, Importance of vSMC abundance predict vCM3 in myogenic, ischaemic and fibrotic groups within a spot (adj. $P$-value, two-sided Wilcoxon rank-sum test; myogenic: $n = 9$, ischaemic: $n = 7$, fibrotic: $n = 4$). n, Deconvoluted abundance of vSMCs or fibroblasts and vCM3 state scores in a border zone (left) and a control human heart (right). o, Importance of myeloid cell abundance in the local neighbourhood for predicting vCM3 in control and remote zone samples (two-sided $t$ test; controls: $n = 3$, remote zones: $n = 3$). For details on visualization, statistics and reproducibility, see Methods.
observed transcription factors (TBX3 and MEF2D) that were associated with pre-stressed stages of cardiomyocyte differentiation (Fig. 4h). Our analysis suggests that MEF2D, a cardiomyocyte factor controlling pacemaker function\textsuperscript{21}, regulates the expression of the sarcomere protein MYBP\textsubscript{C3} (Fig. 4i). MYBP\textsubscript{C3}, in turn, has been reported to regulate cardiomyocyte proliferation postnatally\textsuperscript{2b}. Of note, we identified MYBP\textsubscript{C3} independently in our spatial data as being enriched in molecular niche 1 (Fig. 3e and Extended Data Fig. 3h).

We also identified ANKR\textsubscript{D1}, a mediator of cardiomyocyte response to stress\textsuperscript{22}, as a target of MEF2D, suggesting a key regulatory role of MEF2D in the transition from vCM\textsubscript{1} to vCM\textsubscript{3} (Extended Data Fig. 7f). For vCM\textsubscript{3} (stressed cardiomyocytes), we identified ATP3 as a regulator of the GTPase and Ca\textsuperscript{2+} regulator gene RR\textsubscript{AD} (Fig. 4h). We independently identified RR\textsubscript{AD} in molecular niche 2 (Extended Data Fig. 3h), which supports its relevance as a spatially differentially expressed gene of a distinct cardiomyocyte state, especially in border zone samples (Fig. 4i). We additionally identified the transcriptional regulator JDP2—which has a function in preventing cardiomyocyte hypertrophy and cell death\textsuperscript{23}—as an important regulator of the vCM\textsubscript{3} cardiomyocyte state, with TGFB\textsubscript{2} as one of its target genes (Extended Data Fig. 7g.h).

In summary, our cardiomyocyte states and major transcription factor regulators identified from the integrated snRNA-seq and snATAC-seq data reflect expression patterns associated with molecular niches supporting spatial changes of cardiomyocyte states during remodelling.

We next estimated the cell dependencies of the stressed cardiomyocyte state vCM\textsubscript{3} with other cell types within each spatial spot and its local neighbourhood (radius of five spots) between sample groups (Fig. 4k–o). We observed that there was an increased importance of vSMCs in predicting vCM\textsubscript{3} within a spot in myogenic and ischaemic samples (Fig. 4k), whereas fibroblasts and myeloid cells had a larger role in fibrotic samples (Fig. 4k). The local neighbourhood modelling of vCM\textsubscript{3} revealed that the abundance of fibroblasts better explained vCM\textsubscript{3} in myogenic-enriched samples compared with fibrotic samples (Fig. 4l and Extended Data Fig. 7i). To gain further insight, we visualized the dependencies of vSMCs and fibroblasts on vCM\textsubscript{3} in myogenic-enriched samples and observed that their colocalization occurred in the perivascular niches (Fig. 4n). Overall, this demonstrates that the stressed cardiomyocyte state vCM\textsubscript{3} occurs in the perivascular niche of larger blood vessels, highlighting the interaction of mesenchymal cells\textsuperscript{25} of the perivascular niche with stressed cardiomyocytes in this tissue area. Furthermore, we noticed that when comparing remote zone with control samples, stressed vCM\textsubscript{3}s are best predicted by myeloid cells (Fig. 4o). This underlines the importance of immune–cardiomyocyte interactions that could additionally explain the increased arrhythmia susceptibility in the remote regions of the post-infarct heart, since it has been shown that cardiac macrophages influence normal and aberrant cardiac conduction\textsuperscript{25}. Our results showed that the stressed cardiomyocyte vCM\textsubscript{3} can be found in distinct spatial cell-type neighbourhoods enriched by different compositions of vSMCs, fibroblasts, adipocytes or myeloid cells.

### Cardiac endothelial cell heterogeneity

Co-embedding of snRNA- and snATAC-seq data identified five subtypes of endothelial cells from all major vascular beds, namely capillary endothelial cells, arterial endothelial cells, venous endothelial cells, lymphatic and endocardial endothelial cells (Fig. 5a, Extended Data Fig. 5a–d, and Supplementary Table I2). Subtype-based pseudo-bulk ATAC-seq signals also revealed distinct chromatin accessibility of these marker genes (Extended Data Fig. 8c). Our analysis suggested POST\textsubscript{N} as a characteristic marker for endocardial endothelial cells, which we validated using smFISH (Extended Data Fig. 8e). Analysis of cell proportions among the myogenic-enriched, ischaemic-enriched and fibrotic-enriched samples revealed a reduction of capillary endothelial cells in the ischaemic samples associated with a concordant increase in venous endothelial cells (Fig. 5b and Extended Data Fig. 8f,g). Furthermore, we observed that lymphatic endothelial cells were overall less abundant than the other populations, as expected, but were significantly increased in the ischaemic zone, suggesting an increased abundance of lymphatics modulating the immune response following cardiac injury\textsuperscript{26} (Fig. 5b).

We modelled the association of the different endothelial cell subtypes with the abundances of the other major cell types in spatial transcriptomics. We observed that the markers of arterial endothelial cells were best predicted by vSMCs within a spot and in the local neighbourhood (radius of five spots) reflecting the anatomy of arteries and arterioles in the heart (Fig. 5c,d and Extended Data Fig. 8h). Moreover, the expression of markers of capillary endothelial cells were best predicted by the presence of pericytes in the tissue, in line with the known presence and role of pericytes in direct contact with capillary endothelium\textsuperscript{14} (Extended Data Fig. 5i). The other endothelial subtypes were mainly predicted by the presence of fibroblasts within a spot and in the local neighbourhood (Extended Data Fig. 8h). Additionally, we observed that the abundance of myeloid cells correlated with the expression of markers of lymphatic endothelial cells (Extended Data Fig. 8h). Focusing on molecular niche 10, which contained the highest cell proportion of endothelial cells and additionally pericytes and mast cells (Extended Data Fig. 8j), we observed a significant enrichment of capillary endothelial cells (Extended Data Fig. 8k). Pathway analysis revealed a significantly higher hypoxia and TGF\textsubscript{B} signalling activity in ischaemic and in fibrotic samples, underlining the importance of these processes in chronic fibrotic cardiac remodelling processes.
(Extended Data Fig. 8l). Pathways important for endothelial signalling in homeostasis such as PI3K and TRAIL showed a reduction in the fibrotic and ischaemic groups, respectively, highlighting further the differential endothelial cell signalling changes. Gene set enrichment analysis further revealed an altered metabolism (for example, fatty acid metabolism and oxidative phosphorylation) of this endothelial cell niche in diseased samples which was further associated with an increased inflammatory response via the TNF and Nfkb pathways and increased apoptosis signalling (Extended Data Fig. 8m). In summary, we resolved all major endothelial cell states, localized them in space and described their spatial dependencies. Further, we identified a spatial niche enriched in capillary endothelial cells with complex metabolic and signalling changes.

Cardiac myofibroblast differentiation

To dissect molecular and cellular mechanisms of fibrogenesis in the human heart, we clustered all fibroblasts using the integrated snRNA-seq and snATAC-seq data and identified four sub-clusters (Fib1–4) (Fig. 6a, Extended Data Fig. 9a and Supplementary Table 13). Fib1 was marked by SCARA5, which we recently reported for as a marker for myofibroblast progenitors in the human kidney34. Fib2 was marked by POSTN, COL1A1 and FN1, which, together with the fact that this population expresses most extracellular matrix (ECM)-related genes, suggests that Fib2 indeed comprises terminally differentiated myofibroblasts (Fig 6b and Extended Data Fig. 9a–c). Notably, Fib2 also exhibited an upregulation of RUNX1, which we recently been reported as being involved in kidney myofibroblast differentiation35. Overexpression of RUNX1 in human heart PDGF-Rβ-expressing cells led to increased myofibroblast differentiation and matrix expression (Extended Data Fig. 9d). We validated the presence of high SCARA5 expression in fibroblasts by co-staining with the pan-fibroblast and myofibroblast marker COLISIA1 as well as POSTN and COL1A1 in human heart tissues, and demonstrated that POSTN is significantly enriched in COL1A1+ cells compared with SCARA5+ cells (Extended Data Fig. 9e). Visualization of these markers in our spatial transcriptomics dataset suggested that Fib1 and Fib2 were enriched in mutually exclusive regions of the heart following injury (Fig. 6c and Extended Data Fig. 9f). Additionally, we observed that Fib1 comprised the highest proportion in myogenic-enriched samples, whereas Fib2 (myofibroblasts) were significantly enriched and Fib3 slightly reduced in ischaemic samples (Fig. 6d and Extended Data Fig. 9g–h).

To precisely understand differentiation trajectories of fibroblasts and transfer this knowledge to the human data, we performed inducible lineage tracing in mice using the pan-mesenchymal Cre driver Pdgfrb-CreER (crossed to a R26-tomato reporter) combined with scRNA-seq at different time points following myocardial infarction (Extended Data Fig. 9i–l). We integrated and annotated the cells by label transfer (Fib1–4) from human to mouse (Extended Data Fig. 9m,n). We observed an overall increase of the Fib2 population and collagen and ECM genes over time, whereas the Fib1 proportion was decreased, pointing towards a differentiation trajectory from SCARA5+ fibroblasts (Fib1) to myofibroblasts (Fib2) in mice (Extended Data Fig. 9o,p). Based on these observations, we inferred a pseudotime trajectory from Fib1 (SCARA5+) to Fib2 (myofibroblast) in the human samples, which was further supported by an increased enrichment of the ECM score (Fig. 6e,f) and of ECM biological gene ontology processes consistent with fibroblast-to-myofibroblast differentiation (Extended Data Fig. 9q).

To understand the regulatory mechanisms of these stromal cell differentiation processes we inferred a fibroblast eGRN (Fig. 6g, Extended Data Fig. 10a,b and Supplementary Table 14). Clustering resolved two eGRN modules that each corresponded to a distinct fibroblast state (Extended Data Fig. 10c) and identified potential regulators of myofibroblast differentiation (Fig. 6g). Among the transcription factors regulating the Fib1 module was KLF4, which regulates diverse cellular functions including cellular growth arrest, and is also one of the original reprogramming factors of induced pluripotent stem cells. Our network analysis highlighted the role of KLF4 in regulating SCARA5 and PCOLCE2 expression in Fib1, and it also targets MBLN1, an important regulator of cardiac wound healing36 and fibroblast-to-myofibroblast transition37. Concordantly, we observed reduced KLF4 binding activity and reduced SCARA5 expression in our pseudotime analysis (Fig. 6h), highlighting the role of KLF4 as a putative inhibitor of fibroblast activation. Among the transcription factors identified in the Fib2 module were TEAD3 (an effector of the Hippo pathway), GLI2 (in the hedgehog pathway) and RUNX2, which have been previously identified as regulators of myofibroblast differentiation38 (Fig. 6h and Extended Data Fig. 10d,e). Our network analysis revealed that both TEAD3 and GLI2 regulate bone fide myofibroblast target genes including COL1A1, TGFB1 and POSTN. Additionally, our network analysis identified the key anti-angiogenic regulator THBS139 as a direct target of TEAD3 and the recently identified cardiac fibrosis regulator MOEX1 in human cardiac myofibroblasts40. We next visualized the expression of the KLF4 and TEAD3 target genes in spatial transcriptomics slides and observed gradients and mutually exclusive spatial expression in defined cardiac regions of fibrotic responses, highlighting their differential spatial activity in the human heart (Fig. 6i and Extended Data Fig. 10d).

Fibro-myeloid spatial relations

Myeloid-derived cells have been reported to have key roles in cardiac remodelling following myocardial infarction46. To understand their heterogeneity, we sub-clustered them using the multi-omic data and identified five sub-clusters across all myocardial infarction samples (Fig. 6j,k, Extended Data Fig. 11a–d and Supplementary Table 15). We observed that two clusters showed expression of resident myeloid cell markers42 (LYVE-1 and F4/80 expressing myeloid clusters), as well as a CCL18- and SPP1-expressing macrophage cluster and a monocyte and classical dendritic cell cluster (Fig. 6j and Extended Data Fig. 11b–d). We used an independent snRNA-seq dataset of three acute human myocardial infarction samples as reference for validation and found high concordance in terms of myeloid cell populations based on marker gene expression (Extended Data Fig. 11e). Cell proportion analysis revealed an increased abundance of a macrophage population defined by SPP1 expression in the ischaemic sample group, whereas CCL18+ macrophages were increased in fibrotic samples (Extended Data Fig. 11f). SPP1+ macrophages have been described in pulmonary fibrosis and COVID-1943,44, and recent work suggests a role of these cells in cardiac tissue remodelling in zebrafish45. We observed an upregulation of CD36 in the SPP1+ myeloid population; CD36 encodes a macrophage receptor known to be important for macrophage phagocytosis, binding to apoptotic and dead neutrophils and having a unique role in cardiac remodelling following myocardial infarction46 (Extended Data Fig. 11b). Indeed, smFISH staining of SPP1+ macrophages suggests increased phagocytic activity, since multiple intracellular vacuoles could be observed (Extended Data Fig. 11g,h). Quantification of multiplex in situ hybridization of SPP1, TREM2 and CCR2 in human myocardial infarction tissue specimens revealed that approximately half of all TREM2-expressing myeloid cells also express SPP1, whereas CCR2+ myeloid cells where less frequent (Extended Data Fig. 11i). Cell-dependency analyses of myeloid cell states revealed a close interaction for two identified LYVE-1 resident macrophage populations, whereas the disease-enriched SPP1+ macrophages predicted the presence of CCL18+ macrophages (Extended Data Fig. 11j,k).

