



OPEN

Origin and function of activated fibroblast states during zebrafish heart regeneration

Bo Hu^{1,7}, Sara Lelek^{2,3,7}, Bastiaan Spanjaard^{1,7}, Hadil El-Sammak^{4,5}, Mariana Guedes Simões², Janita Mintcheva¹, Hananeh Aliee⁶, Ronny Schäfer¹, Alexander M. Meyer², Fabian Theis⁶, Didier Y. R. Stainier^{4,5}, Daniela Panáková^{2,3} and Jan Philipp Junker^{1,3}

The adult zebrafish heart has a high capacity for regeneration following injury. However, the composition of the regenerative niche has remained largely elusive. Here, we dissected the diversity of activated cell states in the regenerating zebrafish heart based on single-cell transcriptomics and spatiotemporal analysis. We observed the emergence of several transient cell states with fibroblast characteristics following injury, and we outlined the proregenerative function of collagen-12-expressing fibroblasts. To understand the cascade of events leading to heart regeneration, we determined the origin of these cell states by high-throughput lineage tracing. We found that activated fibroblasts were derived from two separate sources: the epicardium and the endocardium. Mechanistically, we determined Wnt signalling as a regulator of the endocardial fibroblast response. In summary, our work identifies specialized activated fibroblast cell states that contribute to heart regeneration, thereby opening up possible approaches to modulating the regenerative capacity of the vertebrate heart.

Heart injury in adult mammals typically leads to permanent scarring. However, the adult zebrafish heart regenerates efficiently after injury, making zebrafish the preeminent model system for study of the cellular and molecular mechanisms of heart regeneration^{1,2}. After cryoinjury, which mimics aspects of myocardial infarction, the injured zebrafish heart undergoes a transient period of fibrosis³, during which the damaged heart muscle regenerates via dedifferentiation and proliferation of cardiomyocytes^{4,5}. This process is accompanied by hypoxia-induced revascularization of the newly formed heart muscle tissue⁶. Many pathways and factors involved in heart regeneration have been described^{7–11}. Important molecular signals for zebrafish heart regeneration emanate from the epicardium and endocardium^{3,11–14}. Other cell types are required for regeneration: inflammation is an early response to cryoinjury, and depletion of macrophages leads to delayed regeneration^{15,16}. Moreover, fibroblasts (potentially of different origins) are activated after injury, and ablation of fibroblasts leads to reduced cardiomyocyte proliferation¹⁷. Altogether, these data suggest that cell types or transient activated cell states residing in the regenerative niche may be important cellular regulators of regeneration. A number of recent single-cell RNA sequencing (scRNA-seq) studies have determined cell type diversity in the mammalian heart^{18–21}, as well as in the regenerating zebrafish heart^{22,23}. However, to our knowledge, there has been no systematic data-driven attempt to identify proregenerative cell states and their cell type of origin. Consequently, our knowledge of the cellular composition of the regenerative niche and the underlying signalling interactions remains incomplete. Current definitions of activated macrophages and fibroblasts rely heavily on transgenes, may be affected by observation bias and probably underestimate the complexity of cell states involved in regeneration.

Results

The cellular composition of the regenerating heart. For systematic identification of cardiac cell types in the healthy and regenerating zebrafish heart, we performed scRNA-seq of around 200,000 dissociated cells at different stages before and after injury (Fig. 1a). To limit experimental biases, we did not apply any sorting procedure. To include information about the developmental origin of cells, we applied a method for massively parallel lineage tracing based on CRISPR–Cas9 technology^{24–26}. By injecting Cas9 and a single-guide RNA (sgRNA) against a multicopy transgene (*dTomato* in the zebrafish²⁷ line), we recorded lineage relationships in early development by creating ‘genetic scars’ that serve as lineage barcodes²⁴ (Methods).

We first assessed cell type diversity in the healthy and regenerating heart. Clustering of single-cell transcriptomes revealed all major cardiac cell types (Fig. 1b, Supplementary Figs. 1 and 2 and Supplementary Data Files 1 and 2). As expected, we observed a strong increase in fibroblasts and immune cells after injury (Fig. 1b). Closer inspection of the clustering data revealed a substructure among the cell types of the three main layers of the heart: epicardium, myocardium and endocardium (Supplementary Fig. 3a–c and Supplementary Data File 2). We hypothesized that this cell type substructure might correspond to spatial differences due to the functional specifications of these cell types in the atrium and ventricle. Using the tomo-seq method for spatially resolved transcriptomics²⁸, and deconvolving the spatial data into single-cell transcriptional profiles²⁹, we could validate atrial and ventricular enrichment for some of these cell subtypes (Fig. 1c, Supplementary Fig. 3a–c and Methods). We confirmed this finding by physical separation of the atrium and ventricle, followed by scRNA-seq (Fig. 1d).

¹Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Berlin, Germany. ²Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. ³DZHK (German Centre for Cardiovascular Research) partner site, Berlin, Germany. ⁴Department of Developmental Genetics, Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany. ⁵DZHK (German Centre for Cardiovascular Research) partner site Rhine/Main, Frankfurt, Germany. ⁶Helmholtz Center Munich – German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Munich, Germany. ⁷These authors contributed equally: Bo Hu, Sara Lelek, Bastiaan Spanjaard. ✉e-mail: daniela.panakova@mdc-berlin.de; janphilipp.junker@mdc-berlin.de

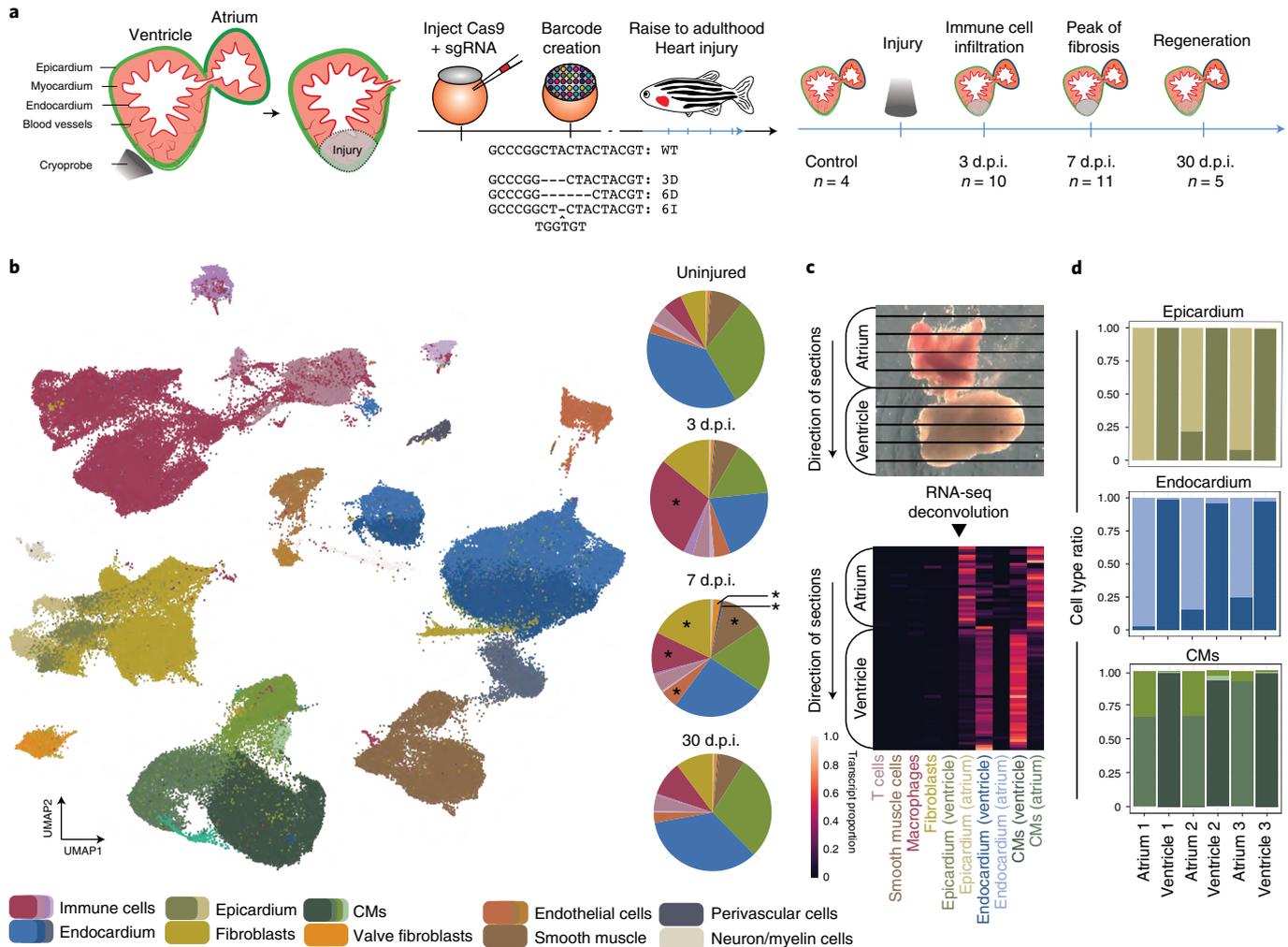


Fig. 1 | The cellular composition of the regenerating heart. **a**, Cartoon of the experimental approach. Cells were barcoded with indels (insertions or deletions) during early development, and fish were raised to adulthood. Hearts were harvested either as an uninjured control or at 3, 7 or 30 d.p.i. **b**, UMAP representation of single-cell RNA-seq data and clustering results. Pie charts show the proportions of different cell types at different time points after injury. In the pie chart representation, similar cell types are grouped and shown by one (representative) colour. Asterisks denote cell types with a statistically significant change in proportions compared with uninjured controls. **c**, Mapping of single-cell data onto a spatially resolved tomoseq data set. A computational deconvolution approach revealed chamber-specific cell subtypes. **d**, Distribution of subtypes of cardiomyocytes (CMs), endocardial cells and epicardial cells for scRNA-seq data sets in which atrium and ventricle were physically separated. Colour scheme as in Supplementary Fig. 3.

Cell type diversity of cardiac fibroblasts. We identified a further transcriptional substructure among the cardiomyocytes (Fig. 2a). In addition to adult cardiomyocytes, which are characterized by expression of genes involved in ATP synthesis and the tricarboxylic acid cycle (*atp5pd* and *aldoaa*), we detected a smaller cluster characterized by genes associated with cardiomyocyte development (*ttn.1*, *ttn.2*, *bves* and *synpo2lb*), as well as *nppa*, which has previously been shown to be a marker for dedifferentiated cardiomyocytes in the border zone²². These dedifferentiated cardiomyocytes, which are a hallmark of the regenerating heart, had increased in number by 3 d.p.i. (days postinjury) (Fig. 2a) and partially colocalized with the established marker gene *nppa*²² at 7 d.p.i. (Supplementary Fig. 3d).

We noticed that three well-established signalling factors in heart regeneration were strongly enriched in fibroblasts: *aldh1a2* (ref. 7), which is the enzyme synthesizing retinoic acid; the cardiomyocyte mitogen *nrg1* (ref. 30); and the proregenerative extracellular matrix (ECM) factor *fn1a*³¹ (Fig. 2b). This prompted us to investigate the diversity of cardiac fibroblasts in more detail. Subclustering revealed an unexpectedly large degree of diversity, with 13 transcriptionally

distinct clusters of fibroblasts (Fig. 2c, Supplementary Fig. 4a,b and Supplementary Data File 3). These 13 clusters exhibited pronounced differences in their expression profiles of ECM-related genes (Fig. 2d), but their transcriptomic diversity extended far beyond that – the same fibroblast subtypes could be identified with high accuracy after removal of ECM-related genes (Supplementary Fig. 5).