Following acute myocardial infarction, an inflammatory response is triggered, resulting in tissue remodelling that can lead to heart failure47. It has been demonstrated that SPP1 itself can activate fibroblasts in vitro48, highlighting the fibro-myeloid signalling interaction as a crucial driver of the cardiac remodelling process. To further gain insights
about the spatial dependencies of the myeloid and fibroblasts states, we modelled their marker expression using the spatial transcriptomics data. We observed that the presence of \textit{SPP1} macrophages better predicted all fibroblasts states compared to other myeloid cell states, with a higher importance for myofibroblasts within a spot and in the local neighbourhood (Fig. 6i and Extended Data Fig. 12a). Myofibroblast...
marker expression aligned with a gradient of expression of the markers of SPP1+ macrophages (Fig. 6m). This pattern was also recovered by our cell-type niche definition, in which the inflammatory niche 5 was surrounded by the fibrotic-rich niche 4 (Extended Data Fig. 12b), which we could confirm by a higher expression of SPP1+ macrophages and myofibroblast marker genes in niche 5 compared with niche 4 (Extended Data Fig. 12c). As our data pointed towards a clear spatial association of myofibroblasts and fibroblasts, and spatially associated cells are presumably more likely to communicate with each other, we next used receptor–ligand interaction analysis to study their cellular crosstalk. We observed an overall complex myofibroblast–fibroblast interaction (Extended Data Fig. 12d), and detected distinct changes in crosstalk between SPP1+ macrophages and Fib2. This included increased PDGF-C, PDGF-D and THBS1 signalling in ischaemic versus myogenic samples and increased ADAM17 and TGFB1 in fibrotic versus myogenic samples (Extended Data Fig. 12e). Of note, we observed enhanced TGFB1 signalling in ischaemic versus myogenic samples towards Fib3 (Extended Data Fig. 12f). To validate the spatial interaction of SPP1+ macrophages and Fib2, we performed RNA in situ hybridization on human cardiac tissues following myocardial infarction and could confirm the spatial interaction and enrichment of SPP1+ macrophages in an independent tissue cohort ($n = 26$ patients) using an orthogonal method (Fig. 6n and Extended Data Fig. 12g).

In summary, we have decoded cellular fibroblast and myofibroblast heterogeneity and spatial modelling of the fibro-myeloid cell states, revealing a unique interaction of SPP1+ macrophages with myofibroblasts across the different stages of human cardiac tissue remodelling.

Discussion

In multicellular organs, such as the human heart, normal cellular function and tissue homeostasis depend on the interaction between neighbouring individual cell types. Single-cell technologies can profile the molecular heterogeneity of the different cell types and changes that occur during disease. However, without spatial context it is unclear how these different cell types interact in space to coordinate tissue functions. Here we provide a comprehensive map of the human heart at early and late stages after myocardial infarction compared to control hearts (non-transplanted donor hearts) by integrating spatial transcriptomics with single-nucleus gene expression and chromatin accessibility data.

Our computational analyses enabled an increased resolution of spatial transcriptomics by estimating cell-type compositions for each location and by estimating pathway activities, mapping transcription factor binding activities, and projecting GWAS SNPs. These different layers of biological information enabled us to link the organization in human heart tissue specimens of different histomorphological regions, different time points after myocardial infarction and different individual to cellular functions. Here we characterized inflammatory and fibrotic remodelling events that differentiated functional myocardium from ischaemic and chronically remodelled tissue. We explored the effects that these remodelling events had on cardiac architecture, specifically on the vasculature and the dependencies between fibroblasts and myeloid cells. Furthermore, we identified spatial enrichment of different functional states of myogenic regions in control, remote and border zones that were not captured by looking at cell-type compositions or histology only.

Analysis of the integrated snRNA-seq and snATAC-seq data identified different cell states and subtypes for cardiomyocytes, endothelial cells, fibroblasts and myeloid cells. We observed distinct cardiomyocyte cell states associated with spatial distribution, pathway activity and disease condition. Leveraging our multi-omic data, we inferred an egRN and identified potential regulators of cardiomyocytes and fibroblasts, which were also reflected in spatial transcriptomics data. Our data revealed a distinct niche of the border zone surrounding the injured myocardium, with a sharp border between injured and uninjured cell types and were marked by a gradient of ANKRD1 and NPPB expression. Late-stage remodelling after myocardial infarction was driven by fibroblasts, with fibroblast-to-myofibroblast differentiation in distinct tissue areas. Our data provide novel insights into myofibroblast differentiation in human hearts after myocardial infarction, with distinct gene expression and gene-regulatory programmes driving this process. In addition, we decoded the fibroblast myeloid cellular heterogeneity after human myocardial infarction and identified a distinct cellular dependency between myofibroblasts and activated phagocytic macrophages (SPP1+CD36+). The combination of spatial technologies with single-cell data represented an opportunity to study how cardiac cell states are influenced by their tissue microenvironment. The identified interactions between cell types largely reflect the spatial organization of the tissue and, although many other factors are involved, these interactions provide hypotheses for further analysis. Of note, we observed high levels of cell death in the ischaemic samples, as expected, and thus also higher levels of ambient RNA, which could introduce a bias in the analyses. Furthermore, we cannot exclude an overestimation of cardiomyocytes in our cell-type proportion analysis, since about 25% of adult human cardiomyocytes are binucleated, although multiple nuclei in a cell are reported to be transcriptionally homogenous.

We envision that our publicly available atlas will serve as a reference for future studies integrating single-cell genomics and epigenomics with spatial gene expression data of the human heart. Furthermore, we believe that our data will facilitate the understanding of spatial gene expression and gene-regulatory networks within the human myocardium and will be a resource for future studies that aim to understand the function of distinct cardiac cell types in cardiac homeostasis and disease.

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/s41586-022-05060-x.

1. Wong, N. D. Epidemiological studies of CHD and the evolution of preventive cardiology. Nat. Rev. Cardiol. 11, 276–289 (2014).
2. Niccoli, G. et al. Optimized treatment of S-t elevation myocardial infarction. Circ. Res. 125, 245–258 (2019).
3. Prabhu Sumanth, D. & Frangogiannis, Nikolaos, G. The biological basis for cardiac repair after myocardial infarction. Circ. Res. 119, 91–112 (2016).
4. Litvivukova, M. et al. Cells of the adult human heart. Nature 588, 466–472 (2020).
5. Wang, L. et al. Single-cell reconstruction of the anterior heart field during mouse embryonic development. Development 131, 3931–3942 (2004).
6. Yan, C., Zhu, M., Stagler, J., Johnson, P. F. & Gao, H. C5a-regulated CCAAT/ enhancer-binding proteins β and δ are essential in Fcγ receptor-mediated inflammatory cytokine and chemokine production in macrophages. J. Biol. Chem. 287, 3317–3320 (2012).
7. Kovary, K. & Bravo, R. The jun and fos protein families are both required for cell cycle progression in fibroblasts. Mol. Cell. Biol. 11, 4466–4472 (1991).
8. Liu, S., Wang, D. Z., Wang, Z., Richardson, J. A. & Olson, E. N. The serum response factor coactivator myocardin is required for vascular smooth muscle development. Proc. Natl Acad. Sci. USA 100, 9366–9370 (2003).
9. Ge, Y. et al. Switching macrophage gene expression from inflammation-resolution to hemorrhage-resolution by redirection of activating transcription factor 1 (ATF1) binding by SMARCA4, BACH1 and histone H3K9 acetylation. Atherosclerosis 315, e2 (2020).
10. Pirruccello, J. P. et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. Nat. Commun. 11, 2524 (2020).
11. Ruetten, H., Dimmeler, S., Gehring, D., Ihling, C. & Zeiher, A. M. Concentric left ventricular hypertrophy and heterogeneity are supported by macrophages during skin repair. Science 362, eaar2971 (2018).

776 | Nature | Vol 608 | 25 August 2022
15. Pakshir, P. et al. Dynamic fibroblast contractions attract remote macrophages in fibrotic collagen matrix. Nat. Commun. 10, 1850 (2019).
16. Aoyagi, T. & Matsui, T. Phosphoinositide-3 kinase signaling in cardiac hypertrophy and heart failure. Curr. Pharm. Des. 17, 1818–1824 (2011).
17. Mak, T. W., Hauck, L., Grothe, D. & Billia, F. p53 regulates the cardiac transcriptome. Proc. Natl Acad. Sci. USA 114, 2331–2336 (2017).
18. Bergmann, M. W. WNT signaling in adult cardiac hypertrophy and remodeling: lessons learned from cardiac development. Circ. Res. 107, 1196–1208 (2010).
19. Ahern, B. M. et al. Myocardial-restricted ablation of the GTPase RAD results in a pro-adaptive heart response in mice. J. Biol. Chem. 294, 10913–10927 (2019).
20. Lotteau, S. et al. Acute genetic ablation of cardiac sodium/calcium exchange in adult mice: implications for cardiomyocyte calcium regulation, cardioprotection, and arrhythmia. J. Am. Heart Assoc. 10, e019273 (2021).
21. Fernandez-Caggiano, M. et al. Mitochondrial pyruvate carrier abundance mediates pathological cardiac cardiac hypertrophy. Nat. Metab. 2, 1223–1231 (2020).
22. Hama, N. et al. Rapid ventricular induction of brain natriuretic peptide gene expression in experimental acute myocardial infarction. Circulation 92, 1558–1564 (1995).
23. Waspe, L. E., Ordahl, C. P. & Simpson, P. C. The cardiac β-myosin heavy chain isogene is induced selectively in α1-adrenergic receptor-stimulated hypertrophy of cultured rat heart myocytes. J. Clin. Invest. 85, 1206–1214 (1990).
24. Beggah, A. T. et al. Reversible cardiac fibrosis and heart failure induced by conditional expression of an antisense mRNA of the mineralocorticoid receptor in cardiomyocytes. Proc. Natl Acad. Sci. USA 99, 7160–7165 (2002).
25. Bakker, M. L. et al. T-box transcription factor TBX3 reprogrammes mature cardiac myocytes into pacemaker-like cells. Cardiovasc. Res. 94, 439–449 (2012).
26. Jiang, J. et al. Cardiac myosin binding protein C regulates postnatal myocardial cytokinesis. Proc. Natl Acad. Sci. USA 112, 9046–9051 (2015).
27. Pirroddi, N. et al. Myocardial overexpression of ANKRD1 causes sinus venous defects and progressive diastolic dysfunction. Cardiovasc. Res. 116, 1458–1472 (2020).
28. Hill, C. et al. Inhibition of AP-1 signaling by JDP2 overexpression protects cardiomyocytes against hypertrophy and apoptosis induction. Cardiovasc. Res. 99, 121–128 (2013).
29. Kraman, R. et al. Perivascular Gfli+ progenitors are key contributors to injury-induced organ fibrosis. Cell Stem Cell 16, 51–66 (2015).
30. Hulsmans, M. et al. Macrophages facilitate electrical conduction in the heart. Cell 169, 510–522.e20 (2017).
31. Vieira, J. M. et al. The cardiac lymphatic system stimulates resolution of inflammation following myocardial infarction. J. Clin. Invest. 128, 3420–3432 (2018).
32. Armulik, A., Abramsson, A. & Betsholtz, C. Endothelial/periocyte interactions. Circ. Res. 97, 512–523 (2005).
33. Deshpande, S. S., Anglew, P., Huang, J., Ozaki, M. & Irani, K. Raci inhibits TNF-α-induced endothelial cell apoptosis: dual regulation by reactive oxygen species. FASEB J. 14, 1705–1714 (2000).
34. Kuppe, C. et al. Decoding myofibroblast origins in human kidney fibrosis. Nature 589, 281–286 (2021).
35. Li, Z. et al. Chromatin-accessibility estimation from single-cell ATAC-seq data with scOpen. Nat. Commun. 12, 6386 (2021).
36. Bugg, B. et al. MBNL1 drives dynamic transitions between fibroblasts and myofibroblasts in cardiac wound healing. Cell Stem Cell 29, 419–433.e10 (2022).
37. Davis, J. et al. MBNL1-mediated regulation of differentiation RNAs promotes myofibroblast transformation and the fibrotic response. Nat. Commun. 6, 10084 (2015).
38. Kramann, R. et al. Pharmacological GLI2 inhibition prevents myofibroblast cell-cycle progression and reduces kidney fibrosis. J. Clin. Invest. 125, 2935–2951 (2015).
39. Lawler, J. Thrombospondin-1 as an endogenous inhibitor of angiogenesis and tumor growth. J. Cell. Mol. Med. 6, 1–12 (2002).
40. Alexanian, M. et al. A transcriptional switch governs fibroblast activation in heart disease. Nature 595, 438–443 (2021).
41. Forte, E. et al. Cross-priming dendritic cells exacerbate immunopathology after ischemic tissue damage in the heart. Circulation 143, 821–836 (2021).
42. Dick, S. A. et al. Three tissue resident macrophage subsets coexist across organs with conserved origins and life cycles. Sci. Immunol. 7, eabb7777 (2022).
43. MacDonald, L. et al. COVID-19 and RA share an SPP1 myeloid pathway that drives PD-L1+ neutrophils and CD14+ monocytes. JCI Insight 6, e147413 (2021).
44. Morse, C. et al. Proliferating SPP1/CD14+ myeloid cells mediate fibrosis in idiopathic pulmonary fibrosis. Eur. Respir. J. 54, 180233 (2019).
45. Blevin, L. et al. Specific macrophage populations promote both cardiac scar deposition and subsequent resolution in adult zebrafish. Cardiovasc. Res. 116, 1357–1371 (2020).
46. DeLeon-Pennell, K. Y. et al. CD36 is a matrix metalloproteinase-9 substrate that stimulates neutrophil apoptosis and removal during cardiac remodeling. Circulation. Cardiovasc. Genet. 9, 14–25 (2016).
47. Ismail, M. A. et al. Remodeling of the mononuclear phagocyte network underlies chronic inflammation and disease progression in heart failure: critical importance of the cardiopulmonary axis. Circulation 134, 266–282 (2014).
48. Shiraishi, H. et al. CX3CR1 and CD14+ monocytes mediate repair of the infarcted adult murine heart. J. Clin. Invest. 126, 2151–2166 (2016).
49. Bergmann, O. et al. Dynamics of cell generation and turnover in the human heart. Cell 161, 1566–1575 (2015).
50. Yekelchyk, M., Guenther, S., Preussner, J. & Braun, T. Mono- and multi-nucleated neutrophils and CD14+ monocytes. JCI Insight 6, e147413 (2021).

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s). Author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022
Methods

Ethics
The local ethics committee of the Ruhr University Bochum in Bad Oeynhausen, the RWTH Aachen University, Utrecht University and WUSTL approved all human tissue protocols (no. 220-640, EK151/09, 12/387 and no. 201104172 respectively). Human myocardial tissue was collected from non-transplanted donor hearts, patients after myocardial infarction undergoing heart transplantation, implantation of a total artificial heart or left ventricular assist device (LVAD) implantation. The study met all criteria of the code of conduct for responsible use of human tissue that is used in the Netherlands. The collection of the human heart tissue was approved by the scientific advisory board of the biobank of the University Medical Center Utrecht, The Netherlands (protocol no. 12/387). All patients provided informed consent and the study was performed in accordance with the Declaration of Helsinki. Written informed consent for collection and biobanking of tissue samples was obtained prior to LVAD implantation.

Human tissue processing and screening
Heart tissues were sampled by the surgeon and immediately frozen in liquid nitrogen. Tissues were homogenized in liquid nitrogen and 7–10 mm³ pieces were embedded in OCT compound (Tissue-Tek) and frozen on dry-ice. Ten-micrometre tissue cryosections were stained with H&E and the appropriate tissue regions were selected for further processing. In total 52 human tissue samples were screened this way and evaluated by a cardiac pathologist. For RNA quality control we minced a 3 × 3 mm² heart tissue piece in liquid nitrogen and isolated the RNA using Qiagen RNeasy Mini kit (Qiagen) using a protease K digestion step as suggested in RNeasy Fibrous Tissue Mini Kit (Qiagen, 74704). RNA integrity number (RIN) analysis (Agilent) was performed using Bioanalyzer RNA 6000 Nano kits (Agilent, No. 5067). RIN ranged from >2 to a maximum of 8.8.

Spatial gene expression assay
Frozen heart samples were embedded in OCT (Tissue-Tek) and cryosectioned (Thermo Cryostar). The 10-µm section was placed on the pre-chilled Optimization slides (Visium, 10X Genomics, PN-1000193) and the optimal lysis time was determined. The tissues were treated as recommended by 10X Genomics and the optimization procedure showed an optimal permeabilization time of 12 or 18 min of digestion and release of RNA from the tissue slide. Spatial gene expression slides (Visum, 10X Genomics, PN-1000187) were used for spatial transcriptomics following the Visium User Guides. Brightfield histological images were taken using a 10X objective on the Nikon Eclipse TiE and a Leica Aperio Versa 200 scanner. Stitching of the raw images was performed using the NIS-Elements software. Next generation sequencing libraries were prepared according to the Visium user guide. Libraries were loaded at 300 pm and sequenced on a NovaSeq 6000 System (Illumina) as recommended by 10X Genomics.

Single-nuclei isolation of human hearts
Single-nuclei isolation was performed as previously described. Briefly, heart tissue was cut into small pieces (0.5 mm³) in a sterile petri dish on ice and transferred to a tissue homogenizer. Nuclei isolation buffer 0.5 ml (EZ lysis buffer, NUC101 Sigma-Aldrich) plus RNase inhibitor (Protector RNase Inhibitor, Roche) were added to the tissue, and 10–15 strokes with pestle A were applied followed by 10–15 strokes of pestle B. The nuclei were stained with DAPI and FANS sorted using a Sony SH800 to enrich the nuclei. Nuclei isolation from three acute myocardial infarction samples from the WUSTL biobank was performed as described.

scRNA-seq
Nuclei suspensions with a concentration ranging from 400–1000 nuclei per µl were loaded into the chromium controller (10X, Genomics, PN-120223) on a Single Cell B chip (10X Genomics, PN-120262) and processed following the manufacturer’s original protocol to generate single-cell gel beads in the emulsion. The sequencing library was generated using the Chromium Single cell 3’ reagent kit v3 (10X, PN-1000092) and Chromium i7 Multiplex kit (10X Genomics, PN-120262). Quality control for the constructed library was performed by Tape Station. Libraries were sequenced on NovaSeq targeting 50,000 reads per cell 2 × 150 paired-end kits using the following read length: 28 bp Read1 for cell barcode and UMl, 8 bp i7 index for sample index, and 91-bp Read2 for transcript.

sc-ATAC-seq
The remaining nuclei after processing for 3’ scRNA-seq assay were centrifuged at 500g at 4 °C for 5 min and resuspended in 10 µl of nuclei suspension buffer. After tagmentation the nuclei suspension was loaded on the Chromium Chip E (10X Genomics, PN-1000082) in the Chromium controller according to the manufacturer’s protocol. The library was sequenced on an Illumina NovaSeq 6000 using the following read length: 50 bp Read1 for DNA fragments, 8 bp for i7 index for sample index, 16 bp i5 index for cell barcodes, and 50 bp Read2 for DNA fragments.

RNA in situ hybridization and image quantification
In situ hybridization was performed using formalin-fixed paraffin embedded tissue samples and the RNAscope Multiplex Detection Kit V2 (RNAscope, cat. no. 323100) and RNAscope 4-Plex Ancillary Kit (RNAscope, cat. no. 323120) following the manufacturer’s protocol with minor modifications. The antigen retrieval was performed for 30 min at 99 °C in a water bath (VWR). Tissue pretreatment and washing was performed as suggested by the RNAscope staining protocol. The following probes were used for the RNAscope assay: Hs-C163 cat. no. 417061-C1, Hs-C221 cat. no. 438221-C1, Hs-ANKRD1 cat. no. 52421-C1, Hs-POSTN cat. no. 575941-C1, Hs-CC21 cat. no. 484001-C2, Hs-CC21 cat. no. 401891-C2, Hs-PSTN cat. no. 487381-C2, Hs-NPPB cat. no. 448511-C2, Hs-TRE2M cat. no. 420409-C1, Hs-SPPL cat. no. 420101-C2, Hs-NPR3 cat. no. 431241-C3, Hs-POSTN cat. no. 409181-C3, Hs-SCARA5 cat. no. 574781-C3, Hs-TNNT2 cat. no. 518991-C3, Hs-SPPL cat. no. 420101-C4 and Hs-NF-221L cat. no. 53850.

Nuclei quantification of H&E stained Visium slides
To quantify nuclei from the H&E staining, we used VistoSeg, an automated MATLAB pipeline for image analysis. Using this pipeline, the individual TIFF files were used for nuclei segmentation using k-means colour-based segmentation in the image processing toolbox. Next, the binary images were refined with the refineVNS() function for accurate detection of the nuclei. Then a CSV and JSON file was generated that contained the metrics to reconstruct the spot grid to allow for nuclei quantification per 10X Visium detection spot. Counting of nuclei was performed with the countNuclei() function. The images were checked individually with the spotspotcheck() function. Code is available at http://research.libd.org/VistoSeg.

Animal model of myocardial infarction
Myocardial infarction was performed as previously described. In brief, 12-week-old male and female C57Bl/6j Pdgfrb-creER;tdTomato mice were subjected to chronic left anterior descending artery ligation. The mice were anaesthetized using isoflurane (2–2.5%). The mice were injected 30 min before surgery with metamizole (200 µg g⁻¹ body weight) subcutaneously. Then they were intubated and ventilated with oxygen using a mouse respirator (Harvard Apparatus). Before incision, we injected bupivacaine (2.5 µg g⁻¹ body weight) subcutaneously and intercostally for local analgesia. Then a left thoracotomy was performed, and myocardial infarction was induced by ligation of the left anterior descending artery with 0/7 silk (Seraflex, IO05171Z). The ribs, muscle layer and skin incision were closed. Metamizole was administered for three days via drinking water (1.25 mg ml⁻¹, 1% sucrose).
post-surgery. All mice were housed under standardized conditions in the Animal Facility of the University Hospital Aachen (Germany). The operating procedure was in accordance with European legislation and approved by local German authorities (LANUV, reference no. 81-02.04.2017.A410.). Mice were euthanized at different time points (sham, 4 days, 7 days and 14 days). As control, hearts from sham-operated, age-matched mice were taken (2 sham female and 2 sham male mice).

**Inducible fate-tracing experiments**

For inducible fate tracing, male and female Pdgfrb-creERT2 Tomato mice (8 weeks of age) received tamoxifen (3 mg intraperitoneally) 4 times followed by a washout period of 21 days and were then subjected to myocardial infarction surgery or sham (12 weeks of age) as described above and euthanized at 4 days, 7 days and 14 days after surgery.

**Echocardiography**

Left ventricular heart function was determined by echocardiography performed on a small-animal ultrasound imager (Vevo 3100 and Mx550D transducer, FUJIFILM Visualsonics). Recordings of short and long cardiac axis were taken in B mode (2D real-time) using a 40 MHz transducer (Mx550D). During the procedure, mice were anesthetized with 1–2% isoflurane. Ejection fraction (EF) and global longitudinal strain (GLS) were recorded and analysed with the VevoLab Software. The Simpson method was used to assess EF. The GLS was measured in the B-mode of the long axis.

**smFISH spot quantification and nuclear segmentation**

Images for smFISH were exported in native Nicon format (.nd2). Images were split by channel using bfconvert for further processing. RNA spots were quantified using the command line version of Radial Symmetry-FISH (RS-FISH). The sigma parameter from RS-FISH, the R parameter from the following values for cardiomyocyte state analysis: channel 1 (TNN1T2) = 0.0107, channel 2 (ANKRD1) = 0.005, channel 3 (NPPB) = 0.0066. To remove spot counts resulting from low signal, high background images, we removed spots with an intensity lower than the 25th percentile of the channel intensity distribution across all images and applied a minimum intensity threshold of 600. For the quantification of CD163 (SPP1) macrophages, while we were not able to perform full cell segmentation, we performed nuclear segmentation using Mesmer with pre-trained nuclear segmentation models to identify all detectable nuclei in each image based on DAPI staining. We subsequently assigned spots to the closest nuclei based on euclidean distance and classified cells as positive or negative for the different markers (POSTN, CD163 and SPP1). Cells with more than 2 spots for a given marker were considered positive for that marker.

**Masson trichrome staining**

Masson's trichrome staining was conducted using aready-to-use kit (Trichrome Stain Masson Kit, HT15, Sigma-Aldrich) as described by the manufacturer.

**Antibodies and immunofluorescence staining**

Heart tissues were fixed in 4% formalin for 4 hours at room temperature and then embedded in paraffin. For staining slides were blocked in 5% donkey serum followed by 1 hour of incubation with the primary antibody, washing 3 times for 5 min in PBS, and subsequent incubation of the secondary antibodies for 45 min. Following DAPI (4',6-diamidino-2-phenylindole) staining (Roche, 1:10,000) the slides were mounted with ProLong Gold (Invitrogen, cat. no. P10144). The following antibodies were used: anti-ACTA2 (aSMa)-Cy3 (C6198, 1:250, Sigma-Aldrich), anti-SEMA3G (HPA001761, 1:100, Sigma-Aldrich), AF647 donkey anti-rabbit (1:200, Jackson ImmunoResearch).

**Confocal imaging**

Acquisition of images was performed using a Nikon AF confocal microscope using 40× and 60× objectives (Nikon). Image processing was performed using the Nikon Software or ImageJ.

**Generation of a human PDGFBc cardiac cell line**

PDGFBc cells were isolated from a 69-year-old male patient, undergoing left ventricular assist device surgery. To generate a single-cell suspension, the tissue was homogenized in gentleMACS dissociator (Miltenyi) and digested with liberase (200 µg ml⁻¹, Roche cat. no. 5401020001) and DNase (60 U ml⁻¹), for 30 min at 37°C. After filtering the cell suspension (70 µm mesh), cells were stained in two steps using a specific PDGFRB- antibody (R&D cat. no. MAB1263 antibody, dilution 1:100) followed by Anti-Mouse IgGl MicroBeads solution (Miltenyi, cat. no. 130-047-102). Following MACS isolation, cells were cultured in DMEM medium (Thermo Fisher cat. no. 31885) for 20 days and immortalized using SV40-LT and HTERT. Retroviral particles were produced by transient transfection of HEK293T cells using TransIT-LT (Mirus). Two types of amphotropic particles were generated by co-transfection of plasmids pBabe-puro-SV40-LT (Addgene #13970) or xol-xDGR-FRT (Addgene #69105) in combination with a packaging plasmid pUMVC (Addgene #8449) and a pseudotyping plasmid pMD2.G (Addgene #12259). Retroviral particles were 100x concentrated using Retro-X concentrator (Clontech) 48 h post-transfection. Cell transduction was performed by incubating the target cells with serial dilutions of the retroviral supernatants (1:1 mix of concentrated particles containing SV40-LT or rather HTERT) for 48 h. Subsequently at 72 h after transduction, the transduced PDGFBc cells were selected with 2 µg ml⁻¹ puromycin for 7 days.