Identification of proregenerative cardiac fibroblasts. To focus our analysis on those fibroblast subtypes that may be part of the regenerative niche, we analysed the dynamics of the cell clusters after injury. Three clusters of fibroblasts, characterized by expression of *coll1a1a*, *coll2a1a* and *nppc*, respectively, were transiently present at the peak of regeneration (3 and 7 d.p.i.) but virtually absent before injury and after regeneration (Fig. 3a, Supplementary Fig. 4c and Supplementary Data File 4). Owing to their transient nature, we refer to these three clusters as cell states instead of cell types. Notably, *coll2a1a*, a nonfibrillar collagen that may act as a matrix-bridging component, has already been shown to be expressed in epicardial and connective tissues following heart injury³² and is known to be

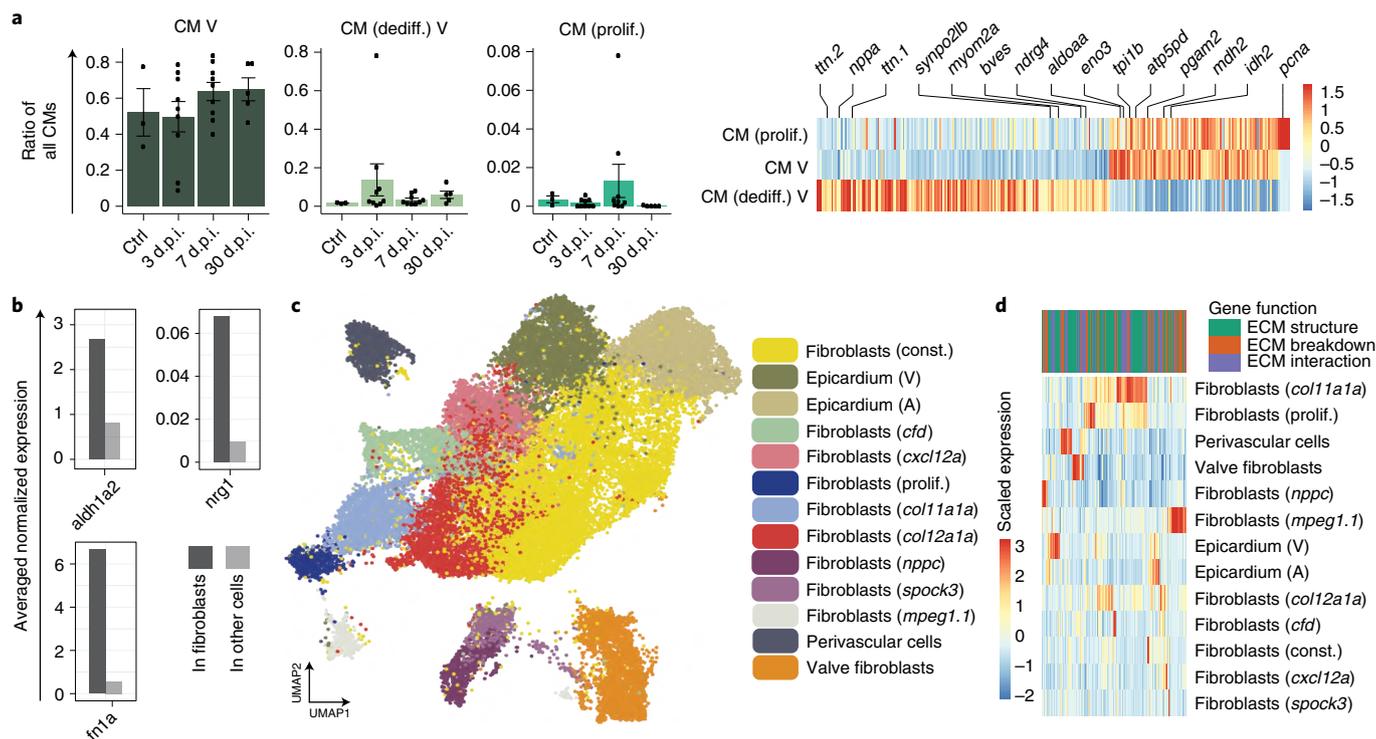


Fig. 2 | Cell type diversity of cardiac fibroblasts. **a**, Left: relative changes in abundance of different subtypes of CMs across the time points ($n=3, 9, 9, 5$ animals; error bars show s.e.m.). Right: differentially expressed genes between subtypes of CMs. **b**, Comparison of average normalized expression of known proregenerative factors in fibroblasts and other cell types. The comparison used data pooled over all time points, and fibroblasts were defined as all cells in the yellow cluster in Fig. 1b. **c**, UMAP representation of the subclustering of *col1a1a*-expressing cells. **d**, Expression of ECM-related genes in different fibroblast cell types. The genes were classified according to their contribution to structure, breakdown or interaction of the ECM. A, atrium; Ctrl., control; const., constitutive; dediff., dedifferentiated; prolif., proliferating; V, ventricle.

involved in regeneration of other organ systems in zebrafish³³. Here, *col12a1a* was not only expressed in *col12a1a* fibroblasts but also in *col11a1a* fibroblasts, and these two transient fibroblast subtypes exhibited similar expression profiles overall (Supplementary Fig. 6). Two other cell types with ECM-related functions, perivascular cells and valve fibroblasts, also showed increased abundance after injury (Fig. 3a, Supplementary Fig. 4c and Supplementary Data File 4). Other fibroblast types showed only moderate changes after injury. Importantly, established markers for ‘activated’ fibroblasts including *postnb*¹⁷ captured some but not all fibroblast clusters that are generated following injury and were also expressed in nonfibroblast populations including epicardial cells (Supplementary Fig. 6), suggesting that previous marker-based analyses have underestimated the transcriptional diversity of the cardiac fibroblast population.

To spatially resolve the identified fibroblasts, we performed fluorescence in situ hybridization (Fig. 3b and Supplementary Figs. 7 and 8). This analysis confirmed the location of the transient fibroblast cell states in the border zone as well as the injury area. We next wanted to understand which of the transient fibroblasts expressed signalling factors involved in heart regeneration (Fig. 3c and Supplementary Fig. 9). We found particularly high expression of *nrg1* in *col12a1a* fibroblasts, whereas *fn1a* was expressed almost exclusively in *col11a1a* fibroblasts. Furthermore, we observed that retinoic acid (*aldh1a2*) was produced at high levels by *nppc* fibroblasts as well as in the epicardium (Supplementary Fig. 8) and at lower levels in the endocardium (Supplementary Fig. 9). The retinoic acid readout gene *stra6* (stimulated by retinoic acid 6)^{34,35} was expressed highly in *col11a1a*, *col12a1a* and proliferating fibroblasts (but not in cardiomyocytes), suggesting possible communication between these cell types. Moreover, we noticed a very specific

interaction between perivascular cells and blood vessel endothelium via *Cxcl12b*–*Cxcr4a* chemokine signalling. Perivascular cells are known regulators of blood vessel formation³⁶, and it was recently shown that *Cxcl12b*–*Cxcr4a* signalling is important for revascularization of the regenerating heart⁶.

We reasoned that a potential proregenerative function of transient cell states might be driven by secreted factors. Bioinformatic analysis revealed that expression of secretome genes increased in regenerating hearts at 3 d.p.i. and 7 d.p.i. compared with uninjured control hearts (Fig. 3d and Supplementary Fig. 10a). The *col11a1a* and *col12a1a* fibroblasts had the highest secretome expression of all detected cell clusters, and their secretome was enriched in genes with known functions in regeneration, morphogenesis and tissue development (Fig. 3e, Supplementary Fig. 10b and Supplementary Data File 5). The secretome at 3 d.p.i. was also enriched in genes with functions in angiogenesis (for example, *angptl1a* and *angptl2a*), suggesting a possible role of these cell types in revascularization. We next performed ligand–receptor analysis of the identified cell clusters (Methods). We found that the number of putative cell–cell interactions increased drastically after injury and peaked at 3 d.p.i., with a noticeable enrichment of both incoming (receptor) and outgoing (ligand) cell interactions in fibroblast subtypes (Supplementary Fig. 11).

Although our spatiotemporal analysis of single-cell gene expression profiles strongly suggested a proregenerative function of transient fibroblasts, additional experiments were needed to validate this hypothesis. To functionally assess the role of the transient *col11a1a* and *col12a1a* fibroblasts, we applied targeted genetic cell ablation using the nitroreductase/metronidazole (NTR/MTZ) system³⁷. We generated and characterized a transgenic line

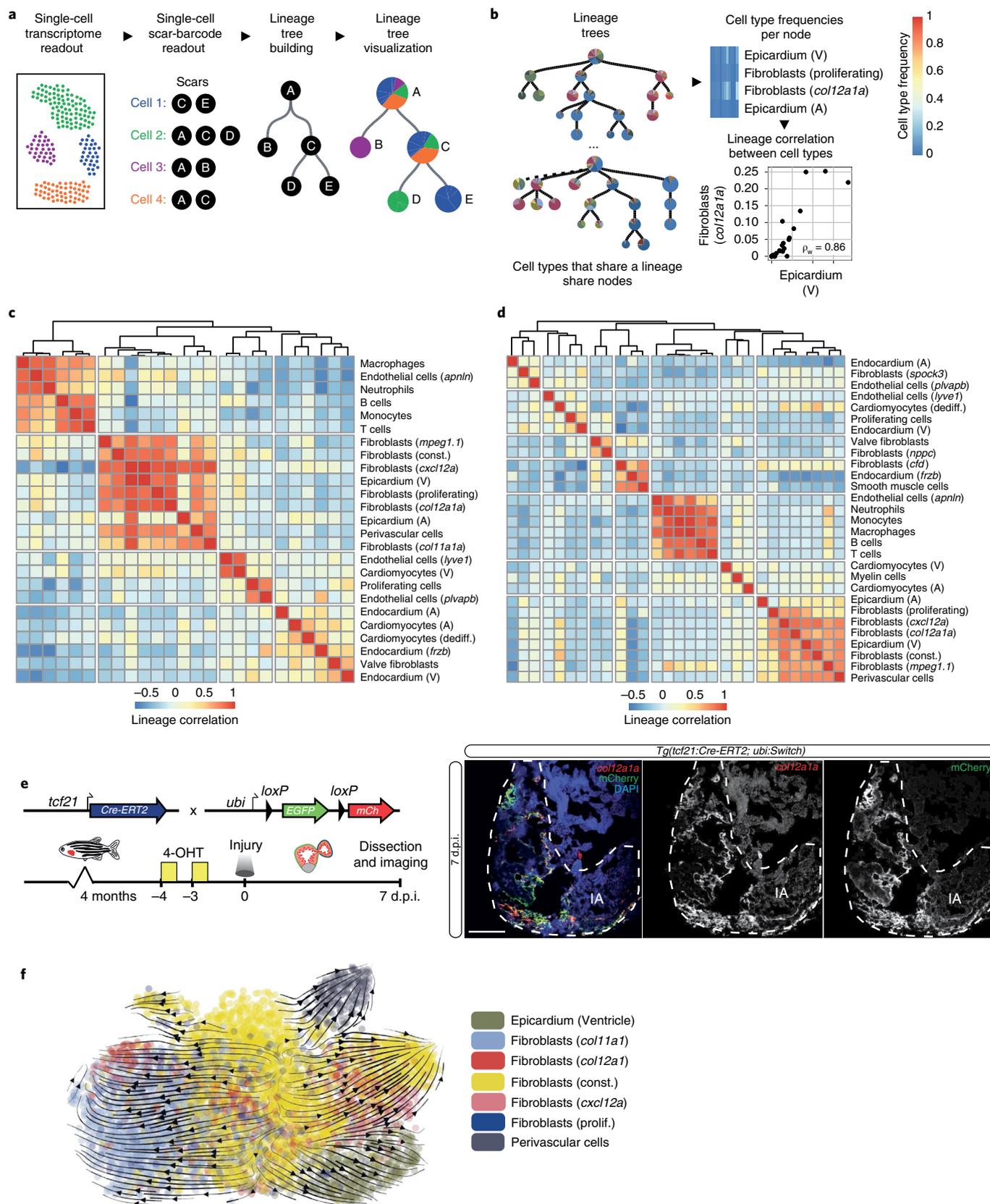


Fig. 4 | Identification of epicardial fibroblasts. a, Cartoon of lineage tree construction using LINNAEUS. **b**, Weighted correlations of cell types over the tree were calculated to quantify lineage similarity. **c,d**, Clustering by lineage correlations at 3 d.p.i. (**c**) and 7 d.p.i. (**d**) revealed the epicardial origin of many niche fibroblasts. **e**, Cre-lox lineage tracing confirmed the epicardial origin of *col12a1a*-expressing cells. Scale bar, 100 μ m. **f**, Trajectory analysis suggested constitutive fibroblasts as the source of *col11a1a* and *col12a1a* fibroblasts at 3 d.p.i.

depletion of *col12a1a*-expressing cells compared with the vehicle control (Fig. 3g,h). At 30 d.p.i., we found that heart regeneration was significantly impaired by *col12a1a*⁺ cell ablation (Fig. 3i,j), suggesting that *col12a1a*-expressing fibroblasts contribute to regeneration. Although it was not possible to formally rule out the possibility that depletion of *col12a1a*-expressing cells may also lead to some tissue degradation of the ventricle independent of a regeneration defect, such effects would be expected to be minor, owing to the lack of *col12a1a* reporter expression in the uninjured ventricle (Supplementary Fig. 12c) and the lack of any visible degeneration outside the injury site (Supplementary Fig. 12e).

In summary, we identified three fibroblast states with potential proregenerative roles during heart regeneration: *col11a1a*, *col12a1a* and *nppc* fibroblasts. Our genetic cell ablation data strongly indicate a role for *col12a1a*-expressing cells in the regenerative niche. However, from which cell types these activated cell states are derived remains to be determined.

Identification of epicardial fibroblasts. We next aimed to elucidate the origins of the transient fibroblast states in order to better understand their mechanism of activation. Trajectory inference using partition-based graph abstraction (PAGA)³⁸ suggested a dense network of possible transitions between all fibroblast subtypes (Supplementary Fig. 13). We reasoned that the large number of putative trajectories might be a computational artifact caused by a multitude of transcriptionally similar cell types, which might be generated by convergent differentiation pathways. We therefore decided to include explicit lineage information and analyse lineage relationships in a high-throughput manner using the LINNAEUS method²⁴. In LINNAEUS, cells are marked by heritable DNA barcodes (genetic scars) that are created by Cas9 during early development (Methods) and whose distinct sequences identify cells that descended from the same parent at the time the scars were created. Here, we injected Cas9 into one-cell-stage embryos in order to record lineage relationships during early development until gastrulation²⁴ (Fig. 1a), which we read out later in the adult heart (all lineage trees are shown in Supplementary Figs. 14–17). By sequencing scars and transcriptomes from the same single cells, we could build lineage trees that revealed the shared developmental origins of the cell types (Fig. 4a and Methods).

We first validated this approach in the adult heart using two control experiments. As expected, when preparing separate single-cell suspensions for the atrium and ventricle, the lineage tree showed separate atrial and ventricular branches, whereas a randomly split suspension did not show branch separation (Supplementary Fig. 18). We next sought to integrate the data across biological replicates in order to infer reproducible lineage relationships between cell types. In a lineage tree, all cells in a node share the same developmental ancestor, and any transient cell originating from a cell in a node can be found in the same node. We calculated correlations between cell type ratios in the different tree nodes to determine which cell types were related by lineage (Fig. 4b and Methods). Hierarchical clustering of these correlations revealed four clusters of cell types at 3 d.p.i. and seven clusters at 7 d.p.i. (Fig. 4c,d, Supplementary Fig. 19 and Methods). At both time points, all immune cells shared a

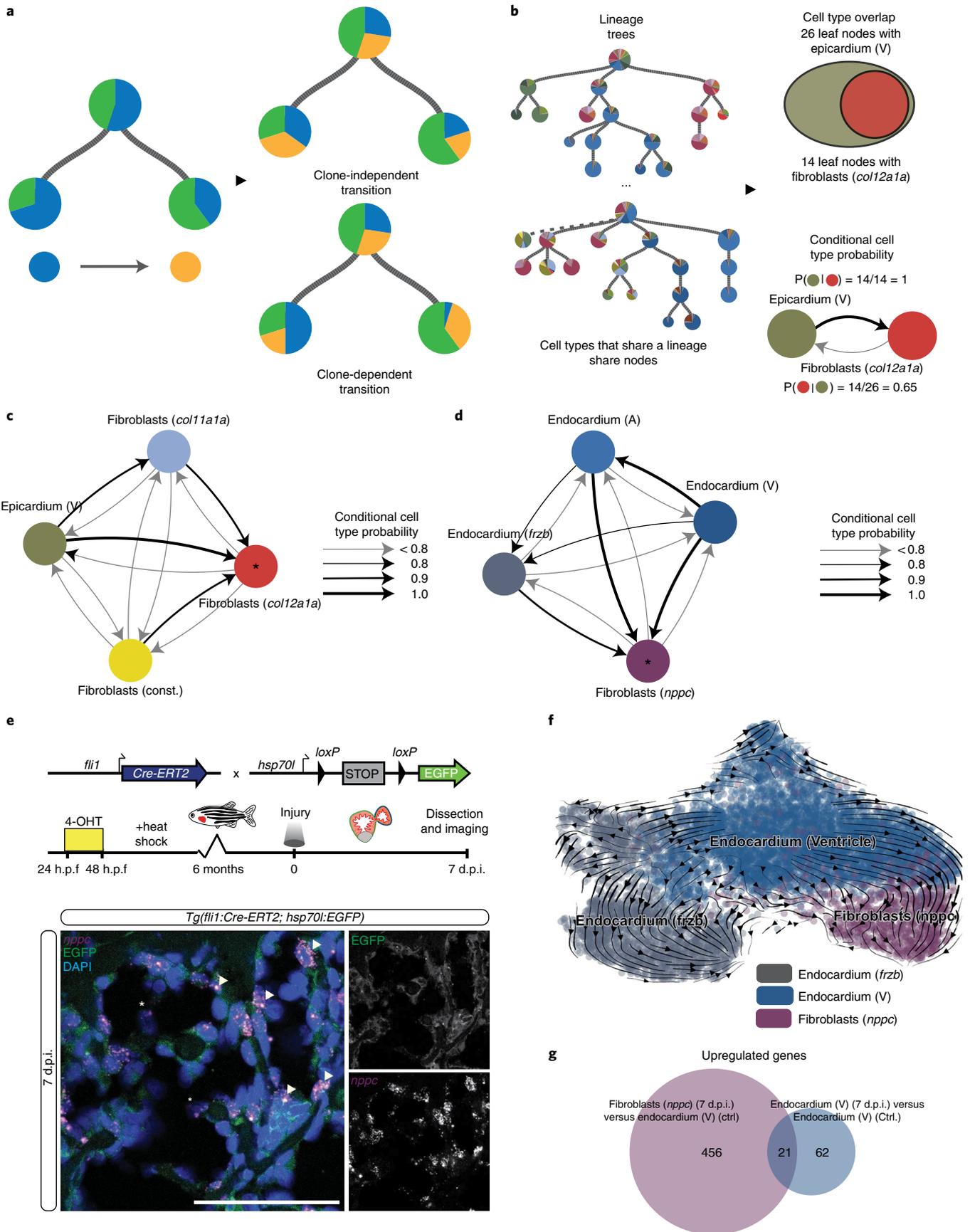
common lineage, validating our approach. We observed a clustering of several fibroblast types, including *col11a1a*, *col12a1a* and constitutive fibroblasts, together with epicardial cells, strongly suggesting that these fibroblast cell types share a developmental origin with the epicardium (Fig. 4c,d).

To validate the lineage origins of *col11a1a* and *col12a1a* fibroblasts, we performed a genetic Cre–lox lineage tracing experiment in regenerating hearts using the transgenic line *TgBAC(tc21:Cre-ERT2;ubi:Switch)*. Our scRNA-seq data showed that *tc21* was expressed in epicardial cells, constitutive fibroblasts and other fibroblast types of the epicardial cluster (Supplementary Fig. 20). After recombination, expressed mCherry at 7 d.p.i. colocalized with endogenously expressed *col12a1a* (Fig. 4e), corroborating the origin of the transient *col11a1a* and *col12a1a* fibroblasts as being the epicardium or epicardial-derived fibroblasts.

LINNAEUS reliably identifies the developmental origins of cell types, but lineage recording is limited to early development. It therefore remained unclear from which source cell type the transient fibroblasts originated following injury in the adult heart – for instance, we could not distinguish whether *col11a1a* and *col12a1a* fibroblasts were derived from epicardial cells, constitutive fibroblasts or any other cell type in the epicardial cluster. Furthermore, expression of *tc21* is not specific enough to address this question using Cre–lox lineage tracing (Supplementary Fig. 20), and reporter-based approaches reach their limits when analysing cell states with high resolution (our data did not identify suitable marker genes that could be used for lineage tracing of constitutive fibroblasts). To further elucidate the origins of the transient epicardial fibroblast states, we therefore applied a transcriptome-based trajectory inference method based on PAGA and RNA velocity³⁹ to all ventricular cell types from epicardial lineage clusters at 3 d.p.i. and 7 d.p.i. (Fig. 4f, Supplementary Fig. 21 and Methods). At both time points, this approach showed that transient *col11a1a* and *col12a1a* fibroblasts originated from constitutive fibroblasts (Fig. 4f and Supplementary Fig. 21a) and suggested that the extremely transient *col11a1a* fibroblasts (which were only present at 3 d.p.i.) turned into *col12a1a* fibroblasts at 7 d.p.i. Genes that were upregulated along these trajectories were those encoding collagens *col11a1a* and *col12a1a*, proregenerative ECM factor *fn1a*, epicardial activation marker *postnb*, retinoic acid signalling response gene *stra6* and cardiomyocyte mitogen *nrg1* (Supplementary Fig. 21b,c).

Identification of endocardial fibroblasts. The transient *nppc* fibroblasts and several other fibroblast subtypes (*spock3*, *cfid* and valve fibroblasts) were not part of the epicardial lineage cluster at 3 d.p.i. or 7 d.p.i. but displayed moderate positive correlations with the endocardium and with each other. We reasoned that our correlation-based analysis had an underlying assumption that all clones would exhibit similar transition rates (Fig. 5a). This assumption may not be true for all cell types, as clones located in the injured part of the heart may not respond to the perturbation. When downsampling to 50% of the clones, we found that the cluster of epicardial fibroblasts remained stable, whereas the correlations among *nppc*, *spock3*, *cfid* and valve fibroblasts varied between iterations (Supplementary Fig. 22 and Methods). Such variability would

Fig. 5 | Identification of endocardial fibroblasts. **a**, Clone-independent transitions (for example 50%, top) showed similar transition rates across clones, leading to a strong correlation between yellow and blue cell types. Clone-dependent transition rates (for example 30% and 80%, bottom) caused yellow and blue cell types to lose correlation, requiring a different approach for lineage analysis. **b**, Conditional cell type probabilities could pinpoint lineage origins despite clone-dependent transitions. **c**, Conditional probabilities reproduced the epicardial lineage origin of *col12a1a* fibroblasts. **d**, Endocardial cell types were present in more than 80% of all nodes containing *nppc*, *spock3* and valve fibroblasts. Asterisks in **c** and **d** indicate the queried cell type. **e**, Cre–lox lineage tracing confirmed the endocardial origin of *nppc*-expressing fibroblasts in cryoinjured hearts at 7 d.p.i. Sections were stained for GFP (green) and *nppc* (magenta). Arrowheads point to *nppc* and EGFP colocalization. Asterisks point to *nppc*-positive delaminating cells. Scale bar, 100 μm. **f**, Trajectory analysis revealed a potential transition from endocardial cells to *nppc* fibroblasts. **g**, Venn diagram of upregulated genes in *nppc* fibroblasts compared with activated endocardium (7 d.p.i.).



be expected if there were considerable differences between the activation profiles of the individual clones that make up these cell types.

We therefore developed a second computational approach, in which we analysed the conditional probabilities of co-occurrence of a cell type of interest with all other cell types (Fig. 5b and Methods). We reasoned that such a targeted approach would yield more reliable results specifically for cell populations of variable clonal behaviour. As validation, we found that *col11a1a* fibroblasts, constitutive fibroblasts and epicardium (ventricle) were present in >90% of the 3 d.p.i. lineage tree nodes that contained *col12a1a* fibroblasts (Fig. 5c). At 7 d.p.i., we found that the different endocardial cell types (endocardium (atrium), endocardium (ventricle) and endocardium (*frzb*)) were present in more than 80% of all nodes containing transient *nppc* fibroblasts, suggesting a lineage relationship between this fibroblast type and the endocardium (Fig. 5d). Similarly, the analysis suggested that *spock3* and valve fibroblasts were also of endocardial origin, whereas *cfid* fibroblasts were of epicardial origin (Supplementary Fig. 23). Hence, our analysis revealed the existence of three endocardial-derived fibroblast subtypes (*nppc*, *spock3* and valve fibroblasts), whereas all other fibroblasts shared a lineage origin with the epicardium. To validate our computational approach, we performed genetic lineage tracing using the transgenic line *Tg(fli1:Cre-ERT2; hsp70l:Switch)*, which, after embryonic recombination, marks endothelial cells in the adult heart^{14,17}. Expressed EGFP at 7 d.p.i. colocalized with endogenously expressed *nppc* (Fig. 5e), corroborating the endocardial origin of the transient *nppc* fibroblasts.

Only ~20% of the nodes with endocardium (ventricle) contained *nppc* fibroblasts (Supplementary Fig. 24a), suggesting that endocardial fibroblasts are only generated in a subset of the endocardium, potentially in the injury area. This clone-specific injury reaction of the endocardium would explain why targeted lineage analysis based on conditional probabilities outperforms correlation-based lineage analysis. We hypothesized that the internal position of the endocardium, as well as the transient nature and low clonality of the *nppc* fibroblasts, might mean that these cells are only generated from damaged endocardial clones. We confirmed that longer contact of the cryoprobe with the heart led to a larger injury area and resulted in much stronger *nppc* expression beyond the border zone, compared with injuries resulting from shorter contact times (Supplementary Fig. 24b,c and Methods).