**Lentiviral overexpression of RUNX1**

The human cDNA of RUNX1 was PCR amplified using the primer sequences 5’-atgcatctacccggtagtgagccacag – 3’ and 5’-tcagtagggcctccacacgg – 3’. Restriction sites and N-terminal IxAHA-Tag were introduced into the PCR product using the primer 5’-cactcgaggccaccatgtacccatacgatgttccagattacgctcg- taccgagtcgtgagcttcg-3’ and 5’-agcaaggattcagggcctccacac-3’. Subsequently, the PCR product was digested with Xhol and EcoRI and cloned into plMIG (pMIG was a gift from W. Hahn) (Addgene plasmid #9044 ; http://n2t.net/omniproc/9044; RRID:Addgene_9044). Retroviral particles were produced by transient transfection in combination with packaging plasmid pUMVC (pUMVC was a gift from W. Weinberg (Addgene plasmid #8449)) and pseudotyping plasmid pMD2.G (Addgene #12259) from D. Trono (Addgene plasmid #12259; http://n2t.net/omniproc/12259; RRID:Addgene_12259) using TransIT-LT (Mirus). Viral supernatants were collected at 48–72 h post-transfection, clarified by centrifugation, supplemented with 10% FCS and Polybrene (Sigma-Aldrich, final concentration of 8 µg ml⁻¹) and filtered with a 0.45-µm PES filter membrane (Millipore; SLHP033RS). Cell transduction was performed by incubating the PDGFBc cells with viral supernatants for 48 h. eGFP-expressing single cells were sorted with a SH800 Cell Sorter.

**Quantitative PCR with reverse transcription**

Cell pellets were collected and washed with PBS followed by RNA extraction using the RNeasy Mini Kit (Qiagen) according to the manufacturer’s instructions. Two-hundred nanograms total RNA was reverse transcribed with High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and quantitative PCR with reverse transcription was carried out as described Data were analysed using the 2⁻ΔΔCt method. The primers used are listed in Supplementary Table 18.

**Preprocessing of snRNA-seq, snATAC-seq and spatial transcriptome data**

For snRNA-seq data, CellRanger software (v6.0.2) was used to perform the alignment with default options. Since the input consists of nuclei, we enabled the option ‘–include-introns’ to include intronic reads.
snATAC-seq data processing
To identify the major lineages representative of all of our specimens, we created a single-nuclei atlas analysing and integrating each snRNA-seq dataset using Seurat\(^63\) (v4.0.1).

Each dataset went through identical quality control processing. We discarded nuclei in the top 1% in terms of the number of genes, (2) with less than 300 genes and less than 500 UMIs, (3) with more than 5% of mitochondrial gene expression, and (4) doublets as estimated using scDbiFinder (v1.4.0)\(^61\) with default parameters. Count matrices were log-normalized for downstream analyses using a scaling factor of 10,000. We calculated a dissociation score for each cell using Seurat’s module score functions with a gene set provided by O’Flanagan et al.\(^61\) and discarded the nuclei that belonged to the top 1%. To generate an integrated atlas of all samples, log-normalized expression matrices were merged and dimensionality reduction was performed on the collection of the top 3,000 most variable genes that were shared with most of the samples using principal component analysis (PCA). To select the collection of shared variable genes between samples, first we estimated the top 3,000 most variable genes per sample and then selected the top 3,000 most-recurrent genes from them across all samples. PCA correction was performed with harmony\(^62\) (v1.0) using as covariates the patient, sample, and batch labels. A shared nearest neighbour (SNN) graph was built with the first 30 principal components using Seurat’s FindNeighbors and the cells were clustered with a Louvain algorithm with FindClusters. A high resolution (1) was selected to generate a large collection of nuclei clusters to capture representative major cell lineages, even if present in low proportions. Cluster markers were identified with Wilcoxon tests as implemented in Seurat’s FindAllMarkers function. Final assignment of cells to major cell lineages was based on literature marker genes. We filtered out small clusters (median number of nuclei across filtered clusters > 269) with low gene count distributions (median counts across filtered clusters = 756) or feature recovery (median number of genes across filtered clusters = 695), with marker genes that could not be assigned to known cell types of the heart. To visualize all nuclei in a two-dimensional embedding, a UMAP was created with Seurat’s RunUMAP function using the first 30 principal components of harmony’s PCA correction embedding. Major cell-type markers were estimated by performing differential expression analysis of cell-type and patient-specific pseudo-bulk profiles. Pseudo-bulk profiles were calculated by summing up the counts of all cells belonging to the same cell type and patient. Profiles coming from less than 10 cells or profiles from which the maximum gene expression was of less than 1,000 counts per library were discarded. Differentially expressed genes were calculated by fitting a quasi-likelihood negative binomial generalized log-linear model as implemented in edgeR (v3.32.1)\(^63\) (false discovery rate (FDR) < 0.15). Each cell type was compared against the rest.

Comparison with independent healthy and ischaemic human heart cell atlases
We compared our generated atlas with another reference human single-nuclei RNA-seq atlas\(^4\) at the molecular and compositional level. The counts matrix was downloaded directly from https://www.heartcellatlas.org and we selected the data coming from single-nuclei and left ventricle samples. Nuclei with fewer than 200 genes, and genes expressed in less than 3 nuclei were excluded. Log normalization with a scaling factor of 10,000 was performed with scumpy’s\(^1\) (v1.7.0) normalize_total function.

To evaluate our major cell-type annotation, we calculated the enrichment of the top 200 marker genes based on log fold change of each cell type defined in the reference atlas in the list of the top 200 marker genes of each of our defined cell types with hypergeometric tests. Marker genes of the reference atlas were calculated with Wilcoxon tests as implemented in scumpy’s\(^4\) (v1.7.0) rank_genes_groups (adj. P < 0.01). Each cell type was compared against the rest. To evaluate the compositional stability of our control samples, we calculated the Pearson correlation between the median proportion of each shared cell type of the reference atlas and our control, border zone, and remote zone samples. Similarly, we compared our atlas to an independent collection of human heart nuclei derived from three ischaemic specimens. First, we analysed and integrated the smaller collection of samples using identical procedures as the ones used in our provided atlas. After nuclei clustering, we assigned each cluster to a cell type using literature markers. Cell-type markers were calculated with Wilcoxon tests (adj. P < 0.01) and the top 200 genes based on log fold change were selected. Marker overlap and compositional stability comparison with ischaemic specimens from our atlas were performed as described previously.

Comparison between snRNA-seq and snATAC-seq data
To control the data quality, the fragment files were used as input for the package ArchR (v1.0.1)\(^61\), and low-quality cells were filtered out based on transcription start site (TSS) enrichment (> 4) and the number of unique fragments (> 3,000 and < 100,000). Doublets were identified and removed by using the functions addDoubletScores and filterDoublets from ArchR with default settings. Next, peaks were identified by using the function addReproduciblePeakSet for each sample. All peaks were merged to create a union peak set of which each peak was annotated as distal, promoter, exonic and intronic. A count matrix was constructed with the function addPeakMatrix. For dimensionality reduction, the method scOpen (v1.0.0)\(^65\) was used to generate a low-dimensional matrix of the cells. The algorithm Harmony\(^62\) was applied to correct the batch effects and integrate the data and UMAP was used to generate a 2D embedding for visualization. Cells were clustered using the Leiden algorithm with a resolution of 1. To annotate the clusters, a gene activity score matrix was created using the function addGeneScoreMatrix and marker genes were detected for each cluster using the function getMarkerFeatures. The same markers from snRNA-seq data were used to annotate the clusters.

Cell-type-specific transcription factor binding and regulon activity
To estimate transcription factor binding activity for each major cell type identified from snATAC-seq data, we first aggregated the reads within each cell type and created a pseudo-bulk profile. Next, we used MACS2 (v2.2.7)\(^67\) to perform peak calling and removed the peaks from chrY, mitochondrial and unassembled ‘random’ contigs. We then predicted the transcription factor binding sites and estimated transcription factor binding activity using HINT-ATAC (v0.13.2)\(^4\) based on the JASPAR2020 database\(^4\). We linked the transcription factor binding sites to the nearest genes to create a cell-type-specific transcription factor–gene interaction. The number of ATAC-seq reads in the region with 100 bp up-stream and downstream of the the transcription factor binding
site were used to indicate how strong the interaction was: each transcription factor–gene interaction was weighted as the ratio between the number of ATAC-seq reads around the transcription factor binding site associated with that gene and the maximum number of reads observed in any binding site of the transcription factor. All interactions with a weight larger than 0.3 were considered in downstream analysis. This generated weighted and filtered cell-type-specific regulons. To infer a transcription factor regulon activity score, we estimated the mean expression of the target genes in each cell-type-specific regulon. Cell-type pseudo-bulk profiles were filtered to contain only genes with at least 10 counts in 5% of the samples, before the estimation of normalized weighted means using decoupleR’s19 (v1.0.0) wmean function with 1,000 permutations. Regulon activities were standardized and correlated with transcription factor binding activities using Spearman correlations. The minimum correlation of 0.5 was used as threshold and the top 5 transcription factors per cell type were selected for visualization.

**Cell-type-specific GWAS signal enrichment**

GWAS summary statistics for 4 MRI based left ventricle function parameters64 were downloaded from the Cardiovascular Disease Knowledge Portal (https://cvd.hugeamp.org/). For each phenotype, GWAS summary statistics were clumped with Plink (v1.9)19 to identify index SNPs (clump-pl = 0.0001, clump-kb = 250, clump-r2 = 0.5) using the European samples from 1000 Genomes as a reference population.

Next, we lifted over the coordinates of index SNPs from hg19 to hg38 using the LiftOver tools. For each major cell type, we generated an average chromatin accessibility profile by using snATAC-seq data from all cells. The cell-type-specific GWAS signal enrichment was performed using gchromVAR (v0.3.2)23 and enrichment scores were normalized to z-scores. P-values were calculated based on the z-scores and were corrected by the Benjamini–Hochberg method.

**Cell-type-specific integration of snATAC-seq and snRNA-seq data and sub-clustering**

For each major cell type that was recovered by both snATAC-seq and snRNA-seq, we aimed to identify sub-clusters spanning multiple samples and modalities. To do so, we devised a multi-step approach to integrate and cluster the data by controlling quality from sample-, cell-type- and modality-specific aspects.

1. To minimize the sample-specific effects, we only considered samples with a minimum number of cells in both snATAC-seq and snRNA-seq: for cardiomyocytes and endothelial cells (n_cells_ATAC > 300 and n_cells_RNA > 400); for fibroblasts (n_cells_ATAC > 100 and n_cells_RNA > 400); and for myeloid (n_cells_ATAC > 50 and n_cells_RNA > 200). This step controls for samples with low recovery of cells in a particular modality.

2. To further filter cell-type-specific low-quality cells from snRNA-seq and snATAC-seq data, we integrated the samples as selected in step 1 using Harmony to correct batch effects from patients and regions based on PCA space (30 dimensions) for snRNA-seq and LSI space (30 dimensions) for snATAC-seq data. We then clustered the cells using Seurat (resolution = 0.4) for each modality independently. We next excluded the clusters that were (i) enriched in a single sample; (ii) showed a lower data quality; (iii) showed a higher doublet score compared with others. Specifically, for cardiomyocytes, we removed 3 clusters from snATAC-seq data: 2 clusters (481 cells) were enriched in a single sample and another cluster (171 cells) showed a low number of unique fragments (average = 8,102). For fibroblasts, we removed 1 cluster (49 cells) from snATAC-seq (98% of cells from a single sample) and 1 cluster (1,172 cells) from snRNA-seq (average doublet score of 0.12). This step controls for cell type and modality-specific low-quality cells.

3. We next integrated the cells from snATAC-seq and snRNA-seq data. To this end, we used the gene activity score matrix of snATAC-seq estimated by ArchR and the gene expression data from snRNA-seq data as input for canonical component analysis by Seurat. The integrated data were projected into a PCA space (30 dimensions) and Harmony was used to correct the batch effects from samples and modalities. This step generated a co-embedded and batch-corrected dataset composed of cells from snRNA-seq and snATAC-seq samples.

4. For each major cell type, we defined the sub-clusters based on the co-embedded data using the Seurat (resolution = 0.9 or 1). Marker genes were identified by using the function FindAllMarkers. We next filtered clusters that were mainly driven by a single sample or modality. Finally, we merged and annotated the clusters based on the markers. The final statistics of the sub-clustering results for each major cell type were provided in Supplementary Table 16.

**Analysis of snRNA-seq data from mouse fibroblasts**

Cellranger mkfastq and count functions (version v6.0.2) with default parameters were used for demultiplexing and aligning the reads, respectively. Reads were aligned to the mouse reference transcriptome (mm10, Version=2020-A). Prior to alignment, reads for tdTomato were added to the reference. Quantified counts from each sample were aggregated and cells with counts < 1,500 and > 20,000 were filtered out. Further, cells with > 5% reads mapped to mitochondrial genes, as well as cells with < 500 genes were removed. Scrublet was used to detect potential doublets and only the resulting 40,495 cells with > 0.2 scrublet score were kept for further analyses. The highly_variable_genes() function with seurat_v3 flavour implemented in Scany (version 1.8.1) was used to obtain the top 2,000 most highly variable genes. Count data was log-normalized using sc.ppnormalize_total(target_sum=1e4) followed by sc.pp.log1p(). The data was subset to the 2,000 genes, unwanted sources of variation from n_umi and mito_fraction were regressed out using sc.pp.regress_out(), and the top 30 principal components were estimated using sc.dl.pca(). Harmony was then used to account for large differences across samples using ‘sample’ as the batch indicator. Network neighbourhood graph was constructed using the function sc.pp.neighbors() with 30 adjusted principal components, cosine distance and n_neighbors = 10. Leiden clustering with resolution 1.0 was used to cluster the cells into 17 clusters. Marker genes were identified using the Wilcoxon test implemented in sc.tl.rank_genes_groups() function in Scany. Clusters were manually annotated using the marker genes. We next cleaned up the data by removing clusters with low data quality and re-clustered the data with resolution of 0.2. To annotate the cells, we used the label transfer approach from Seurat based on the sub-clustering results from human fibroblasts.