Trajectory analysis revealed transcriptional similarity and a potential transition path between ventricular endocardium and *nppc* fibroblasts (Fig. 5f). We observed that *nppc* fibroblasts continued to express endothelial genes (for example *vwf* and *fli1a*) in addition to ECM genes, suggesting that they maintain at least part of their endocardial gene expression after turning on a fibrotic gene expression program. This prompted us to investigate in more detail to what degree the gene expression profile of these cells had fibroblast characteristics. In particular, we considered whether the *nppc* fibroblasts might correspond to the activated endocardium that was previously described after injury¹³. We observed from our single-cell transcriptomic data that the endocardium underwent an activation response at 7 d.p.i. compared with healthy control hearts (Fig. 5g). However, this response was weaker than and only partially overlapping with the gene expression profile of *nppc* fibroblasts. Hence, the endocardium responds to injury with two separate transcriptional responses, which differ in the magnitude of their expression changes and which have not been systematically distinguished in the previous literature. Differentially expressed genes of *nppc* fibroblasts include many ECM-related factors (which led us to identify this cluster as a fibroblast state as shown in Fig. 2c) (Supplementary Data File 3). Furthermore, we noticed that *nppc* fibroblasts had upregulation of *twist1b* and express *snail2*, both of which are classical markers of epithelial-to-mesenchymal transition. This, together with the observation that some *nppc*⁺ cells appeared to be

delaminating in our microscopy images (Fig. 5e), suggested that *nppc* fibroblasts may undergo at least a partial transition toward a mesenchymal state. However, as we did not observe a clear downregulation of VE-cadherin (*cdh5*) or upregulation of N-cadherin (*cdh2*), we speculate that *nppc* fibroblasts do not acquire full mesenchymal characteristics. Of note, it was previously reported that endocardial cells express collagen following injury but do not undergo a complete epithelial-to-mesenchymal transition¹⁷, whereas another publication showed that a minor proportion of endocardial cells appear to undergo a more complete transition¹⁴.

In summary, our analyses revealed a clear separation of epicardial- and endocardial-derived fibroblasts, with both lineages giving rise to distinct transient fibroblast states following injury. An epicardial origin for cardiac fibroblasts has also recently been established in zebrafish^{14,40,41}, and our analysis provides strong evidence for the existence of fibroblasts of endocardial origin in the atrium and ventricle, as previously suggested in mouse^{42,43} and zebrafish¹⁴.

Cellular dissection of the role of canonical Wnt signalling. We noticed that fibroblasts expressed many genes related to Wnt signalling (ligands, receptors and modulators) (Fig. 6a and Supplementary Fig. 25), which inspired us to investigate the role of canonical Wnt signalling in this system. The role of Wnt signalling in heart regeneration remains an important open question⁴⁴. On the one hand, Wnt is generally considered to be a proproliferative factor, and Wnt activation has been shown to be beneficial for zebrafish fin and spinal cord regeneration^{33,45}. On the other hand, the role of Wnt signalling in cardiomyocyte dedifferentiation and proliferation is controversial^{16–48}. We hypothesized that Wnt might exert its complex functions by regulating activation of proregenerative cell states, and we investigated whether cells with a shared lineage origin exhibited similar changes following a stimulus, that is, whether lineage relationship is predictive of perturbation response.

We therefore inhibited canonical Wnt signalling after cryoinjury using the well-characterized Wnt/ β -catenin-dependent signalling inhibitor IWR-1 and observed the effects at 3, 7, 15 and 30 d.p.i. (Fig. 6b and Methods). Wnt/ β -catenin signalling inhibition led to a significant delay in heart regeneration, with prolonged fibrosis and increased size of injury area compared with the control (Fig. 6b). Single-cell transcriptomics of IWR-1-treated hearts at 3 d.p.i. and 7 d.p.i. revealed that cardiomyocyte dedifferentiation was delayed (Fig. 6c). Compared with control samples, numbers of dedifferentiated cardiomyocytes were reduced at 3 d.p.i., and they did not localize to the injury area at 7 d.p.i. (Fig. 6c).

Levels of perivascular cells and all endocardial fibroblasts (*nppc*, *spock3* and valve fibroblasts) were strongly reduced after Wnt inhibition, whereas the other transiently increased fibroblasts remained at similar levels overall (Fig. 6d, Supplementary Fig. 26 and Supplementary Data File 4). Specifically, the overall abundance of *col11a1a* and *col12a1a* fibroblasts did not change (Supplementary Fig. 26), even though *col12a1a* fibroblast numbers peaked earlier after Wnt inhibition (at 3 d.p.i. instead of 7 d.p.i.); this may have been an indirect effect due to the observed impairment of heart regeneration. We confirmed these findings by fluorescence in situ hybridization (Fig. 6e). These observations strongly suggest that Wnt signalling is required for the activation of endocardial fibroblasts, akin to the described role of Wnt in inducing endothelial-to-mesenchymal transition in mice^{49,50}.

We next wanted to understand how the absence of these cell types could mediate the observed delay in regeneration. Perivascular cells are known to exert a proangiogenic function through signalling and mechanical stabilization of nascent vasculature⁵¹. High expression of the chemokine *cxcl12b* in perivascular cells in our data set (Fig. 3c) further supported the putative proangiogenic role of these cells. Furthermore, NPPC signalling has been shown to have a complex range of functions in the cardiovascular system^{52,53} and has

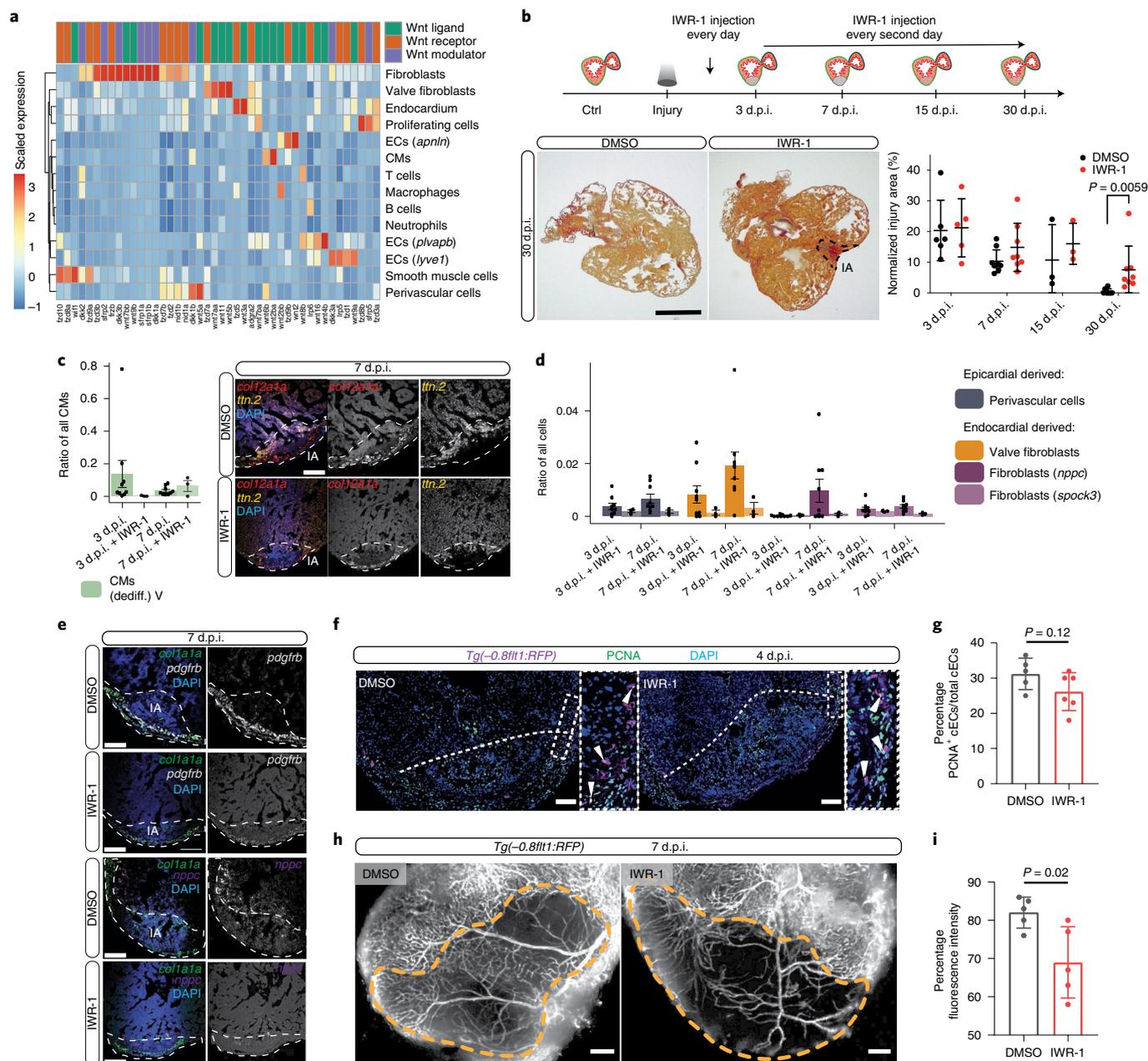


Fig. 6 | Cellular dissection of the role of canonical Wnt signaling. **a**, Expression of Wnt signalling factors in different cell types of the zebrafish regenerating heart. **b**, Upper panel: cartoon summary of IWR-1 Wnt inhibition experiments. Lower left: histological comparison of the IA at 30 d.p.i. with intraperitoneal (IP) injections of IWR-1 or DMSO. In the paraffin AFOG-stained sections, fibrin-red and collagen-blue are clearly visible in the IA in IWR-1-treated samples at 30 d.p.i. Scale bar, 300 μ m. Lower right: relative size of the IA area as a percentage across all histological replicates; data shown as mean and s.d. Two-tailed unpaired Student's *t* test between groups at each time point, $P = 0.0059$. **c**, Left: changes in relative numbers of dedifferentiated CMs at 3 and 7 d.p.i. between IWR-1- and DMSO-treated hearts (error bars indicate s.e.m.). Right: localization of dedifferentiated (*ttn.2*) cardiomyocytes at 7 d.p.i. with and without Wnt inhibition. Scale bar, 100 μ m. **d**, Changes in relative numbers of non-CMs following Wnt inhibition at 3 and 7 d.p.i. (error bars indicate s.e.m.). **e**, Fluorescence in situ hybridization of perivascular cells (*pdgfrb*, white) and *nppc* fibroblasts (*nppc*, purple) at 7 d.p.i. with and without Wnt inhibition. Scale bar in **c** and **e**, 100 μ m. White dashed lines indicate the IA. **f**, Immunostaining of *Tg(-0.8ft1:RFP)* hearts at 4 d.p.i. with and without Wnt inhibition; RFP (coronaries, magenta), PCNA (proliferation marker, green) and DNA (DAPI, blue). Arrowheads point to PCNA⁺ cECs, white dashed lines indicate the IA. Scale bar, 100 μ m. **g**, Percentages of PCNA⁺ cECs in DMSO-injected ($n = 5$) and IWR-1-injected ($n = 6$) fish at 4 d.p.i. Data are shown as mean and s.d. Two-tailed unpaired Student's *t* test, $P = 0.12$. **h**, Whole-mount images of *Tg(-0.8ft1:RFP)* hearts at 7 d.p.i. with and without Wnt inhibition. Orange dashed lines indicate injury area. Scale bar, 100 μ m. **i**, Percentages of RFP fluorescence intensity in injured tissue of DMSO-injected ($n = 5$) and IWR-1-injected ($n = 5$) fish at 7 d.p.i. Data are shown as mean and s.d. Two-tailed unpaired Student's *t* test, $P = 0.02$.

recently been suggested to be involved in fin regeneration⁵⁴. We therefore hypothesized that Wnt inhibition might delay heart regeneration via reduced revascularization following injury.

To assess revascularization, we quantified the proliferation of coronary endothelial cells (cECs) at 4 d.p.i. as well as the coverage of coronaries in the injury area at 7 d.p.i. (Methods). Following

administration of IWR-1, we observed a marginal but not significant reduction in cEC proliferation at 4 d.p.i. (Fig. 6f,g). However, at 7 d.p.i., we observed a clear and statistically significant reduction in the coverage of coronary vessels in the injury area (Fig. 6h,i). In summary, these results suggest that the observed reduction in heart regeneration following Wnt inhibition is at least partially mediated by a revascularization defect. This defect does not appear to be caused by reduced proliferation of cECs and might instead be linked to other aspects of revascularization, such as migration of coronaries or interactions with perivascular cells.

Discussion

Here, we used a combination of single-cell transcriptomics and spatiotemporal information to characterize putative proregenerative cell states in the adult zebrafish heart in a data-driven manner. This approach allowed us to identify three transient cell states with fibroblast characteristics (*col11a1a*, *col12a1a* and *nppc* fibroblasts) that are major sources of known proregenerative genes and also express additional secreted factors with potential proregenerative functions. These fibroblast cell states are ideally suited to be cellular drivers for regeneration based on their expression profiles, their location in the injury area and their transient appearance after injury. By combining high-throughput lineage tracing with RNA velocity and single-cell trajectory inference, we were able to systematically identify the origin of these cell states. This approach effectively combines information about two time points – CRISPR barcoding during early development and trajectory inference at the time of analysis. Furthermore, we developed two complementary computational approaches for integrating CRISPR barcoding data across biological replicates in order to obtain statistically sound information on lineage relationships between cell types. Taken together, the results obtained with these approaches can serve as a blueprint for understanding the origin of transient cell states that are generated under disease conditions.

Several lines of evidence (CRISPR lineage analysis, Cre-*lox* lineage tracing, transcriptome profiles and long versus short injury) showed that *nppc* fibroblasts were derived from the endocardium. The endocardium exhibited two separate activation programs that had not previously been systematically distinguished: *nppc* fibroblasts and a less pronounced endocardial activation profile. Functional experiments showed that activation of endocardial fibroblasts depends on canonical Wnt signalling. Wnt inhibition leads to a delay in regeneration which is accompanied by a revascularization defect. Although it is tempting to link the observed depletion of *nppc* fibroblasts and perivascular cells following Wnt inhibition to reduced revascularization, further experiments would be needed to substantiate this point. A combination of CRISPR lineage tracing, trajectory inference and Cre-*lox* lineage tracing showed that the *col11a1a* and *col12a1a* fibroblasts shared their lineage origin with the epicardium and were generated from constitutive fibroblasts following injury. *Col11a1a* and *col12a1a* fibroblasts have very similar transcriptional profiles, and our RNA velocity analysis suggested that *col11a1a* fibroblasts are precursors of *col12a1a* fibroblasts. Targeted cell type depletion of *col12a1a*-expressing cells after injury led to impaired heart regeneration, supporting the proregenerative function of these cells. In summary, our detailed cellular dissection of the fibrotic response in the zebrafish heart characterized sources and mechanisms of proregenerative programs. The large diversity of fibroblast cell populations found in the regenerating zebrafish heart suggests that fibrosis should be seen as a collective phenomenon to which a variety of cell types and states contribute.

The pathways that contribute to the regenerative capacity of the zebrafish heart are conserved in humans. Thus, our analysis may lead us to better understand the limited regenerative capacity of the human heart, while at the same time opening up an exciting strategy to identify potential therapeutic approaches. Our data set and

the experimental and computational approaches presented here will serve as a powerful resource for identification of additional candidate interactions between cell states involved in heart regeneration.

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/s41588-022-01129-5>.

Received: 15 February 2022; Accepted: 9 June 2022;

Published online: 21 July 2022

References

- Poss, K. D., Wilson, L. G. & Keating, M. T. Heart regeneration in zebrafish. *Science* **298**, 2188–2190 (2002).
- Gonzalez-Rosa, J. M., Burns, C. E. & Burns, C. G. Zebrafish heart regeneration: 15 years of discoveries. *Regeneration* **4**, 105–123 (2017).
- Gonzalez-Rosa, J. M., Martin, V., Peralta, M., Torres, M. & Mercader, N. Extensive scar formation and regression during heart regeneration after cryoinjury in zebrafish. *Development* **138**, 1663–1674 (2011).
- Kikuchi, K. et al. Primary contribution to zebrafish heart regeneration by *gata4*⁺ cardiomyocytes. *Nature* **464**, 601–605 (2010).
- Jopling, C. et al. Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* **464**, 606–609 (2010).
- Marín-Juez, R. et al. Coronary revascularization during heart regeneration is regulated by epicardial and endocardial cues and forms a scaffold for cardiomyocyte repopulation. *Dev. Cell* **51**, 503–515.e4 (2019).
- Kikuchi, K. et al. Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration. *Dev. Cell* **20**, 397–404 (2011).
- Han, Y. et al. Vitamin D stimulates cardiomyocyte proliferation and controls organ size and regeneration in zebrafish. *Dev. Cell* **48**, 853–863.e5 (2019).
- Wu, C.-C. et al. Spatially resolved genome-wide transcriptional profiling identifies BMP signaling as essential regulator of zebrafish cardiomyocyte regeneration. *Dev. Cell* **36**, 36–49 (2016).
- Ogawa, M. et al. Krüppel-like factor 1 is a core cardiomyogenic trigger in zebrafish. *Science* **372**, 201–205 (2021).
- Fang, Y. et al. Translational profiling of cardiomyocytes identifies an early Jak1/Stat3 injury response required for zebrafish heart regeneration. *Proc. Natl Acad. Sci. USA* **110**, 13416–13421 (2013).
- Lepilina, A. et al. A dynamic epicardial injury response supports progenitor cell activity during zebrafish heart regeneration. *Cell* **127**, 607–619 (2006).
- Münch, J., Grivas, D., González-Rajal, Á., Torregrosa-Carrión, R. & de la Pompa, J. L. Notch signalling restricts inflammation and *serpine1* expression in the dynamic endocardium of the regenerating zebrafish heart. *Development* **144**, 1425–1440 (2017).
- Allanki, S. et al. Interleukin-11 signaling promotes cellular reprogramming and limits fibrotic scarring during tissue regeneration. *Sci. Adv.* **7**, eabg6497 (2021).
- de Preux Charles, A.-S., Bise, T., Baier, F., Marro, J. & Jazwińska, A. Distinct effects of inflammation on preconditioning and regeneration of the adult zebrafish heart. *Open Biol.* **6**, 160102 (2016).
- Lai, S.-L. et al. Reciprocal analyses in zebrafish and medaka reveal that harnessing the immune response promotes cardiac regeneration. *eLife* **6**, e25605 (2017).
- Sánchez-Iranzo, H. et al. Transient fibrosis resolves via fibroblast inactivation in the regenerating zebrafish heart. *Proc. Natl Acad. Sci. USA* **115**, 4188–4193 (2018).
- Paik, D. T., Cho, S., Tian, L., Chang, H. Y. & Wu, J. C. Single-cell RNA sequencing in cardiovascular development, disease and medicine. *Nat. Rev. Cardiol.* **17**, 457–473 (2020).
- Tucker, N. R. Transcriptional and cellular diversity of the human heart. *Circulation* **142**, 466–482 (2020).
- Litviňuková, M. et al. Cells of the adult human heart. *Nature* **588**, 466–472 (2020).
- Farbehi, N. et al. Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *eLife* **8**, e43882 (2019).
- Honkoop, H. et al. Single-cell analysis uncovers that metabolic reprogramming by ErbB2 signaling is essential for cardiomyocyte proliferation in the regenerating heart. *eLife* **8**, e50163 (2019).
- Ma, H. et al. Functional coordination of non-myocytes plays a key role in adult zebrafish heart regeneration. *EMBO Rep.* **22**, e52901 (2021).

24. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
25. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 1–22 (2018).
26. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
27. Pan, Y. A. et al. ZebraBow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish. *Development* **140**, 2835–2846 (2013).
28. Junker, J. P. et al. Genome-wide RNA tomography in the zebrafish embryo. *Cell* **159**, 662–675 (2014).
29. Aliee, H. & Theis, F. AutoGeneS: automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *Cell Syst.* **12**, 706–715 (2021).
30. Gemberling, M., Karra, R., Dickson, A. L. & Poss, K. D. Nrg1 is an injury-induced cardiomyocyte mitogen for the endogenous heart regeneration program in zebrafish. *eLife* **4**, e05871 (2015).
31. Wang, J., Karra, R., Dickson, A. L. & Poss, K. D. Fibronectin is deposited by injury-activated epicardial cells and is necessary for zebrafish heart regeneration. *Dev. Biol.* **382**, 427–435 (2013).
32. Marro, J., Pfefferli, C., de Preux Charles, A.-S., Bise, T. & Jazwińska, A. Collagen XII contributes to epicardial and connective tissues in the zebrafish heart during ontogenesis and regeneration. *PLoS ONE* **11**, e0165497 (2016).
33. Wehner, D. et al. Wnt signaling controls pro-regenerative collagen XII in functional spinal cord regeneration in zebrafish. *Nat. Commun.* **8**, 126 (2017).
34. Bouillet, P. et al. Efficient cloning of cDNAs of retinoic acid-responsive genes in P19 embryonal carcinoma cells and characterization of a novel mouse gene, *Stral* (mouse LERK-2/Eplg2). *Dev. Biol.* **170**, 420–433 (1995).
35. Isken, A. et al. RBP4 disrupts vitamin A uptake homeostasis in a STRA6-deficient animal model for Matthew-Wood syndrome. *Cell Metab.* **7**, 258–268 (2008).
36. Bergers, G. & Song, S. The role of pericytes in blood-vessel formation and maintenance. *Neuro Oncol.* **7**, 452–464 (2005).
37. Curado, S. et al. Conditional targeted cell ablation in zebrafish: a new tool for regeneration studies. *Dev. Dyn.* **236**, 1025–1035 (2007).
38. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
39. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 1–25 (2018).
40. González-Rosa, J. M., Peralta, M. & Mercader, N. Pan-epicardial lineage tracing reveals that epicardium derived cells give rise to myofibroblasts and perivascular cells during zebrafish heart regeneration. *Dev. Biol.* **370**, 173–186 (2012).
41. Kikuchi, K. et al. tcf21+ epicardial cells adopt non-myocardial fates during zebrafish heart development and regeneration. *Development* **138**, 2895–2902 (2011).
42. Zeisberg, E. M. et al. Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nat. Med.* **13**, 952–961 (2007).
43. Tombor, L. S. et al. Single cell sequencing reveals endothelial plasticity with transient mesenchymal activation after myocardial infarction. *Nat. Commun.* **12**, 681 (2021).
44. Ozhan, G. & Weidinger, G. Wnt/ β -catenin signaling in heart regeneration. *Cell Regen.* **4**, 3 (2015).
45. Chen, B. et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat. Chem. Biol.* **5**, 100–107 (2009).
46. Zhao, L., Ben-Yair, R., Burns, C. E. & Burns, C. G. Endocardial notch signaling promotes cardiomyocyte proliferation in the regenerating zebrafish heart through Wnt pathway antagonism. *Cell Rep.* **26**, 546–554.e5 (2019).
47. Peng, X. et al. Induction of Wnt signaling antagonists and p21-activated kinase enhances cardiomyocyte proliferation during zebrafish heart regeneration. *J. Mol. Cell. Biol.* **13**, 41–58 (2021).
48. Bertozzi, A., Wu, C.-C., Hans, S., Brand, M. & Weidinger, G. Wnt/ β -catenin signaling acts cell-autonomously to promote cardiomyocyte regeneration in the zebrafish heart. *Dev. Biol.* **481**, 226–237 (2022).
49. Aisagbonhi, O. et al. Experimental myocardial infarction triggers canonical Wnt signaling and endothelial-to-mesenchymal transition. *Dis. Models Mech.* **4**, 469–483 (2011).
50. Duan, J. et al. Wnt1/ β catenin injury response activates the epicardium and cardiac fibroblasts to promote cardiac repair: Wnt1/ β -catenin injury response regulates cardiac repair. *EMBO J.* **31**, 429–442 (2012).
51. Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nat. Med.* **6**, 389–395 (2000).
52. Bubb, K. J. et al. Endothelial C-type natriuretic peptide is a critical regulator of angiogenesis and vascular remodeling. *Circulation* **139**, 1612–1628 (2019).
53. Moyes, A. J. & Hobbs, A. J. C-type natriuretic peptide: a multifaceted paracrine regulator in the heart and vasculature. *IJMS* **20**, 2281 (2019).
54. Shi, L. et al. Systematic profiling of early regulators during tissue regeneration using zebrafish model. *Wound Repair Regen.* **29**, 189–195 (2021).

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



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

© The Author(s) 2022

Methods

Ethics statement. This study complied with all relevant ethical regulations. Zebrafish were bred, raised and maintained in accordance with the FELASA guidelines⁵⁵, the guidelines of the Max Delbrück Center for Molecular Medicine and the Max Planck Society, and the local authorities for animal protection (Landesamt für Gesundheit und Soziales, Berlin, and The Veterinary department of the Regional Board of Darmstadt, Darmstadt, Germany) for the use of laboratory animals, based on the current version of German law on the protection of animals and EU directive 2010/63/EU on the protection of animals used for scientific purposes. In addition, housing and breeding standards followed the international 'Principles of Laboratory Animal Care' (NIH publication no. 86-23, revised 1985).

Zebrafish. Experiments in this study used the zebrafish wild-type strain *AB* and transgenic lines *Tg[ubi:zebrabow-M]^{g1317g}* (ref. ²⁷), *TgBAC[cryaa:EGFP, tg21:Cre-ERT2]^{pd4}* (ref. ⁴¹), *Tg[-3.5ubi-loxP-EGFP-loxP-mCherry] (ubi:Switch)³⁶*, *Tg[(fli1:Cre-ERT2)^{ns9}* (ref. ¹⁷), *Tg[-1.5hsp70l:loxP-STOP-loxP-EGFP, cryaa:Venus]^{zh701}* (ref. ⁵⁷), *TgUAS:NTR-mCherry)³⁸*, *Tg(-0.8flt1:RFP)^{hnc5333}* (ref. ⁵⁹) and *Tg(-4kbccl12a1a:GAL4VP16, ins:YFP)* (this study). Adult zebrafish of random sex, aged between 4 months and 1 year and with a length of at least 3 cm were used in all experiments. Animals were randomly assigned to the test groups in all experiments.