**Gene-regulatory network inference for cardiomyocytes and fibroblasts**

We inferred an eGRN for cardiomyocytes and fibroblasts using a multi-step approach including modality pairing, transcription factors and genes selection, and network construction.

1. We first paired the cells between snATAC-seq and snRNA-seq based on the previously described co-embedding space using an optimal matching approach19. This method returns a matching of a snATAC-seq cell to a unique cell in snRNA-seq. Next, we produced a diffusion map and created trajectories in this space using the function addTrajectory from ArchR (v1.0.1)46. For cardiomyocytes, we inferred a trajectory from clusters vCM1, vCM2 and vCM3, where vCM1 were considered as roots and vCM3 as the terminal state. For fibroblasts, we built a trajectory with SCARA5 fibroblasts as root and myofibroblasts as terminal state.

2. Next, we predicted a single-cell-specific transcription factor binding activity score using the R package chromVAR (v1.16)39 from the snATAC-seq data based on motif from the JASPAR2020 database60. In contrast to HiNT-ATAC, chromVAR provides transcription factor activity scores at single-cell level. We next selected transcription factors that display concordant binding activity (snATAC-seq) and its gene expression (snRNA-seq) (Pearson correlation > 0.1). This
Different data modalities, we calculated Spearman correlations between
per spot, and 2 combinations per spot. Additionally, for each spot we
matrix by taking the mean estimation. Each slide was later deconvoluted
log-normalized (size factor = 10,000) and scaled matrices were calcu-
and mitochondrial genes were excluded from this analysis. Individual
transcriptomics analysis using Seurat (v4.0.1). Spots with less than 300
count matrices were normalized with sctransform79, and additional
transcription factors along the trajectory as described65. We kept the top 10% variable genes and considered them as potential transcription factor targets.

(3) To associate regulators with targets (that is, transcription fac-
tors with genes), we explored the correlation of peak accessibility
and gene expression to identify peak-to-gene links. Specifically,
for each gene, we consider peaks that are within 125 kb on either
side of the transcription start sites, while excluding the promoter
regions. This analysis generated a list of enhancer-to-promoter
links. We only considered significantly correlated links (FDR <
0.0001) with a positive correlation as before46. Finally, we associ-
ated a transcription factor with a target gene if this gene was linked
to an enhancer and this enhancer was predicted to be found by this
transcription factor.

(4) To build a quantitative transcription factor–gene-regulatory
network, we estimated the correlation of the transcription fac-
tor binding activity from snATAC-seq and target gene expression
from snRNA-seq data and only considered those interactions with
Pearson correlation >0.4. We visualized the network based on a
force-layout, which places transcription factors (or target genes)
with similar interactions close together. We coloured transcription
factor nodes in the networks using the assigned pseudotime labels
as inferred in step 2. To characterize the importance of transcription
factors, we computed two measures: node betweenness (denoted by
b)77 and pagerank (denoted by p)78. A final importance score for
transcription factor i was calculated as:

\[
\text{Importance}_i = (b_i - \text{min}(b))^2 + (p_i - \text{min}(p))^2
\]

(5) Finally, to map the inferred GRN into spatial transcriptomics data,
we used the target genes for each transcription factor and calcu-
lated a module score by using the function AddModuleScore from
Seurat (v4.1.0).

**Characterization of spatial transcriptomics datasets**

**Single-slide processing.** Filtered feature-barcode expression matri-
ces from SpaceRanger (v1.3.2) were used as initial input for the spatial
transcriptomics analysis using Seurat (v4.0.1). Spots with less than 300
measured genes and less than 500 UMIs were filtered out. Ribosomal
and mitochondrial genes were excluded from this analysis. Individual
count matrices were normalized with sctransform79, and additional
log-normalized (size factor = 10,000) and scaled matrices were calcu-
lated for comparative analyses using default settings.

Cell-type compositions were calculated for each spot using cell2lo-
ication80 (v0.05). Reference expression signatures of major cell types
were estimated using regularized negative binomial regressions and
our integrated snRNA-seq atlas. We fitted a model in six downsampled
iterations of our snRNA-seq atlas (30%) and generated a final reference
matrix by taking the mean estimation. Each slide was later deconvoluted
using hierarchical bayesian models as implemented in run_cell2location.
We provided the following hyperparameters: 8 cells per spot, 4 factors
per spot, and 2 combinations per spot. Additionally, for each spot we
calculated cell-type proportions using the cell-type-specific abundance
estimations. Cell-type compositions of the complete slide were calcu-
lated adding the estimated number of cells of each type across all spots.
To compare the stability of estimated cell compositions between our
different data modalities, we calculated Spearman correlations between
the estimated cell type proportions of each slide and the observed cell
type proportions in its corresponding snRNA-seq and snATAC-seq
dataset.

**Estimation of functional information from spatial data.** For each
spot, we estimated signalling pathway activities with PROGENy81,82
(v1.12.0) model matrix using the top 1,000 genes of each transcriptional
footprint and the sctransform normalized data. Spatially vari-
able genes were calculated with SPARKX83 (v1.1.1) using log-normalized
data (FDR < 0.001). To obtain overrepresented biological processes
from each list of spatially variable genes, we performed hypergeometric
tests using the set of canonical pathways provided by MSigDB84
(FDR < 0.05).

**Estimation of cell death molecular footprints from spatial data.** To
associate the differences in nuclei capture in snRNA-seq between
the different samples to cell death processes, we leveraged the informa-
tion from spatial transcriptomics to estimate the general expression
of genes associated to cell death for each sample. For each unfiltered slide
we estimated per spot the normalized gene expression of BioCarta’s84
‘death pathway’ and Reactome’s87 ‘regulated necrosis pathway’ using
the decoupleR (v1.1.0) wmean method and the sctransform normalized
data. To have a final pathway score per slide, we calculated for each slide
the mean ‘pathway expression’ across all spots.

**Mapping transcription factor binding activity and GWAS enrich-
ment to spatial data.** To visualize the transcription factor binding ac-
tivity estimated from snATAC-seq data in space, we used the estimated
cell type proportion calculated from cell2location scores for mapping.
Specifically, for each spot i and transcription factor j, we calculated the
transcription factor binding activity as follows:

\[
\text{ACT}_{ij} = \sum_{k=1}^{K} \text{Proportion}_{ik} \times \text{ACT}'_{kj},
\]

where Proportion_{ik} is the estimated proportion of cell type k, K is the
number of cell types, and ACT’_{kj} is the binding activity of transcription
factor j in cell type k from snATAC-seq data. An equivalent approach
was used to map GWAS scores into space.

**Cell-state spatial mapping.** To map the functional states of each cell
type into spatial locations, we leveraged the deconvolution results of
each slide and the set of differentially expressed genes of each recov-
ered cell state. Given the continuous nature of cell states, we assumed
that the collection of up and downregulated genes of a cell state rep-
resented its transcriptional fingerprint and could be summarized in
a continuous score in locations where we could reliably identify the
major cell type from which the state was derived. For a given major
cell type of interest k, we identified spots where its inferred abundance
was of at least 10%. To estimate state scores associated with cell type k,
we used decoupleR’s (v1.1.0) normalized weighted mean method
(wmean) and the set of the upregulated genes of each state defined with
snRNA-seq and snATAC-seq (log fold change > 0; Wilcoxon tests,
FDR < 0.05). The log fold change of each selected gene was used as the
weight in the wmean function.

**Analysis of ion channel-related genes.** We related the expression
of ion channel-related gene sets to the different cardiomyocyte cell states
and their location in spatial transcriptomics. First we selected two dif-
gene sets containing ion channel-related genes: (1) Reactome’s81
‘ion channel transport’ and a curated list of transmural ion channels
from Grant et al.86. Gene sets are provided in Supplementary Table 17.
First, we calculated gene set scores for each spatial transcriptomics
spot using decoupleR’s wmean function. Then we correlated these
gene set scores to the spatial mapping of cardiomyocyte cell states
in regions where we observed at least 10% of cardiomyocytes. Addi-
tionally, we evaluated if any of the genes belonging to these gene sets
were differentially expressed between the vCMI and stressed vCM3
population using Wilcoxon tests as implemented in scran’s (v1.18.5) findMarkers function (area under the curve (AUC) < 0.4, AUC > 0.6, FDR < 0.05).

Spatial map of cell dependencies. We used MISTy’s \(^5\) implementation in mistyR (v1.2.1) to estimate the importance of the abundance of each major cell type in explaining the abundance of the other major cell types. Cell-type cell2location estimations of all slides were modelled in a multi-view model using three different spatial contexts: (1) an intrinsic view that measures the relationships between the deconvolution estimations within a spot, (2) a juxta view that sums the observed deconvolution estimations of immediate neighbours (largest distance threshold = 5), and (3) a para view that weights the deconvolution estimations of more distant neighbours of each cell type (effective radius = 15 spots).

The aggregated estimated standardized importances (median) of each view of all slides were interpreted as cell-type dependencies in different spatial contexts, such as colocalization or mutual exclusion. Nevertheless, the reported interactions did not imply any causal relation. Before aggregation, we excluded the importances of all predictors of target cell types whose \(R^2\) was less than 10\% for each slide.

To associate tissue structures with tissue functions, we fitted a MISTy model to explain the distribution of PROGENy’s pathway activities standardized scores. The multi-view model consisted of the following predictors: (1) an intrinsic view to model pathway crosstalk within a spot, (2) a juxta view to model pathway crosstalk between neighbouring spots (largest distance threshold = 5), (3) a para view estimating pathway relations in larger tissue structures (effective radius = 15), (4) an intrinsic view and (5) a para view containing cell2location estimations (effective radius = 15). These last two views model explicitly the relations between cell-type compositions of spots and pathway activities. Cycling cells and TNF were not included in the described analyses. Before aggregation, we excluded the importances of all predictors of target pathway activities whose \(R^2\) was less than 10\% for each slide.

Niche definitions from spatial transcriptomics data. To identify groups of spots in the different samples that shared similar cell-type compositions, we transformed the estimated cell-type proportions of each spatial transcriptomics spot and slide into isometric log ratios (ILR)\(^6\), and clustered spots into groups. These niches represent groups of spots that are similar in cell composition and represent potential shared structural building blocks of our different slides; we refer to these groups of spots as cell-type niches. Louvain clustering of spots was performed by first creating a shared nearest neighbour graph with \(k\) different number of neighbours (10, 20, 50) using scran’s \(^7\) (v1.18.5) buildSNNGraph function. Then, we estimated the clustering resolution that maximized the mean silhouette score of each cluster. We assigned overrepresented cell types in each structure by comparing the distribution of cell-type compositions within a cell-type niche versus the rest using Wilcoxon tests (FDR < 0.05). We tested if a given cell state was more representative of a cell-type niche by performing Wilcoxon tests between each niche and the rest (FDR < 0.05). Only positive state scores were considered in this analysis.

Additionally, to complement the repertoire of niches identified with cell-type compositions, we integrated and clustered the Visium spots of all slides using their log-normalized gene expression. We called these clusters molecular niches. Integration and clustering of spots was performed with the same methodology as the one used to create the snRNA-Seq atlas. A low resolution was used (0.2) to have a similar number of molecular niches as cell-type niches. Cell-type and cell-state enrichment was performed as mentioned before.

Differential expression analysis of molecular niches enriched with cardiomyocytes. Differential expression analysis between molecular niches enriched in cardiomyocytes (niche 0, niche 1, niche 3) was performed using the log-transformed expression of all spots belonging to a given niche. Wilcoxon tests were performed with scran’s \(^8\) (v1.18.5) findMarkers function. Genes with a summary AUC > 0.55 and FDR < 0.05 were considered upregulated genes.

Differential molecular profiles of the molecular niche 10 enriched with capillary endothelial cells. Differential expression analysis between ischaemic, fibrotic and myogenic-enriched spatial transcriptomic spots was performed with Wilcoxon tests as implemented in scran’s \(^9\) (v1.18.5) findMarkers function. To obtain overrepresented biological processes from upregulated genes, we performed hypergeometric tests using the set of hallmark pathways provided by MSigDB \(^10\). Normalized PROGENy’s pathway activities for each spot were calculated using decoupleR’s wsum method with 100 permutations on log-transformed data. Mean normalized pathway scores were calculated per slide and comparisons between groups were performed with Wilcoxon tests. Reported \(P\) values were adjusted for multiple testing using the Benjamini–Hochberg procedure.

General differences in tissue organization. We annotated the different spatial transcriptomic slides into three groups based on histological differences with the help of pathologists: myogenic-enriched, fibrotic-enriched and ischaemic-enriched. A general comparison of the sampled patient specimens was performed at the compositional and molecular level.

Hierarchical clustering, with euclidean distances and Ward’s algorithm, was used to cluster the pseudo-bulk profiles of the spatial transcriptomics datasets (replicates where merged, \(n = 27\)). Genes with less than 100 counts in 85\% of the sample size were excluded for this analysis. Log normalization (scale factor = 10,000) was performed. To visualize the general molecular differences between our samples, log-normalized pseudo-bulk profiles of the spatial transcriptomics datasets were projected in an UMAP embedding.

To identify compositional differences between our sample groups, we compared cell-type and niche compositions. To identify cell-type composition changes associated to the sample groups, mean cell-type compositions across single-cell and spatial datasets were compared with Kruskal–Wallis tests (FDR < 0.1). Pairwise comparisons of sample groups were performed with the Wilcoxon test. Additionally, to test which cell-type and molecular niches had different distributions between our sample groups, we performed Kruskal–Wallis tests over the compositions of cell-type or molecular niches (FDR < 0.1). Additional pairwise comparisons were performed with Wilcoxon tests (\(P\) values adjusted with Benjamini–Hochberg procedure). For this, we only consider slides where no single niche represents more than 80\% of the spots. Also, we only consider niches representing more than 1\% of the composition of at least 5 slides.