Cryo-injury procedure. Cryoinjury was performed as previously described^{60,61}. Briefly, fish were preanesthetized in water containing 0.03 mg ml⁻¹ Tricaine (PHARMAQ, pH 7). The concentration was then increased to 0.168 mg ml⁻¹ for anesthesia. Fish were placed with the ventral side facing up in a foam holder under a dissecting scope. To access the heart, a small incision was made through the body wall and the pericardium, using microdissection forceps and scissors. Once the pericardial sac was opened, the heart ventricle was exposed by gently compressing the abdomen. Excess water was carefully removed by blotting with tissue paper, not allowing the fish skin to dry. Then, a stainless steel cryoprobe precooled in liquid nitrogen was applied to the ventricular wall for 20 s. Fish were then placed in a tank of fresh system water with 1.5 mg l⁻¹ morphine sulfate for 6 h⁶²; for the analysis of scRNA-seq data presented here, we also included two morphine-treated and two control hearts from Lelek et al.⁶² (~20,000 cells, about 10% of the total data set). Reanimation was enhanced by gill oxygenation, in which water around the gills was aerated by pipetting for a couple of minutes. To investigate the effects of the cryoinjury depth, fish were injured using the same procedure, but the cryoprobe was applied for 5 s, 10 s, 15 s, 20 s and 25 s, respectively. Established protocols use 20 s of cryoprobe application^{60,61}. For all injuries, a timer was used to assure the reproducible timing of the cryoprobe contact with the heart tissue. Hearts were analysed at 3, 7, 15 and 30 d.p.i.

Cas9 and sgRNA injections for high-throughput lineage tracing. For CRISPR-Cas9 lineage tracing, we used the same experimental approach as previously described²⁴. In brief, we set up crosses between zebrafish *M²⁷* adults with high dTomato fluorescence, and we injected the embryos at the one-cell stage with 2 nl Cas9 protein (NEB, final concentration 350 ng µl⁻¹) in combination with a sgRNA targeting dTomato (final concentration 50 ng µl⁻¹, sequence: GGTGTCCACGTTAGTAGCGTTT TAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAGTGGCACCGAGTCGGTCTTTT). Loss of dTomato fluorescence was used as a marker of successful generation of genetic scars. Injected fish were raised to adulthood for scRNA-seq analysis of the heart.

Preparation of single-cell suspensions. Adult zebrafish (injured and uninjured) were humanely killed by immersion in ice-cold water (0–4 °C) for 20 min. The heart was dissected from the fish and transferred into cold Hanks' Balanced Salt Solution (HBSS). The dissection included the atrium, ventricle and bulbus arteriosus, except for samples in which only the atrium or the ventricle was isolated. A needle and a syringe filled with cold HBSS were used to pierce the lumen of the heart and thoroughly wash away most of the erythrocytes. Afterwards, the heart was opened carefully with forceps, and the tissue was incubated at 37 °C for 30 min in 500 µl HBSS containing Liberase enzyme mix (Sigma-Aldrich, 0.26 U ml⁻¹ final concentration) and Pluronic F-68 (Thermo Fisher Scientific, 0.1%) while shaking at 750 r.p.m. on an Eppendorf ThermoMixer F2.0 with intermittent pipette mixing. After most of the tissue had been dissociated, the reaction was stopped by adding 500 µl cold HBSS supplemented with 1% bovine serum albumin (BSA). The suspension was centrifuged at 250g at 4 °C and washed two times with 500 µl cold HBSS containing 0.05% BSA, then filtered through a cell strainer of 35 µm diameter. The quality of the single-cell suspension was then confirmed under a microscope, and cells were counted before scRNA-seq library preparation. We noticed some fluctuations in cell type ratios detected by scRNA-seq (Figs. 2a and 3a), which we speculate could have been due to variability in the cryoinjury procedure or in the timing of the repair process. Furthermore, our dissociation protocol might have led to a depletion of cardiomyocytes, which, as they are large cells, tend to break relatively easily. Indeed, cardiomyocytes on average have lower transcript counts than nonmyocytes (~1,600 compared with ~4,500 transcripts per cell), which may indicate that we detected nuclei rather than cells in the case of cardiomyocytes.

Single-cell RNA-seq and scar detection. Single cells were captured using a Chromium Single Cell 3' kit (10X Genomics, PN-1000075), according to the manufacturer's recommendations. We aimed for 10,000 cells per library whenever possible. Both v2- and v3-chemistry were used for the data presented here. For simultaneous detection of transcriptome and CRISPR-induced genetic scar sequences for lineage tracing, we proceeded as previously described²⁴. In brief, the scar sequences were enriched by a two-round nested PCR approach, using 10 µl of the 10X Genomics cDNA and target site specific primers (dTomato transgene). The scar library was then indexed using the indexing primers of the Chromium kit to ensure they could be sequenced and demultiplexed together with the transcriptome. Samples were sequenced on an Illumina NextSeq 500 150 bp and an Illumina HiSeq 2500 200 bp after successful quality control by Bioanalyzer (DNA HS kit, Agilent).

Mapping and clustering of single-cell mRNA data. A zebrafish transcriptome was created with Cell Ranger 3.0.2 from GRCz11, release 92. Alignment and transcript counting of libraries were done using Cell Ranger 3.0.2. Library statistics are summarized in Supplementary Data File 1. The transcriptome data were filtered, clustered and visualized using Seurat 3.0 (ref. ⁶³). We excluded the cluster of *mpeg1.1* fibroblasts (Fig. 2c,d) from further analysis, as we could not rule out the possibility that the expression of the macrophage marker *mpeg1.1* in this fibroblast cluster was an artifact caused by, for example, fragmentation of macrophages⁶⁴.

Histological staining, analysis and imaging. For the analysis of the injury areas, animals were humanely killed at different times after injury by placing them in ice-cold water at 0–4 °C for 20 min. Hearts were dissected and incubated in 2 U ml⁻¹ heparin and 0.1 M KCl in phosphate-buffered saline (PBS) for around 30 min. Cryosamples were fixed in 4% paraformaldehyde (PFA) in PBS overnight at 4 °C, washed in PBS for 3 × 10 min and incubated overnight in 30% sucrose in PBS. Samples were then frozen in Tissue-Tek O.C.T. Compound (Sakura) on dry ice. Tissue was cut at 7 µm on a cryostat (Leica) using Superfrost slides (Thermo Fisher Scientific). For paraffin sections, tissue was fixed in 4% PFA in PBS overnight at 4 °C, washed in PBS for 3 × 10 min and incubated in an ethanol series at 30% for 10 min, 50% for 20 min and 70% overnight at 4 °C. Next, samples were dehydrated in an ethanol series at 80% for 30 min, 90% for 30 min, 96% for 1 h, 100% for 5 min and 100% for 1 h, then in Toluol for 45 min and Toluol for 15 min, and finally embedded in paraffin. Connective tissue was stained using acid fuchsin orange G (AFOG). In brief, slides were dried for 30 min at room temperature. Next, slides were incubated in Bouin Solution (Sigma-Aldrich) for 2 h at 60 °C and left for overnight incubation under the hood. Slides were washed for 30 min under running double-distilled water (ddH₂O) and incubated for 7 min in 1% phosphomolybdic acid (Sigma-Aldrich). Samples were washed for 3 min in running ddH₂O and incubated with AFOG solution (made in our laboratory, Sigma-Aldrich) for 3 min. Slides were washed until clear with running ddH₂O and rehydrated with 70%, 94% and 2 × 100% ethanol and for 2 × 5 min with xylol (Sigma-Aldrich). Slides were mounted with xylene mounting medium (Merck) and allowed to dry overnight under the hood. For the analysis of injury size, the total ventricular tissue area and injury area on multiple slides were measured. Injury area was identified by visualization of collagen (blue) and fibrin (red). Images were obtained with a Keyence Microscope BZX800 using Viewer and analysed with ImageJ/Fiji (v2.1.0).

RNAscope Multiplex Fluorescent v2 assay for immunofluorescence in situ hybridization and imaging. The RNAscope assay (ACD) for fluorescence in situ hybridization was used to localize different messenger RNAs (mRNAs) expressed by various cell types during heart regeneration. The technique was performed according to the manufacturer's instructions for fixed frozen tissues (323100-USM) or in combination with antibody staining (MK 51-150). Cryosections (7 µm) were dried for 30 min at room temperature before the experiment. The following changes were made: slides were incubated at 99 °C for 15 min with an antigen retrieval solution (ACD) using a steamer (WMH). Signal development was done using Tyramide Signal Amplification (TSA) plus fluorophores (PerkinElmer) fluorescein, cyanine 3 and cyanine 5 at 1:1,000 dilution. The probes used are listed in Supplementary Table 1. In addition, anti-EGFP (Sigma-Aldrich, catalog no. G1544, 1:500) and secondary antibody AlexaFluor 488 goat anti-rabbit IgG (Thermo Scientific, catalog no. A-11008, 1:500) were used in the protocol combining immunohistochemistry with in situ hybridization. The fluorescence in situ hybridization for the whole-mount embryo was performed according to the manufacturer's instructions (MK 50-016). Images were obtained with a Zeiss LSM880 confocal microscope using Zen 2.3 SP1 FP3 (v.14.0.0.0) software and analysed using ImageJ/Fiji (v2.1.0) and Photoshop (v20.0.6) software.

Microscopy analysis of cardiomyocyte proliferation. Zebrafish hearts were extracted and fixed in 4% PFA in PBS overnight, washed in PBS for 3 × 10 min and incubated overnight in 30% sucrose in PBS. Samples were then frozen in Tissue-Tek O.C.T. Compound (Sakura) on dry ice. Tissues were cut at 7 µm on a cryostat (Leica) using Superfrost slides (Thermo Fisher Scientific). For immunofluorescence, samples were washed with PBST (1 × PBS, 0.1% Triton X-100) before permeabilization with Antigen Unmasking Solution (Vector Laboratories, Inc. (H-3300), 1:100) at 80 °C for 20 min. Samples were then washed

twice with PBST and incubated in blocking solution (3% BSA, 5% goat serum, 0.2% Triton X-100 in PBS), then incubated with primary antibodies overnight at 4°C, followed by two PBST washes and incubation with secondary antibodies for 2 h at room temperature. Slides were washed again with PBST and mounted with mounting medium Vectashield with DAPI (Vector Laboratories, Inc.). The primary antibodies used were anti-Mef2C (Santa Cruz Biotechnology, catalog no. sc-313, 1:200) and anti-PCNA (Dako, catalog no. M0879, 1:500). The secondary antibodies were AlexaFluor 488 goat anti-rabbit IgG (Thermo Scientific, catalog no. A-11008, 1:500) and AlexaFluor 633 goat anti-mouse IgG (Thermo Scientific, catalog no. A-21050, 1:500). Confocal images were taken using an LSM880 microscope (Zeiss). The percentage of proliferating cardiomyocytes was calculated as a ratio of the number of (PCNA + Mef2C)⁺ particles to the total number of Mef2C⁺ particles in the injury area and in 200 μm of the injury border zone using ImageJ (v.2.1.0). Briefly, the total number of cardiomyocytes was analysed in the Mef2C⁺ channel, as the total number of particles that were segmented and counted using the particle analysis tool after thresholding and applying a despeckling filter. The number of (PCNA + Mef2C)⁺ particles was determined using the particle analysis tool after merging the green and red (identified using look-up tables) channels and selecting a colour threshold for yellow. Thresholding was verified visually using the original image. For each biological replicate, three nonconsecutive midsagittal sections were used.

Microscopy analysis of revascularization. Zebrafish ventricles were extracted from adults that had not been treated with morphine and fixed in 4% PFA for 1 h at room temperature. Ventricles were embedded in OCT (Tissue-Tek) and sectioned (8 μm). For each biological replicate, three nonconsecutive midsagittal sections were used. Immunostaining was performed as previously described⁶⁵. The primary antibodies used were anti-RFP (Rockland, catalog no. 600-401-379, 1:200) and anti-PCNA (Santa Cruz Biotechnology, catalog no. sc-56, 1:500). The secondary antibodies were AlexaFluor 488 goat anti-mouse IgG (H+L) (Thermo Scientific, catalog no. A-11029, 1:500) and AlexaFluor 647 goat anti-rabbit IgG (H+L) (Thermo Scientific, catalog no. A-21244, 1:500). Confocal images were obtained using an LSM700 microscope (Zeiss). The percentage of proliferating cECs (PCNA⁺) was calculated as the ratio of the total numbers of cECs in the injury area and in 200 μm of the injury border zone using ZEN 3.2 Blue software. For fluorescence intensity analysis, the percentage fluorescence was calculated from whole-mount images as a ratio to the background fluorescence in the injury area using Fiji software.