To identify differences between the structurally similar tissues captured in the myogenic-enriched group, we separated the samples into remote, border, and control zones and repeated the niche composition comparison described previously.

To identify patterns of tissue organization associated with a sample group, we tested if differential cell dependencies were captured by the MISTy models used to predict cell-type abundance (see ‘Spatial map of cell dependencies’). First, we filtered the standardized importance matrices of each sample’s MISTy model fitted to predict the abundance of major cell-types to contain only the values of target cell types predicted with an \(R^2\) greater than 0.05. Then, for each slide we created a spatial dependency vector where each element contains the importance of each possible pair of target and predicted cell types. Finally, we tested which cell interactions had higher importance in one of the sample groups compared to the rest using Wilcoxon tests (FDR < 0.25). To prioritize interactions, we only performed pairwise comparisons between sample groups for cell-type dependencies from which the maximum median importance across all groups was greater than 0.
Omnipath’s ligand–receptor database92 implemented in LIANA90. This framework that compiles the results of state-of-the-art cell communication inference methods, to infer ligand–receptor interactions. We next used LIANA (v0.0.9)90, a group (that is, myogenic, ischaemic and fibrotic) and combined the matrices from all sub-populations. We hypothesized that the distribution of specific cell states in the spatial transcriptomics slides could be modelled by the cell-type composition or cell-state presence of individual spots and their neighbourhood.

For a given collection of cell states of interest, we first defined regions of interest in every single slide as the collection of spots where the inferred abundance of the cell type from which the cell state was derived was at least 10%. These regions limit the target spots used in the MISTY model, however the whole slide is used to spatially contextualise the predictors. We used as predictors the abundances of cell types estimated with cell-2location or cell states scores. To account only for the effects of the activation of a cell state, the state scores of predictor cell states were masked to 0 whenever their score was lower than 0. In all models we included two classes of spatially contextualized predictive views: an intrinsic (intra) and a local neighbourhood view (para, effective radius = 5).

Specifically we fitted the following models to answer four questions:

1. What are the main cell types whose abundance within a spot or in the local neighbourhood predict the stressed vCM3?
   - vCM3 - intra(cell-type abundance) + para(cell-type abundance)

2. What are the main cell types whose abundance within a spot or in the local neighbourhood predict the endothelial subtypes? How do the different subtypes relate to each other?
   - ECsubtypes - intra(ECsubtypes) + para(ECsubtypes) + intra(cell-type abundance) + para(cell-type abundance)

3. What are the myeloid cells within a spot or in the local neighbourhood that better predict fibroblasts cell states? How do fibroblasts cell states relate to each other?
   - FibroblastStates - intra(FibroblastStates) + para(FibroblastStates) + intra(MyeloidStates) + para(MyeloidStates)

4. What are the main cell types whose abundance within a spot or in the local neighbourhood predict the myeloid cell states? How do the different states relate to each other?
   - MyeloidStates - intra(MyeloidStates) + para(MyeloidStates) + intra(cell-type abundance) + para(cell-type abundance)

Specific view importances were compared between patient groups as described previously with an $R^2$ filter of 0.1.

**Cell–cell communication analysis**

To estimate ligand–receptor interactions between the sub-populations of fibroblasts and myeloid cells, we extracted gene expression matrix from the integrated snRNA-seq and snATAC-seq data for each sample (that is, myogenic, ischaemic and fibrotic) and combined the matrices from all sub-populations. We next used LIANA (v0.0.9)90, a framework that compiles the results of state-of-the-art cell communication inference methods, to infer ligand–receptor interactions. We focused on the CellPhoneDB41 ligand–receptor method with Omnipath’s ligand–receptor database42 implemented in LIANA40. This was done by combining snRNA-seq samples of myogenic, ischaemic and fibrotic groups and subsetting only the fibroblasts and myeloid cells sub-states. Next, we used CrossTalker (v1.3.1)40 to find changes in cell–cell communication by contrasting ligand–receptor interactions predicted in myogenic vs. ischaemic samples and myogenic vs. fibrotic samples. The interactions considered by CrossTalker were obtained by filtering the output of LIANA40 ($P > 0.01$).

**Visualization, statistics, and reproducibility**

In data represented as box plots (Figs. 2f, 4c,d,m,o, 5b and 6d,n) the middle line corresponds to the median, the lower and upper hinges describe the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × inter-quartile range (IQR) from the hinge and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, and data beyond the end of the whiskers are outlying points that are plotted individually. In Figs. 4b and 5b,k, Colours refer to gene-weighted kernel density as estimated by using R package Nebulosa44. All reported $P$-values based on multi-comparison tests were corrected using the Benjamini–Hochberg method45. The depicted immunofluorescence micrographs are representative (Figs. 4c and 6n). The number of samples for each group was chosen on the basis of the expected levels of variation and consistency. The depicted RNAscope, immunofluorescence micrographs are representative and were performed at least twice, and all repeats were successful. Fig. 1a contains a panel from BioRender.com.

**Data availability**

Processed snRNA-seq, snATAC-seq, and spatial transcriptomics data are available at cellxgene https://cellxgene.ciszience.com/collections/8191e283-0816-4249-9a61-c31e6258a77 and at the Zenodo data archive (https://zenodo.org/record/6578047). Raw data generated by CellRanger and SpaceRanger pipelines are available through the Human Cell Atlas Data Portal at https://data.humancellatlas.org/explore/projects/ef936305-d857-44a3-9390-df4e6007dc97 and at the Zenodo data archive (https://zenodo.org/record/6578553, https://zenodo.org/record/6578617 and https://zenodo.org/record/6580069). Source data are provided with this paper.

**Code availability**

All code used for analysis is available at https://github.com/saezlab/visium_heart and https://github.com/KramannLab/visium_heart.

**Reporting summary**

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

**Acknowledgements**

We thank RECOVAM, for access to facilities to perform the experiments. Additional information on research design is available in the Nature Methods Reporting summary.

**Code availability**

All code used for analysis is available at https://github.com/saezlab/visium_heart and https://github.com/KramannLab/visium_heart.

**Conflict of interest**

J.L., L.D., C.S.T., D.R., O.E., A.B., C.R., R.B., M.G., I.G., S.A., J.K., T.M., L.K., E.N., L.A., and M.G.D. declare no conflict of interest. J.C., I.P., J.B., J.M., T.M., R.L., and J.B. declare no conflict of interest. J.L., L.D., C.S.T., D.R., O.E., A.B., C.R., R.B., M.G., I.G., S.A., J.K., T.M., L.K., E.N., L.A., and M.G.D. declare no conflict of interest.

**Further information**

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

**Code availability**

All code used for analysis is available at https://github.com/saezlab/visium_heart and https://github.com/KramannLab/visium_heart.

**Acknowledgements**

We thank RECOVAM, for access to facilities to perform the experiments. Additional information on research design is available in the Nature Methods Reporting summary.

**Conflict of interest**

J.L., L.D., C.S.T., D.R., O.E., A.B., C.R., R.B., M.G., I.G., S.A., J.K., T.M., L.K., E.N., L.A., and M.G.D. declare no conflict of interest. J.C., I.P., J.B., J.M., T.M., R.L., and J.B. declare no conflict of interest. J.L., L.D., C.S.T., D.R., O.E., A.B., C.R., R.B., M.G., I.G., S.A., J.K., T.M., L.K., E.N., L.A., and M.G.D. declare no conflict of interest.

**Further information**

Further information on research design is available in the Nature Methods Reporting summary linked to this article.
63. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq data with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

64. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

65. Granja, J. M. et al. ArchiFl is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

66. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

67. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

68. Li, Z. et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).

69. Fornes, O., A. L., & J. D. S. T. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Cellphone DB* **15**, 15 (2018).

70. Grant, A. O. Cardiac ion channels. *F1000Res.* **13**, 2017 (2017).

71. Schubert, M. et al. Perturbation-response genes reveal signaling footprints in cancer transcriptomics data, cell sub-state definitions and gene-regulatory networks. *R.R., M.H., J.T., J.N. and Z.L. carried out snATAC-seq data analysis, its integration with snRNA-seq and spatial transcriptomics data, cell sub-state definitions and gene-regulatory networks. R.R., M.H., J.T., J.N. and Z.L. carried out snATAC-seq data analysis, its integration with snRNA-seq and spatial transcriptomics data, cell sub-state definitions and gene-regulatory networks.

72. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

73. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Networks* **1**, 215–239 (1978).

74. Page, L. et al. The PageRank Citation Ranking: Bringing Order to the Web. (Stanford Holoab, 1999).

75. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

76. Kleschevnikov, V. e. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).

77. Schubert, M. et al. Perturbation-response genes reveal signaling footprints in cancer transcriptomics data, cell sub-state definitions and gene-regulatory networks. *R.R., M.H., J.T., J.N. and Z.L. carried out snATAC-seq data analysis, its integration with snRNA-seq and spatial transcriptomics data, cell sub-state definitions and gene-regulatory networks. R.R., M.H., J.T., J.N. and Z.L. carried out snATAC-seq data analysis, its integration with snRNA-seq and spatial transcriptomics data, cell sub-state definitions and gene-regulatory networks.

78. Holland, C. H., Szalay, A. B. & Saez-Rodriguez, J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochem. Biophys. Res. Commun.* **1863**, 194431 (2020).

79. Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol.* **22**, 184 (2021).

80. Liberez, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

81. Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2021).
Extended Data Fig. 1 | Computational workflow and snRNA-Seq data analysis. 

a, Schematic of the computational workflow for the main analyses of snRNA-seq, snATAC-seq, and spatial transcriptomics data. 
b, UMAP embedding of snRNA-seq data from all samples for patients, regions, and clusters. 
c, UMAP embedding of snRNA-seq data showing G2-M-phase cell-cycle score (left) and S-phase score (right). 
d, Barplots showing cell-type proportions of snRNA-seq data across all samples. Colours indicate different major cell types. 
e, Comparison of the generated snRNA-seq atlas to a previously published human heart cell atlas (HCA) for molecular profiles (left) and cell-type proportion (right). Left panel shows the adjusted $P$-value of the overlap of top gene markers of each cell-type between the different data sets (hypergeometric test). Right panel shows the Pearson correlation between the median proportion of each shared cell-type of the reference atlas and our control, border zone, and remote zone samples. 
f, UMAP embedding and annotation of an external dataset of three ischaemic zone samples following MI. 
g, Comparison of the generated snRNA-seq atlas to the external ischaemic data for molecular profiles (left) and cell-type proportion (right). Left panel shows the adjusted $P$-value of the overlap of top gene markers of each cell-type between the different data sets (hypergeometric test). Right panel shows the Pearson correlation between the median proportion of each shared cell-type of the external data set and our ischaemic samples.
Excised Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | snATAC-seq and spatial transcriptomics data analysis.

a, UMAP showing snATAC-seq data for all patients, regions, and clusters.
b, Validation of snATAC-seq cell type annotation using snRNA-seq data.
Left: UMAP showing the predicted labels for snATAC-seq data. Right: heatmap showing evaluation results by adjusted rand index (ARI).
c, snATAC-seq cell-type proportions of all samples.
d, Spearman correlations of cell-type compositions of samples estimated from snRNA-seq and snATAC-seq data. Box-whisker plots showing median and IQR (n = 3 for BZ, n = 4 for control, n = 4 for RZ, n = 5 for FZ, and n = 9 for IZ).
e, Cell type detection for snRNA-seq and snATAC-seq data across samples. The colour represents in which modality was cell type detected: snRNA-seq and snATAC-seq, snRNA-seq only, or not recovered.
f, Transcription factor (TF) binding and TF target expression per major cell type based on the snATAC-seq and snRNA-seq data.
g, Overrepresented spatially variable biological processes (myocardium related, immune related and fibrosis related) across regions. Each cell contains the mean adjusted P-value of a hypergeometric test across all spatial transcriptomics samples belonging to each region.
h, Spearman correlations of cell-type compositions of samples estimated from spatial transcriptomics and snRNA-seq (left) or snATAC-seq (right). Box-whisker plots showing median and IQR (n = 3 for BZ, n = 4 for control, n = 4 for RZ, n = 5 for FZ, and n = 9 for IZ).
i, GWAS SNP enrichment score across major cell types.
j, Visualization of GWAS SNP enrichment (left ventricular ejection fraction) in spatial transcriptomics data. In d, h, each spot is a patient sample (n = 3 for borderzone (BZ), n = 4 for controls, n = 6 for fibrotic zone (FZ), n = 9 for ischaemic zone (IZ), n = 5 for remote zone (RZ)). Data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually. In a-c the number of spots of the bottom panels correspond to the barplots in the upper panel.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Cell-type niches and spatially contextualized views analysis. **a-b**, UMAP embedding of spatial transcriptomics (ST) spots based on major cell-type compositions coloured by the patient (a) or region (b). **c**, Cell-type niche compositions across patient sample. **d**, UMAP embedding of ST spots based on major cell-type compositions colored by the compositions of cycling cells, pericytes, adipocytes, endothelial cells and myeloid cells and their respective marker genes (*Ryr2* cardiomyocytes, *Pdgfra* fibroblasts, *Mki67* cycling cells, *Abcc9* pericytes, *Fasn* adipocytes, *Vwf* endothelial cells, *Il7r* myeloid cells, and *Myh11* vSMCs). **e**, Standardized mean PROGENy pathway activities across different niches. **f**, H&E staining and cell2location cell-type abundance estimations of cardiomyocytes, fibroblasts and myeloid cells. Delineated areas represent myogenic or fibroblast/myeloid cell enriched tissue areas. **g**, Median standardized importances (> 0) of cell-type abundances in the prediction of other cell types within the immediate neighbourhood (upper part) and the extended neighbourhood (effective radius of 15 spots) (lower part) inferred from spatially contextualized models. Cell-type abundances of the immediate (upper panels) and extended neighbourhood of cardiomyocytes, fibroblasts and myeloid cells. **h**, Visualization examples of the dependencies between the abundance of endothelial cells and the abundance of pericytes in the immediate neighborhood, and vascular smooth muscle cells in the extended neighbourhood (effective radius of 15 spots) visualized on three tissues.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | Spatial pathway and cell type dependency analysis.

a, Median standardized importances (> 0) of cell-type abundances within the spot and in the extended neighbourhood (effective radius = 15), and PROGENy pathway activities within a spot, and in the immediate or extended neighbourhood on the prediction of pathway activities inferred from spatially contextualized models. b, Standardized importances of cardiomyocyte abundances in the extended neighbourhood (radius of 15 spots) in predicting hypoxia pathway activity. Comparison was performed between two sample groups: control, border and remote zones samples (n = 10), and fibrotic and ischaemic samples (n = 15). Two-sided Wilcoxon rank sum test. c–e, Visualization examples of pathway to pathway and pathway to cell-type dependencies in spatial transcriptomics. (c) p53 and PI3K spatial pathway distribution. (d) JAK-STAT and NFkB activities’ dependencies to myeloid and lymphoid cells’ abundance. (e) Cardiomyocyte to WNT and hypoxia pathway dependency. f, Hierarchical clustering of the pseudo-bulk transcriptional profile of spatial transcriptomics slides of patient samples. Colour represents the three major sample groups. g, Comparisons of patient mean major cell-type proportions (inferred from snRNA, snATAC-seq data, and spatial transcriptomics) between myogenic, ischaemic and fibrotic groups. Two-sided Wilcoxon rank sum test. Each dot represents a patient sample (n=13 for myogenic group, n=9 for ischaemic group, n=5 for fibrotic group). h, Pairwise comparison between patient groups of the standardized importances of cell-type abundances within the same spot (intra), and in the immediate (juxta) or extended neighborhood (para) to predict other cell types’ abundances (two-sided Wilcoxon rank sum test). Myogenic vs ischaemic (left), myogenic vs fibrotic (middle), fibrotic vs ischaemic (right). * = adj. P-value <= 0.15. i, Comparison of standardized importances of lymphoid cells’ abundances in the immediate neighborhood (juxta) to predict myeloid cells between myogenic, ischaemic, and fibrotic groups (two-sided Wilcoxon rank sum test). Visualization of lymphoid and myeloid cells in an ischaemic (up) and control (down) sample. j, Comparison of standardized importances of pericytes’ abundances to predict cardiomyocytes’ abundances within the same spot (two-sided Wilcoxon rank sum test). Each dot represents a sample (n = 14 for myogenic group, n=9 for ischaemic group, n = 5 for fibrotic group). Spatial distributions of pericyte and cardiomyocytes abundances estimated from deconvolution in a myogenic sample (left) and a fibrotic sample (right). Arrows in the fibrotic enriched sample show a strong colocalization event. k, Comparison of the compositions of cell-type niches between myogenic, fibrotic, and ischaemic groups (Kruskal-Wallis test, line denotes an adj. P-value < 0.1). Comparison of the proportions of cell-type niches 4, 5, 7, and 8 between groups (adj. P-value estimated from two-sided Wilcoxon rank sum test) and visualization example. Each dot represents a patient sample (n for myogenic group = 13, n for ischaemic group = 7, n for fibrotic group = 5). l, Comparison of the proportions of cell-type niche 9 between groups (two-sided Wilcoxon rank sum test). Each dot represents a patient sample (n for myogenic group = 13, n for ischaemic group = 7, n for fibrotic group = 5). m, Pairwise comparison of the compositions of the cell-type niches between control, border and remote zone samples (two-sided Wilcoxon rank sum test). * = reflect changes with adj. P-value < 0.1. n for CTRL = 4, n for RZ = 5, n for BZ = 3. In b, g, i, j, k, l data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | Characterization of molecular niches. a, UMAP embedding of ST spots based on gene expression coloured by patients (left) and regions (right). b, Gene expression of RYR2, ABCG9, MYH11, PDGERa, CD8A and IL7R. Colours refer to gene-weighted kernel density as estimated by using R package Nebulosa. c, Bar plots visualizing molecular niches proportion per patient. d, Standardized mean PROGENy pathway activities across different molecular niches. e, Visualization of molecular niches in control, IZ, and FZ samples. f, Comparison between patient groups of the proportions of molecular niches 5, 6, and 1 (adj. P value from a two-sided Wilcoxon rank sum test). Each dot represents a patient sample (n for myogenic group = 13, n for ischaemic group = 9, n for fibrotic group = 5). g, Comparison between control (CTRL), border zone (BZ) and remote zone (RZ) of the proportions of molecular niches 1, 2, 3, and 4 (adj. P value, two-sided Wilcoxon rank sum test, Box-Whisker plots showing median and IQR. Maximum and minimum values as described previously). Each dot represents a patient sample (n for myogenic group = 13, n for ischaemic group = 9, n for fibrotic group = 5). h, Top five differentially upregulated genes between spots belonging to molecular niches enriched with cardiomyocytes. Summary area under the curve (AUC) is used as a size effect of comparing the expression of one molecular niche against the rest. * = reflect FDR < 0.05 and summary AUC > 0.55 (n for mol. niche 0 = 30,058, n for mol. niche 1 = 19,958, n for mol. niche 3 = 7,360). Spatial distribution of the expression of MYLK3 in a control slide (upper) and a border zone (lower). i, Comparison between patient groups of the proportions of molecular niche 2 (adj. P value, two-sided Wilcoxon rank sum test, Box-Whisker plots showing median and IQR. Maximum and minimum values as described previously). Each dot represents a patient sample (n for myogenic group = 13, n for ischaemic group = 9, n for fibrotic group = 5). j, Same as (i) for niche 3. In f, g, i, j data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually.
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Characterization of cardiomyocyte state clusters.

a, UMAP embedding of the integrated snRNA-seq and snATAC-seq data coloured by patients, regions, modalities, and clusters (resolution= 1). b, Dot plot showing the top 10 upregulated genes for each CM-state. c, State-specific pseudo bulk ATAC-seq profiles showing distinct chromatin accessibility for NPPB between different states. d, In situ RNA hybridization of NPPB, ANKRDI and TNNT2 on human cardiac control tissue (upper panel). Note that only a single NPPB/ANKRD1 positive cardiomyocyte could be detected. (below) Representative image of the same in-situ staining of a human myocardial infarction sample. Note the NPPB and ANKRDI expression in several cardiomyocytes (arrows). Scale bars: 25 µm (upper) + 50 µm (lower).

e, Spearman correlation between the spatial expression of Ion-channel transport and transmural-ion channels related genes to the expression of the marker genes (state scores) of vCM1-3 (two-sided Wilcoxon rank sum test). Each dot represents a visium slide (n = 28). f, Differential gene expression of ion channel related genes in vCM1 contrasted to vCM3 cardiomyocytes (two-sided Wilcoxon rank sum test, area under the curve (AUC) is shown as size effect). Colours refer to the gene set membership of each gene.

g, Visualization of vCM1 gene marker expression mapping (state scores) and the expression of RYR2 and ATP2B4 in a border zone sample. h, Comparison of vCM1 and vCM2 cell proportion between patient groups. P-values were calculated using Wilcoxon Rank Sum test (unpaired, two-sided). Each dot represents a sample (n = 13 for myogenic group, n = 7 for ischaemic group, and n = 4 for fibrotic group).

i, Distribution of vCM3 marker gene expression (state-score) across molecular niche 1, 2, and 4. Each dot represents a spatial transcriptomics spot belonging to a molecular niche across samples (n = 30,058 for niche 1, n = 19,958 for niche 2, niche 4 = 7,360). Two-sided Wilcoxon rank sum test, adj. p-values (niche 1 vs niche 4 = 0, niche 2 vs niche 4 = 2e-09, niche 1 vs niche 2 = 0). j, Visualisation of the expression of ANKRDI and NPPB, the spatial distribution of TGFβ, hypoxia, p53 and PI3K signaling activities, the spatial distribution of the expression of vCM-states marker genes (state score) for a border zone sample and the distribution of molecular niches 1, 2, and 4. Box-Whisker plots showing median and IQR. Maximum and minimum values as described previously. In e, h, i data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually. In a-c the number of spots of the bottom panels correspond to the barplots in the upper panel.
Extended Data Fig. 7 | See next page for caption.
Extended Data Fig. 7 | Cardiomyocyte pseudotime and gene-regulatory network analysis. 

a, Diffusion map embedding of pseudotime cardiomyocyte clusters vCM1, vCM2 and vCM3. 
b, Computational workflow for building gene regulatory network (upper) and schematic of linking TF to target genes through peak-to-gene links and predicted TF binding sites (lower). 
c, Heatmap showing TF binding activity and gene expression along the pseudotime trajectory. 
d, Heatmap of gene expression across pseudotime vCM1–vCM3. 
e, Heatmap showing the correlation between TF binding activity and gene expression along the pseudotime trajectory from vCM1 to vCM3. Colors on the top refer to pseudotime point where the TF showed the highest binding activity. 
Clustering analysis identified three modules.
f, ANKRD1 and NPPB CM-stress marker gene expression along pseudotime vCM1–vCM3. 
g, Peak-to-gene links showing that JDP2 regulates TGFB2. Each loop represents a putative link between TGFB2 and a peak. Loop height represents the significance of the correlation and dash line represents threshold of significance ($P = 0.05$).