Wnt inhibition. To investigate the role of Wnt signalling in heart regeneration, the Wnt antagonist IWR-1-endo (Sigma-Aldrich) was used^{66,68}. IWR-1 was dissolved in dimethyl sulfoxide (DMSO) to prepare a 10 mM stock solution. Wild-type fish were injected intraperitoneally with 25 μl of 10 μM IWR-1 in PBS or DMSO (0.1% in PBS)⁶⁷. The injection was administered at 1 d.p.i. and 2 d.p.i. for the 3 d.p.i. analysis and, for later time points, once every 2 days from 2 d.p.i. until the day fish were euthanized.

Cloning and transgenesis of *-4kbcoll12a1a:GAL4VP16*. The possible 5'-upstream promoter region upstream of the predicted transcription start site of *coll12a1a* was identified using *Danio rerio* strain Tuebingen chromosome 17, GRCz11 Primary Assembly NCBI Reference Sequence: NC_007128.7, and served as a target template for primer design. Forward (5'-GAGAGAGAGAAAGCACCATTCTG-3') and reverse (5'-GTTTACACACACAGTCAGCAG-3') primers were designed using plasmid editor ApE and used to amplify the promoter region by PCR from genomic DNA of *D. rerio*. PCR was carried out in a 25 μl reaction using GoTaq Long PCR Mastermix (Promega, cat no. M4021) according to the manufacturer's protocol and TOPO cloned into pENTR5' (Invitrogen) to create pAM222. *Tg(-4kbcoll12a1a:GAL4VP16)* is a multisite gateway assembly of pAM222, Tol2kit no. 387 (pME-GAL4VP16), no. 302 (p3E_SV40polyA) and pAM57 (pDestTol2 *insulin:YFP*). To obtain pAM57, the *crystallin* promoter was exchanged with the minimal *insulin* promoter. First, using site-directed mutagenesis, pAM58, an empty pDestTol2 vector containing YFP with an added restriction site for insertion of any promoter of interest, was generated. Deletion primers were engineered by designing standard, nonmutagenic forward and reverse primers that flanked the region to be deleted (5'-GGTACCCGTAACACGACGCCAGTGAATTATC-3' and 5'-GATATCATGGTGAACAAGGGCGAGGAGCTGTTC-3', respectively). In the forward primer, the sequence for the restriction site EcoRV was included. The minimal *insulin* promoter was amplified from genomic DNA and inserted using In-Fusion Cloning technology (Takara). The primers were designed using the In-Fusion Cloning Primer Design Tool (www.takarabio.com): minimal *insulin* promoter forward primer 5'-TTTTACGGTACCGATCTTCAGCCACA GTCTAGTTTAG-3' and reverse primer 5'-CTTGCTCACCATGATCGAAGCAG AGCCGAGGAATG-3'. PCR was carried out in a 25 μl reaction using Clone Amp HiFi PCR premix. The In-Fusion HD cloning reaction was performed with cleaned products of pDestTol2 *GATATC-YFP* (EcoRV-digested) and minimal *insulin* promoter PCR product using In-Fusion HD enzyme premix. The minimal *insulin* promoter is ideal as a selection transgenesis marker because of its small size and very early activity (from 24 h postfertilization; h.p.f.).

For Tol2-mediated zebrafish transgenesis, 25 ng μl⁻¹ Tol2 mRNA was injected with 25 ng μl⁻¹ plasmid DNA. F0 founders were screened for specific *insulin:YFP*

expression, raised to adulthood and screened for germline transmission. The activity of the isolated 4-kb upstream region of zebrafish *coll12a1a* corresponded to that of the endogenous *coll12a1a* expressed in embryos of various stages (Supplementary Fig. 12a). In uninjured adult hearts, RFP expression in *coll12a1a>NTR:RFP*-labelled cells colocalized with endogenous *coll12a1a* expressed in the epicardium (Supplementary Fig. 12b). As expected, cryoinjury transiently induced both endogenous *coll12a1a* and RFP expression in the injury area (Supplementary Fig. 12c).

Genetic ablation of *coll12a1a*-expressing cells using the NTR/MTZ system. To investigate the role of *coll12a1a*-expressing fibroblasts in heart regeneration, we used the fish line *Tg(-4kbcoll12a1a:GAL4VP16; UAS:NTR:RFP)*. The functionality of the cell ablation approach was verified in 48 h.p.f. embryos by treatment with a 5 mM solution of MTZ (Sigma-Aldrich) dissolved in DMSO that led to formation of RFP⁺ aggregates (Supplementary Fig. 12d), indicative of cell death as previously reported¹⁷. To ablate *coll12a1a*-expressing cells in adult hearts, the injured fish were immersed 2 d.p.i. in a system of water containing 10 mM MTZ (or 0.2% DMSO as a vehicle control) and kept in the dark. Water with the fresh drug was exchanged twice a day until the day fish were euthanized. Fish were treated with MTZ from day 2 until day 6 for the 7 d.p.i. time point and also from day 14 to day 16 for the 30 d.p.i. time point. We observed increased mortality in *coll12a1a>NTR:RFP*-ablated injured fish compared with DMSO-treated controls. Additional experiments would be needed to determine whether this mortality is related to heart regeneration defects.

Cre-lox lineage tracing. For epicardial Cre-lox lineage tracing, *TgBAC(cryaa:EGFP; tcf21:Cre-ERT2; -3.5ubi-loxP-EGFP-loxP-mCherry)*-transgenic fish were treated with 10 μM 4-OHT (Sigma-Aldrich) dissolved in 100 ml of water per fish, 4 and 3 days before cryoinjury for 12 h as described¹⁷. Before systemic administration, the 10 mM stock (dissolved in ethanol) was heated for 10 min at 65°C. Hearts were harvested 7 d.p.i. For endocardial Cre-lox lineage tracing, *Tg(fli1:Cre-ERT2; -1.5hsp70l:loxP-STOP-loxP-EGFP; cryaa:Venus)*-transgenic fish were treated as embryos at 24 h.p.f. with 5 μM 4-OHT (Sigma-Aldrich) for 24 h and heat shocked for 1 h at 37°C. Adult fish were injured and hearts were harvested at 7 d.p.i.

Statistics and reproducibility. Data collection and analysis were not performed blind to the conditions of the experiments. Sample collection and genomics analyses were performed by two independent experimenters. Sample sizes (*n*) are indicated either in each figure legend including the number of independent experiments or in this section. Quantification of each experiment is described in the respective section of the Methods together with the software used. No data points were excluded from the analysis. Graphs and statistical analyses shown in Figs. 3b,j and 6b,g,i and in Supplementary Figs. 12f and 24b were performed in GraphPad Prism (v.7 and v.9). Graphs show mean and s.d.; the statistical tests with *P* values are indicated in figure legends. The experiments in Fig. 3g used DMSO-treated fish (*n* = 3) and MTZ-treated fish (*n* = 6) at 7 d.p.i.; *n* represents biologically independent samples from two independent experiments. The experiments in Fig. 3i used DMSO-treated (*n* = 4) and MTZ-treated (*n* = 3) fish; results are from two independent experiments. All RNAscope experiments in Figs. 3b and 6c,e and in Supplementary Figs. 7, 8 and 12 were performed at least three times using one or two independent experiments, unless stated otherwise. Fig. 4e (*n* = 3) shows the results of one independent experiment, Fig. 5e (*n* = 2) shows the results of one independent experiment; for the experiments in Fig. 6c,d, *n* = 9, 3, 9, 9 animals; Fig. 6b, lower right panel: DMSO was injected at 3 d.p.i., *n* = 6; 7 d.p.i., *n* = 9, 15 d.p.i., *n* = 3, and 30 d.p.i., *n* = 8, and IWR-1 was injected at 3 d.p.i., *n* = 5; 7 d.p.i., *n* = 9; 15 d.p.i., *n* = 3, and 30 d.p.i., *n* = 8; *n* represents biologically independent samples over four independent experiments. In Fig. 6g, *n* represents biologically independent samples over two independent experiments; in Fig. 6i, *n* represents biologically independent samples over two independent experiments. The results in Supplementary Fig. 3d are based on *n* = 2 animals in one independent experiment, those in Supplementary Fig. 24c (*n*₅₅ = 3, *n*₁₀₅ = 3, *n*₁₅₅ = 3, *n*₂₀₅ = 4, *n*₂₅₅ = 3) are based on two independent experiments. The number of animals was determined based on previous literature^{6,10,17,41,68,69} for the experiments in Figs. 3b, 4e, 5e and 6f-1 and in Supplementary Figs. 3d, 7, 8 and 12b,c or calculated using G*power calculator⁷⁰ by two-tailed *t* test with α err prob = 0.05 and power(1 - β err prob) = 0.80; effect sizes *f* were determined as follows: Fig. 3g-j and Supplementary Fig. 12e,f: *f* = 2; Fig. 6b,c,e: *f* = 1.5; Supplementary Fig. 24c: *f* = 1.48.

Mapping and filtering of single-cell scar data. Scar filtering was performed as described by Spanjaard et al.²⁴. Briefly, scar sequences were aligned using *bwa mem*³ (v.0.7.12) to a reference of dTomato. Valid cell barcodes were identified based on the single-cell transcriptome data (see previous paragraph). We removed reads that were unmapped, had an incorrect barcode or did not start with the exact PCR primer we used. We truncated all scar sequences to 75 nucleotides and removed shorter sequences.

In further filtering steps, we consecutively filtered out scars in order to remove all scars that did not fulfill the following quality control conditions. As a first step, we required all molecules to be sequenced at least twice to remove straightforward sequencing errors; this also reduced complexity for the subsequent filtering steps.

To remove molecules caused by incorrect annealing, we selected only the most prevalent scar for each combination of barcode and unique molecular identifier (UMI), only the most prevalent UMI for each barcode and scar, and only the most prevalent barcode for each UMI and scar. We then assessed similar (Hamming distance of 2 or less) scar sequences in each cell, only keeping both if they met criteria designed to test whether both sequences were verifiably correct (see Spanjaard et al.²⁴ for details). Finally, we removed cells that looked like doublets by comparing their numbers of different scars with the scar number distribution for that cell type.

Tree building. Tree building was done as described in Spanjaard et al.²⁴. Briefly, the lineage tree building algorithm was designed to build lineage trees with high numbers of cells, despite the low detection rate inherent in single-cell sequencing. It consisted of two steps: we first determined the lineage tree in terms of scars created and afterwards placed the single cells in their appropriate positions in the tree.

To build the scar lineage, we used the data set of scars in single cells and determined which scars could be found in the same branch of the lineage trees: those scars were observed together in the same cell. We created a graph of scars with connections between scars, indicating that they were together in a branch. A scar that had been created early would be present in many branches and connected to many other scars. As a corollary, the scar that was connected to most other scars in this graph was created before all other scars. This scar became the root of the lineage tree; it was removed from the scar graph and we identified the most highly connected scar in the resulting graph. This scar was placed underneath the first scar in the lineage tree and was again removed from the graph. If at any point the graph split into disconnected parts after removing a scar, those parts were considered to be distinct branches of the lineage tree and were treated separately. This process was repeated until all scars had been placed in the lineage tree.

The above tree construction algorithm is valid for high, nonvariable scar-detection rates with perfect filtering and no doublets; for more realistic scenarios, computations are more complicated but follow the general scheme as outlined above. Refer to Spanjaard et al. for details²⁴.

Once the shape of the lineage tree and the location of all scars in the lineage tree had been determined, we placed the cells in the tree. Cells were placed using the scars that were found in them, at the lowest possible position in the tree. Importantly, this means that some cells may have been placed in leaf nodes, but typically some cells in which a leaf node scar had not been detected would be placed in internal nodes.

Lineage determination. To determine lineage relationships between cell types, we contracted all vertical branches (that is, removed all nodes that were not part of a lineage split) and then calculated cell type ratios – the amount of cells of a specific type divided by the node size – for each node. We calculated the lineage similarity between two cell types as the correlation between their cell type ratios, weighted by the size of the nodes. Importantly, this approach allowed us to integrate lineage data for multiple biological replicates. Using cell type ratios emphasizes nodes that are lower in the trees and contain fewer cell types, and weighting by node size emphasizes larger nodes to maximize reproducibility.

We determined the stability of clusters in two different ways. A thousand-fold repetition of downsampling all lineage trees to half of the cells was used to estimate the stability of lineage clusters against low cell numbers. After calculating the correlations and clustering the cell types, we recovered the same clusters in the majority of cases, indicating that these clusters were not influenced by low cell numbers (Supplementary Fig. 19). A thousand-fold repetition of downsampling all lineage trees to half of the nodes was used to estimate the stability of lineage clusters against low clonality or clone-specific transition rates. After calculating the correlations and clustering the cell types, we found that *nppc*, *spock3*, *cfb* and valve fibroblasts were attributed to varying clusters, indicating that differences between clones were too large for the correlation analysis to yield reliable answers (Supplementary Fig. 22).