ATAC-seq tracks were generated from pseudo-bulk chromatin profiles of vCM1, vCM2, and vCM3. Binding sites of JDP2 are highlighted. 
h, Line plots showing TF activity and expression after z-score normalization (y-axis) over pseudotime (x-axis) for JDP2 (left), its corresponding target gene expression after z-score normalization (y-axis) over pseudotime (x-axis) for TGFB2 (middle), and visualization of all target genes in the BZ sample (right). 
i, Comparison of standardized importances of myeloid abundances within the spot and fibroblast abundances in the local neighbourhood (radius of 5 spots) to predict vCM3 between patient groups samples. Each dot represents a sample (n = 9 for myogenic group, n = 7 for ischaemic group, n = 4 for fibrotic group). 
(lower panel) Spatial distribution of the state score of vCM3 and myeloid cell abundances in a RZ slide with histological evidence of a scar. 
(Two-sided Wilcoxon rank sum test, Box-Whisker plots showing median and IQR. Maximum and minimum values as described previously.) 
Data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually. 
In a–c the number of spots of the bottom panels correspond to the barplots in the upper panel.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Endothelial cell heterogeneity. a. UMAP embedding of integrated snRNA-seq and snATAC-seq data coloured by patients, regions, modalities and clusters (resolution= 0.9). b. Gene expression of NRG3, FLT4, SEMA3G and ACKR1. Colours refer to gene-weighted kernel density as estimated by using R package Nebulosa. c. Sub-cluster-specific pseudo bulk ATAC-seq tracks showing the chromatin accessibility of the genes from (b). d. Marker dot plot showing the DEGs for each endothelial cells state and subtype. e. Gene expression of SEM3G and POSTN. Colors refer to gene-weighted kernel density as estimated by using R package Nebulosa. Immunofluorescence of SEMA3G and ACTA 2 (upper image). In situ mRNA (RNAscope) for POSTN and PECAM1 (magenta). Arrows highlight endocardial endothelial cells. Scale bars: 75 µm (upper) and 10 µm (lower). f. Endothelial cell state compositions across all patient samples. g. Comparison between patient groups of the venous endothelial cell proportion (two-sided Wilcoxon rank sum test). Each dot represents a patient sample (n for myogenic group = 13, n for ischaemic group = 9, n for fibrotic group = 5). h. Mean Pearson correlation between the abundance of major cell-types and endothelial cell-state scores across all spatial transcriptomics slides. i. Visualization of the spatial distribution of the abundances of pericytes and the state score of capillary endothelial cells sample. Arrows point at colocalization events. j. Visualization of the spatial distribution of molecular niches, the state score of capillary endothelial cells and the abundance of pericytes and cardiomyocytes. Arrows point at locations where molecular niche 10 is present together with high abundances of cardiomyocytes and pericytes. k. Endothelial cell state score distributions in all spots belonging to the molecular niche 10 across all slides (n = 1,874, P-value = 3.19e-298, obtained from a two sided Wilcoxon signed-rank test). l. Mean signalling pathway activities in the molecular niche 10 across different patient groups (adj. P-value of two-sided Wilcoxon Rank Sum test). Each dot represents a slide (n = 14 for myogenic group, n = 9 for ischaemic group, n = 5 for fibrotic group). m. Overrepresented hallmark pathways in the differentially upregulated genes of molecular niche 10 across different patient groups (hypergeometric tests, adj. P-values). In g, k, l data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the interquartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually. In a–c the number of spots of the bottom panels correspond to the barplots in the upper panel.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Fibroblast heterogeneity and PDGFRβ lineage tracing in murine MI. **a**, UMAP embedding of integrated snRNA-seq and snATAC-seq data coloured by patients, regions, modalities and clusters (resolution = 0.9). **b**, Marker dot plot showing the DEGs for each fibroblast state. **c**, Box plots showing module score of NABA-collagen and NABA-core-matrisome score per fibroblast state. *P*-values were calculated using Wilcoxon rank sum (unpaired and two sided). **d**, Expression of COL1A1, ACTA2 and FN1 by RNA qPCR after RUNX1 overexpression with and without TGFβ compared to empty vector (EV) (n = 6). One-way ANOVA followed by Bonferroni correction. Error bars = S.D. **e**, In situ hybridization (RNAscope) on human myocardial tissue of SCARA5 and COL15a1 and quantification and comparison of SCARA5+/POSTN+ cells vs. POSTN+/COL1A1+ cells in human heart failure tissues (n = 7). Mann-Whitney test. Error bars = S.D. **f**, Visualization of SCARA5, POSTN, COL1A1 and FN1 on spatial transcriptomics slides from IZ and FZ human MI samples. **g**, Cell-type state compositions across patient sample. **h**, Comparison of Fib3 cell proportion between patient groups. *P*-values were calculated using Wilcoxon rank sum test (unpaired, two-sided). Each dot represents a sample (n = 13 for myogenic group, n = 7 for ischaemic group, and n = 4 for fibrotic group). **i**, Time course of lineage tracing experiment using PDGFRβCreER-tdTomato mice. **j**, Results of echocardiographic measurements EF (in %), and global longitudinal strain (GLS, in %), for each of the time points. One-way-ANOVA. **k**, Quality measurements of single-cell RNA-Seq data from mouse MI experiment. **l**, UMAP representation of time points, cluster and gene expression (POSTN, SCARA5, COL1A1 and FN1) from mouse MI experiment. **m**, Confusion matrix comparing predicted fibroblasts states and obtained clusters for mouse dataset. **n**, UMAP representation of species cross-annotation of fibroblast clusters. **o**, Cell proportion of mouse fibroblast state Fib1 (SCARA5+) and Fib2 (POSTN+) per MI time-point. **p**, Box plots showing NABA collagens scores and NABA core matrisome score per fibroblast state (mouse Fib1 vs. Fib1) per time point. *P*-values were calculated using Wilcoxon rank sum (unpaired and two sided) (Sham: n = 578 for Fib1 and n = 685 for Fib2; Day 4: n = 1688 for Fib1 and n = 2498 for Fib2; Day 7: n = 203 for Fib1 and n = 330 for Fib2; Day 14: n = 2232 for Fib1 and n = 7684 for Fib2). **q**, Gene set enrichment analysis per human fibroblast-cell state. Scale bar: 10 µm. In c, d, k, p data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually.
Extended Data Fig. 10 | Gene regulatory network analysis of human cardiac fibroblasts. **a**, Pseudotime heatmap showing gene expression (left) and TF binding activity (right) along the trajectory Fib1 to Fib2 (myofibroblasts). **b**, Pseudotime heatmap showing highly variable genes along the trajectory. **c**, Heatmap showing the correlation between TF binding activity and gene expression each TF from (a) and gene from (b). Each column represents a TF and each row represents a TF. Colours on the top refer to pseudotime labels for each TF estimated based on binding activity. **d**, Upper: line plots showing TF activity and expression after z-score normalization (y-axis) over pseudotime (x-axis) for TGFB1. Lower: visualization of target gene expression of KLF4 and GLI2, and TGFB pathway activity in the BZ sample. **e**, Upper: line plots showing TF activity and expression after z-score normalization (y-axis) over pseudotime (x-axis) for RUNX2 and its corresponding target gene expression after z-score normalization (y-axis) over pseudotime (x-axis) for COL1A1. Visualization of RUNX2 target genes in the BZ sample (right).
Extended Data Fig. 11 | Myeloid cell heterogeneity and spatial mapping in human myocardial infarction. a, UMAP embedding of integrated snRNA-seq and snATAC-seq data coloured by patient contribution, region, modality and clusters (resolution = 1). b, Marker dot plot showing the DEGs for each fibroblast state. c, Gene expression of MRC1, ITGAX, FLT3 and CD163. Colors refer to gene-weighted kernel density as estimated by using R package Nebulosa. d, Bar plots visualizing myeloid cell sub-population proportion per patient. e, Gene expression of MRC1, ITGAX, SPP1, CCL18, FLT3, ZBTB46 in an external dataset of ischaemic patients. Colors refer to gene-weighted kernel density as estimated by using R package Nebulosa. f, Comparison of myeloid cell proportion between patient groups. P-values were calculated using Wilcoxon Rank Sum test (unpaired, two-sided) (n = 13 for myogenic, n = 8 for ischaemic, and n = 5 for fibrotic group). g, In situ hybridization of CCR2, SPP1 and TREM2 on human myocardial infarction section IZ. Arrows point to phagocytotic vacuoles in the magnification. h, In situ hybridization of SPP1, POSTN and CD163 on human myocardial infarction section IZ. Note that SPP1+ macrophages colocalized in distinct tissue regions and appeared to have an enlarged cell-size. i, In situ hybridization quantification of cell type proportion per tissue section. n = 7 patient tissues. Adjusted P-values were calculated using Wilcoxon signed-rank test. j, Median standardized importances (>0) of the cell-state scores of myeloid cells within the spot in the prediction of other myeloid cell-state scores in spatial transcriptomics. k, Mean pearson correlation between myeloid cell’s state scores across all spatial transcriptomics slides. Data in f, i are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually.
Extended Data Fig. 12 | See next page for caption.
Extended Data Fig. 12 | Myeloid cell spatial modelling and cell-cell communication. a, Median standardized importances (>0) of the cell-state scores of myeloid cells (colocalization) in the prediction of fibroblast cell-state scores in spatial transcriptomics. b, Spatial mapping of cell-type niches in a control and an ischaemic sample. Arrows in the IZ sample show niche 4 and how it surrounds niche 5. c, Distribution of myofibroblasts and SPP1+ macrophages marker gene expression (state-score) across cell-type niche 4 and 5. Each dot represents a spatial transcriptomics spot belonging to a molecular niche across samples (n = 12,427 for niche 4, n = 4,375 for niche 5) (two-sided Wilcoxon rank sum test, adj P-value = 6.02e-133 for myofibroblast, adj P-value = 0 for SPP1+ monocytes). Data are represented as boxplots where the middle line is the median, the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends from the hinge to the largest value no further than 1.5 × IQR from the hinge (where IQR is the inter-quartile range) and the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge, while data beyond the end of the whiskers are outlying points that are plotted individually. d, Cell-cell communication network representing number of ligand-receptor interactions (edge richness), expression of ligand-receptor pairs (LR scores; colour gradients) and cell centrality (Pagerank; node size) as estimated in snRNA-seq of myogenic, ischaemic and fibrotic samples. e, Sankey plots summarizing the top 50 ligand-receptor interactions for selected source and target cells and contrast. These ligand-receptor pairs are selected by absolute value of the difference in LR score, as provided by CellPhoneDB method implemented in Liana. f, Sankey plot summarizing top 50 TGFbeta ligand-receptor interactions from Fib2 to SPP1+ Mac. cells. g, In situ hybridization mRNA (RNAscope) staining of CD163 (myeloid), POSTN (myofibroblast) and SPP1 on human cardiac MI tissue (crop-out in Fig. 6n). Arrows indicate CD163+SPP1+ macrophages near myofibroblasts. Scale: 25 µm.
Reporting Summary

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

Statistics

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

☐ n/a  Confirmed

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ The statistical test(s) used AND whether they are one- or two-sided

☐ Only common tests should be described solely by name; describe more complex techniques in the methods section.

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) and variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

☐ Give P values as exact values whenever possible.

☐ 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

Software used include: Sony FACS software (Version 2.1.5), Vevo 3100 (Version 5.5.1) NIS-Elements AR 5.11.03, Bio-Rad CFX96 Real Time System (Bio-Rad), Chromium Controller Firmware version XYZ, NanoZoomer HT (Hamatsu Photonics, Hamamatsu, Japan), Graphpad (Version 8.4.3), Flowl (10.6.1), Adobe Photoshop (21.2.4), Adobe Illustrator (24.3), ImageJ (1.53)

Data analysis

All software/scripts used in data analysis is described here: https://github.com/saezlab/vision_heart.

Software used include: python (3.7.8), R (v4.0.2, v4.1.0), CellRanger software (v6.0.2), CellRanger ATAC pipeline [v2.2.0], SpaceRanger software (v1.3.2), Seurat (v4.0.1), scDbFinder (v1.4.0), edgeR (v3.32.1), scmapy (v1.7.0), ArchR (v1.0.1), MACS2 (v2.2.7), HINT-ATAC (v0.13.2), decoupler (v1.1.0), Plink (v1.9), gchromVAR (v0.3.2), chromVAR (v1.16), cell2location (v0.05), PROGENy (v1.12.0), SPARXX (v1.1.1), mistyR (v2.2.1), scran (v1.19.5), Mesmer (v0.3.0), RS-FISH (v2.3.1)

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/reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

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

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

The data is made available through the Human Cell Atlas (HCA) Data Coordination Platform (DCP) and can be accessed here: https://data.humancellatlas.org/explore/projects/e9f36305-d857-4aa3-930d-0f4e6007dc97. Additionally, the data can be explored through the interactive data explorer cellxgene (czisciencee) here: https://cellxgene.cziscience.com/collections/8191c283-0816-424b-9b61-c3e1d6258a77/private.
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-list.pdf

Life sciences study design

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

**Sample size**
For an overview on sample please see supplementary table 1. No statistical methods were used to predetermine sample size. The patients were selected to give an equal representation of cells from healthy and myocardial infarction tissue from different zones. RT-qPCR and histological analysis were performed on five independent replicates with a minimum of 3 (n=shown in the figure and described in legends).

**Data exclusions**
For full details see description in Methods. For single-cell RNA-seq experiments, low-quality droplets were excluded based on standard, pre-established approaches: (1) transcript count, (2) doublet score, (3) mitochondrial transcript content and (4) gene recovery. For spatial transcriptomics experiments, spots were excluded based on their (1) transcript count and (2) gene recovery. For single-cell ATAC-seq experiments, low-quality droplets were excluded based on: (1) number of unique fragments; (2) transcription start site (TSS) enrichment.

**Replication**
The reported findings were replicated across multiple biological samples (n reported in each figure or Methods). Immunofluorescent imaging was performed on 3 heart samples minimum.

**Randomization**
Randomisation was not relevant due to the study design where human control and acute myocardial infarction tissues were used on availability.

**Blinding**
Blinding of the tissue was not possible. All analyses were performed in an automated manner across conditions.

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 | Methods |
|----------------------------------|---------|
| n/a | n/a |
| ☑ Antibodies | ☑ ChIP-seq |
| ☑ Eukaryotic cell lines | ☑ Flow cytometry |
| ☑ Palaeontology | ☑ MRI-based neuroimaging |
| ☑ Animals and other organisms | |
| ☑ Human research participants | |
| ☑ Clinical data | |

**Antibodies**

| Antibodies used |
|-----------------|
| Antibody, Company, Catalog number, Clone, Dilution: |
| anti-ACTA2(aSMα)-Cy3 (Sigma-Aldrich, C6198, 1:250) |
| anti-SEMA3G (Sigma Aldrich, HPA001761, 1:100) |
| anti-PDGFRb, R&D, MA51263, PR7212, BQF0715561, 1:100 |
| A647 anti-rabbit Jackson ImmunolResearch, 711-605-152, 1:200 |
| anti-mouse IgG1-MicroBeads solution, Miltenyi, 130-047-102 |

**Validation**

All antibodies used in this study are commercially available. They are validated by the vendors for the specific assay and species used. The validation is available on the vendors website. Primary antibodies for immunostaining were validated by performing comparisons to species-matched isotype antibodies and unstained controls. For the primary flow antibodies the specificity was validated by staining directly against species-matched isotype and unstained controls.
Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)  The immortalized human PDGFRb+ heart cell line from heart failure patients was generated as described in the methods.

Authentication  None of the cell lines used were authenticated.

Mycoplasma contamination  Mycoplasma tests showed no evidence of contamination by PCR.

Commonly misidentified lines (See ITAC register)  No commonly misidentified cell lines were used.

Animals and other organisms

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

Laboratory animals  This study involved transgenic mice (Mus musculus). Used strains were PDGFRBCreERT2 (i.e. B6-Cg-Gt(1PdgfrB-LacZ-Tg(ROSA26Sortm1(CAG-Cat))/JAX Stock #029684), Rosa26tdTomato (i.e. B6-Cg-Gt(Rosa26Sortm1(CAG-tdTomato)LacZf/JAX Stock #007909). PDGFRBCreERT2tdTomato mice were males and females with an age of 8 weeks. Mice were housed two to five animals per cage with a 12-h light-dark cycle (lights on from 0700 to 1900 h) at sustained temperature (20 °C±0.5°C) and humidity (~50%±10%) with ad libitum access to food and water.

Wild animals  This study did not involve any wild animals.

Field-collected samples  This study did not involve any samples collected from the field.

Ethics oversight  LANUV (Landesamt für Natur, Umwelt und Verbraucherschutz) North Rhine-Westphalia (Germany) approved the study protocol with reference No. 81-02.04.2017.A410.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics  Please see Extended Data table 1 for the clinical characteristics of the patients used in this study for snRNA, snATAC and spatial transcriptomics. Extended Data table 5 described cohort of patients used for in-situ studies. Ethnicity was caucasian.

Recruitment  Patients were recruited as described in the Methods. All tissue samples used for this study were obtained with written informed consent from all patients in accordance with the guidelines of The Declaration of Helsinki 2000.

Ethics oversight  The local ethics committee of the Ruhr University Bochum in Bad Oeynhausen, the RWTH Aachen University, Utrecht University and WUSTL approved all human tissue protocols (No. 220-640, EK151/09, 12/387, No. 201104172 respectively).

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration  n/a

Study protocol  n/a

Data collection  n/a

Outcomes  n/a
Flow Cytometry

Plots

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

| Sample preparation | FACS protocol and flow cytometry and detailed sample preparation are described in the Methods. |
|--------------------|--------------------------------------------------------------------------------------------------|
| Instrument         | SH800 Sorter (SONY)                                                                               |
| Software           | SONY Sorter software.                                                                            |
| Cell population abundance | Post-Sort purity was confirmed over 95%.                                                       |
| Gating strategy    | Please see Methods and Fig. 1b. for gating on DAPI+ nuclei.                                      |

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