To determine possible sources of *nppc*, *spock3*, *cfb* and valve fibroblasts, we determined the overlap of cell types with these target cell types in tree leaf nodes. We calculated the conditional probability of cell types as the number of leaf nodes that contained cells of both source and target type divided by the number of leaf nodes that contained cells of the target type. For example, at 7 d.p.i., ventricular endocardium had a conditional probability of 1 with regard to *nppc* fibroblasts, meaning that at 7 d.p.i., every leaf node that contained *nppc* fibroblasts also contained ventricular endocardium. However, *nppc* fibroblasts had a conditional probability of 0.2 with regard to the ventricular endocardium, indicating an asymmetric lineage relationship between these cell types.

Trajectory analysis and RNA velocity. We used PAGA, implemented in scanpy³⁸ (v.1.8.2) and scvelo⁷¹ (v.0.2.4) for integrated trajectory and RNA velocity analysis. First, we extracted velocity information from reads aligned to a transcriptome based on GRCz11, release 92, using Cell Ranger 3.1 and velocity³⁹. Using cell type annotations determined previously, we selected relevant cells; regressed out total counts, numbers of genes and mitochondrial percentages; determined highly variable genes and principal components; and used bbknn to integrate data sets. We then reclustered the data using the Leiden algorithm and ran PAGA on the

Leiden clusters. For the integrated data set, we calculated RNA velocities using the stochastic model implemented in scvelo and plotted those on the UMAP representation of the data.

Secretome analysis. We performed an ortholog conversion of the Vertebrate Secretome Database⁷² using orthology data from the Alliance of Genome Resources, release 3.2.0 (ref. ⁷³). For the ortholog conversion, we used all gene pairs with 'IsBestScore=Yes' and kept all unique zebrafish gene names. After normalizing all cells to 10,000 transcripts, we calculated the percentage of transcripts that were in the transcriptome for each cell type at each time point. We then calculated $\log(1+x)$ of the average expression of secretome genes per cell type and used the clustermap function in seaborn to perform hierarchical clustering using Ward's method for the cell types and secretome genes that were expressed at an average of 10 in at least one cell type. We similarly performed hierarchical clustering of cell type averages of z transformed secretome expressions (capped at 5) for genes that had a z score greater than 2 in at least one cell type. We excluded neuronal cells and myelin cells from these analyses.

Interaction analysis. We used CellPhoneDB⁷⁴ (v.2.1.4) to analyse interactions between cell types at each time point after ortholog conversion of the genes in CellPhoneDB, similar to the conversion of the secretome. We included all cell types with more than 50 cells or whose cell count proportion of the total data set exceeded 0.01 and counted the number of significant outgoing and incoming interactions for each cell type.

tomo-seq data deconvolution. We used AutoGeneS²⁹ (v.1.0, <https://github.com/theislab/AutoGeneS>) to deconvolve the tomo-seq data. As input, we used the single-cell data constrained to the 5,000 most variable genes using scanpy⁷⁵ (pp.highly_variable_genes). AutoGeneS was then used to select a total of 400 informative genes from among the highly variable ones that differentiated the cell types. The highly variable genes were selected based on normalized dispersion. The proportions were inferred using nonnegative least squares⁷⁶ (scipy.optimize.nnls) based on the cellular mean expression of the informative genes. For each bulk sample, negative proportions were set to zero and the rest were normalized to sum to one.

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

Data availability

Source data are provided with this paper. Sequencing data have been deposited in the Gene Expression Omnibus with accession numbers GSE159032 and GSE158919. Transcriptome data were aligned to a zebrafish transcriptome created with Cell Ranger 3.0.2 from GRCz11, release 92. We performed an ortholog conversion of the Vertebrate Secretome Database VerSeDa using orthology data from the Alliance of Genome Resources, release 3.2.0.

Code availability

Custom code is provided at https://github.com/Bastiaanspanjaard/Heart_regeneration.

References

- Aleström, P. et al. Zebrafish: housing and husbandry recommendations. *Lab. Anim.* **54**, 213–224 (2020).
- Mosimann, C. et al. Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169–177 (2011).
- Felker, A. et al. Continuous addition of progenitors forms the cardiac ventricle in zebrafish. *Nat. Commun.* **9**, 2001 (2018).
- Davison, J. M. et al. Transactivation from Gal4-VP16 transgenic insertions for tissue-specific cell labeling and ablation in zebrafish. *Dev. Biol.* **304**, 811–824 (2007).
- Bussmann, J. et al. Arteries provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk. *Development* **137**, 2653–2657 (2010).
- González-Rosa, J. M. & Mercader, N. Cryoinjury as a myocardial infarction model for the study of cardiac regeneration in the zebrafish. *Nat. Protoc.* **7**, 782–788 (2012).
- Chablais, F., Veit, J., Rainer, G. & Jazwińska, A. The zebrafish heart regenerates after cryoinjury-induced myocardial infarction. *BMC Dev. Biol.* **11**, 21 (2011).
- Lelek, S. et al. Morphine alleviates pain after heart cryoinjury in zebrafish without impeding regeneration. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.10.01.322560> (2020).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
- Millard, S. M. et al. Fragmentation of macrophages during isolation confounds analysis of single cell preparations from mouse hematopoietic tissues. *Cell Rep.* **37**, 110058 (2021).

65. Marín-Juez, R. et al. Fast revascularization of the injured area is essential to support zebrafish heart regeneration. *Proc. Natl Acad. Sci. USA* **113**, 11237–11242 (2016).
66. Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smothened activity. *Proc. Natl Acad. Sci. USA* **99**, 14071–14076 (2002).
67. Kinkel, M. D., Eames, S. C., Philipson, L. H. & Prince, V. E. Intraperitoneal injection into adult zebrafish. *J. Vis. Exp.* **30**, 2126 (2010).
68. Koth, J. et al. Runx1 promotes scar deposition and inhibits myocardial proliferation and survival during zebrafish heart regeneration. *Development* **147**, dev186569 (2020).
69. Lowe, V. et al. Neuropilin 1 mediates epicardial activation and revascularization in the regenerating zebrafish heart. *Development* **146**, dev174482 (2019).
70. Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* **41**, 1149–1160 (2009).
71. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
72. Cortazar, A. R., Oguiza, J. A., Aransay, A. M. & Lavin, J. L. VerSeDa: vertebrate secretome database. *Database* **2017**, baw171 (2017).
73. Alliance of Genome Resources Consortium. Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Res.* **48**, D650–D658 (2020).
74. Efremova, M. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 25 (2020).
75. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
76. Lawson, C & Hanson, R.J. *Solving Least Squares Problems* (SIAM, 1987).

Acknowledgements

We acknowledge support by MDC/BIMSB core facilities (zebrafish, light microscopy, genomics, bioinformatics), and we thank J. Richter for help with zebrafish experiments, K. Poss (Duke University), N. Mercader (University of Bern) and C. Mosimann (University of Colorado) for sharing the *tgf21:Cre-ERT2*, *fli1:Cre-ERT2* and *hsp70l:Switch* transgenic lines, and S. Sawamiphak (Max Delbrück Center, Berlin) for Mef2c and PCNA antibodies. Work in D.P.'s laboratory was supported by the DZHK (Berlin partner site project 81Z0100103) and by the Helmholtz Association (ERC-RA-0047). Work in J.P.J.'s laboratory was funded by a European Research Council Starting Grant (ERC-StG

715361 SPACEVAR), a Fondation Leducq Transatlantic Networks Grant (16CVD03), a Helmholtz Incubator grant (Sparse2Big ZT-I-0007) and the DZHK (Berlin partner site project 81Z0100112). B.H. was supported by a PhD fellowship from Studienstiftung des deutschen Volkes. Work in D.Y.R.S.'s laboratory was supported in part by funds from the Max Planck Society, the DFG (project number 394046768) SFB 1366/project A4 and the Leducq Foundation. H.E.-S. was funded by a Boehringer Ingelheim Fonds PhD fellowship.

Author contributions

B.H., S.L., B.S., D.P. and J.P.J. conceived and designed the project. B.H. performed single-cell experiments and mRNA and cell type analysis, with support from J.M. and B.S. S.L. and M.G.S. did cryoinjury experiments. S.L. and A.M.M. generated and characterized the *col12a1a>NTR:RFP* line. S.L. performed and analysed histology, RNAscope, Cre–lox lineage tracing, pathway inhibition experiments and ablation experiments. B.S. performed secretome and cell–cell interaction analyses, built single-cell lineage trees, developed tree analysis methods, and performed trajectory and RNA velocity analyses. H.E. performed microscopy analysis of revascularization, with guidance from D.Y.R.S. H.A. adapted bulk deconvolution to tomo-seq data, with guidance from B.S. and F.T. D.P. and J.P.J. guided experiments, and J.P.J. guided data analysis. B.H., B.S. and J.P.J. wrote the manuscript, in close interaction with S.L. and D.P. and with input from all other authors. All authors discussed and interpreted results.

Funding

Open access funding provided by Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-022-01129-5>.

Correspondence and requests for materials should be addressed to Daniela Panáková or Jan Philipp Junker.

Peer review information *Nature Genetics* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

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 our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

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

n/a Confirmed

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Zen 2.3 SP1 FP3 (version 14.0.0.0) and Zen 3.2 blue edition were used to acquire confocal images. BIOREVO Keyence BZ-9000 Viewer was used to acquire histological images.

Data analysis

Cell Ranger (version 3.0.2) was used to align and demultiplex single-cell transcriptomics data. bwa mem (version 0.7.12) was used to map scar data. Velocity (version 0.17) was used to extract mRNA velocity data from single-cell transcriptomics data. All data analysis was done in R (version 3.3 and higher), except when indicated. The R package Seurat (version 3.0) was used to cluster single-cell transcriptomics data. Interactions between cell types were calculated using CellPhoneDB (<https://github.com/Teichlab/cellphonedb>) (version 2.1.4) and further analyzed in custom scripts provided at https://github.com/Bastiaanspanjaard/Heart_regeneration. Analysis of cellular trajectories and mRNA velocity were done in Python (3.8) using scanpy (version 1.8.2) and scvelo (version 0.2.4). Deconvolution of TOMO-seq data was done in Python (3.8) using AutoGeneS (<https://github.com/theislab/AutoGeneS>) (version 1.0). We developed custom code to filter, analyze and visualize scar data. The custom code is provided at https://github.com/Bastiaanspanjaard/Heart_regeneration. Microscopy data was analyzed in Fiji/ImageJ including the Particle analyzer plug-in (version 2.1.0) and Adobe Photoshop (version 20.0.6). GraphPad Prism 7/9 (version 7.0d/9.3.1) was used for plotting graphs and statistical analysis.

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

Sequencing data are deposited on Gene Expression Omnibus, accession numbers GSE159032 and GSE158919. Transcriptome data was aligned to a zebrafish transcriptome created with Cell Ranger 3.0.2 from GRCz11, release 92. We performed an orthologue conversion of the Vertebrate Secretome Database VerSeDa, using orthology data from the Alliance of Genome Resources, release 3.2.0.

Field-specific reporting

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

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	The number of cells to be analyzed was based on the maximum number attainable. Since the various detected cell types have vastly different abundances, we focused the analysis on those cell types that were detected at high enough numbers for statistical analysis. Sample size for animal test groups was determined using G*Power calculator (G*Power 3.1).
Data exclusions	No data was excluded. However, the following filtering steps were applied to the scar sequencing data in order to mitigate the effect of sequencing errors, PCR amplification errors and formation of chimeric reads: removal of all sequences that were detected only once; removal of molecules with the same BC+UMI but different scar, same BC+scar but different UMI, and same UMI+scar but different BC; removal of scars with a Hamming distance ≤ 2 to a more highly detected scar; removal of the most lowly detected scars; and removal of cells that express more scars than expected (putative doublets). For lineage analysis, only scars with creation probability < 0.01 (as determined by bulk experiments) were used in order to minimize artifacts caused by repeated creation of the same scar sequence in one animal.
Replication	All attempts at replication were successful. The number of replicate experiments is provided in figure legends and/or the section "Statistics and reproducibility" in the Methods
Randomization	Formal randomization of samples was not applicable, since no manual selection of single cells was performed during experiments and analysis. Animals were randomly assigned to test groups.
Blinding	Formal blinding was not applicable, since no manual selection of single cells was performed during experiments and analysis. Data collection and analysis were not performed blind to the conditions of the experiments. Sample collection and genomics analysis were performed by two independent experimenters.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

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

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Zebrafish (<i>Danio rerio</i> , strain AB). Adult zebrafish of random sex, aged between 4 months and a year, and with a length of at least 3 cm were used.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field collected samples were used in this study.
Ethics oversight	Zebrafish were bred, raised and maintained in accordance with the FELASA guidelines, the guidelines of the Max-Delbrück Center for Molecular Medicine and the Max Planck Society, and the local authorities for animal protection (Landesamt für Gesundheit und Soziales, Berlin, and The Veterinary department of the Regional Board of Darmstadt, Darmstadt, Germany) for the use of laboratory animals based on the current version of German law on the protection of Animals and EU directive 2010/63/EU on the protection of animals used for scientific purposes. In addition, housing and breeding standards followed also the international 'Principles of Laboratory Animal Care' (NIH publication no. 86-23, revised 1985).

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