Single-cell dissection of live human hearts in ischemic heart disease and heart failure reveals cell-type-specific driver genes and pathways

Suvi Linna-Kuosmanen[1,2,3,*✉], Eloi Schmauch[1,2,3,*], Kyriakitsa Galani[1,2], Carles A. Boix[1,2], Lei Hou[1,2], Tiit Örd[3], Anu Toropainen[3], Lindsey K. Stolze[4], Elamaran Meibalan[2,10], Julio C. Mantero[1,2], Ashley Renfro[1,2], Johannes Ojanen[3], Leandro Z. Agudelo[1,2], Maija Hollmen[5], Juho Jalkanen[6], Jarmo Gunn[6,7], Pasi Tavi[3], Casey E. Romanoski[4], Calum A. MacRae[2,8,9,10], Minna U. Kaikkonen[3], Guillermo Garcia-Cardena[2,9,11], Tuomas Kiviniemi[6,7,8], and Manolis Kellis[1,2✉]

[1] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA. [2] Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA. [3] A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, 70211 Kuopio, Finland. [4] Department of Cellular and Molecular Medicine, The College of Medicine, The University of Arizona; Tucson, AZ 85721, USA. [5] Medicity Research Laboratories, University of Turku, 20500 Turku, Finland [6] Heart Center, Turku University Hospital, 20521 Turku, Finland. [7] Department of Medicine, University of Turku, 20500 Turku, Finland. [8] Cardiovascular Medicine and Network Medicine Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [9] Harvard Stem Cell Institute, Cambridge, MA 02138, USA. [10] Department of Medicine, Harvard Medical School, Boston, MA 02115, USA. [11] Center for Excellence in Vascular Biology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. [*] These authors contributed equally: Suvi Linna-Kuosmanen, Eloi Schmauch. ✉ Corresponding authors, email: suvil@mit.edu, manoli@mit.edu

Abstract

Ischemic heart disease is the single most common cause of death worldwide with an annual death rate of over 9 million people. Genome-wide association studies have uncovered over 200 genetic loci underlying the disease, providing a deeper understanding of the causal mechanisms leading to it. However, in order to understand ischemic heart disease at the cellular and molecular level, it is necessary to identify the cell-type-specific circuits enabling dissection of driver variants, genes, and signaling pathways in normal and diseased tissues. Here, we provide the first detailed single-cell dissection of the cell types and disease-associated gene expression changes in the living human heart, using cardiac biopsies collected during open-heart surgery from control, ischemic heart disease, and ischemic and non-ischemic heart failure patients. We identify 84 cell types/states, grouped in 12 major cell types. We define markers for each cell type, providing the first extensive reference set for the live human heart. These major cell types include cardiovascular cells (cardiomyocytes, endothelial cells, fibroblasts), rarer cell types (B lymphocytes, neurons, Schwann cells), and rich populations of previously understudied layer-specific epicardial and endocardial cells. In addition, we reveal substantial differences in disease-associated gene expression at the cell subtype level, revealing t arterial pericytes as having a central role in the pathogenesis of ischemic heart disease and heart failure. Our results demonstrate the importance of high-resolution cellular subtype mapping in gaining mechanistic insight into human cardiovascular disease.

Introduction

Ischemic heart disease affects an estimated 126 million people globally and is the leading cause of death

worldwide[1]. The underlying atherosclerotic process involves multiple cell types, and though the principal focus of the work to date has been on endothelial, smooth muscle, and immune cells in the arterial wall, the involvement of the microvasculature, hepatocytes, muscle cells and adipocytes is also acknowledged. The mechanisms involved cause fundamental defects in myocardial physiology through inflammation, atherosclerotic lesion formation and impaired microvascular function, ultimately leading to myocardial ischemia, myocardial infarction and heart failure[2,3]. Although multiple risk factors for ischemic heart disease have been described, the proximate causal mechanisms remain largely unknown, thereby hampering the design of safe and effective therapeutics.

Recent genetic studies provide invaluable resources for understanding the mechanistic basis of cardiovascular disease and for discovering new targets for therapeutics[4–6]. Nonetheless, definitive mechanistic insights remain elusive, as 90% of Genome-wide Association Study (GWAS) hits do not affect protein coding regions directly, but concentrate in gene-regulatory regions, making it difficult to identify their target genes and cell types of action[7,8]. Moreover, currently-discovered loci explain only a small fraction of disease heritability, indicating that many additional contributors exist and that new approaches are needed for their discovery[8–10]. Multiple studies and ongoing collaborations have examined gene expression, epigenomics and genetic variation across individuals and helped to reveal the regulatory code of our genome, demonstrating that expression of nearly all genes is impacted by genetic variation, most of which is proximal to the affected gene[9–12]. However, current data originate predominantly from *postmortem* samples that fail to adequately capture the *in vivo* properties of living tissue and from bulk tissues, which represent averaged signals over complex and heterogeneous populations of cells, thus masking the cellular heterogeneity and cell-type-specific genetic effects that likely underlie complex disorders. To trace cell-type-specific altered gene expression patterns to their upstream regulatory regions and to the genetic variants that impact them, we need to detect altered gene activity at single cell resolution.

Ischemic heart disease, whether it is caused by atherosclerosis or microvascular dysfunction, affects the function of the heart as a whole. The right atrium, in addition to liver, coronary arteries and adipose tissue was shown to capture significant proportions of the genetic signal from ischemic heart disease in the recently published EpiMap analysis[11] highlighting the importance of the discrete tissue compartments as a source of meaningful information content on the core biology. While single-cell dissection of coronary arteries has been addressed in several previous studies in both human and mouse[12], right atrium in disease context remains unexplored. To resolve the effects of the underlying biology more broadly, we need to collect data also from other implicated sites. Here, we focus on living right atrial tissue, which contains many of the cell types and the genetic components affecting coronary microvascular function in ischemic heart disease, but also captures orthogonal information from the surrounding tissue and its various cell types.

We present a comprehensive map of the right atrium in ischemic heart disease and heart failure providing, to our knowledge, the first map from the live human heart. To ensure the capture of both cardiomyocyte and non-cardiomyocyte populations, we used single-nuclei RNA-seq (snRNA-seq) to identify the main cell types and cell subtypes. We determined the changes in cell function in ischemic heart disease compared to control and

followed the changes to ischemic heart failure. In addition, we characterized the differences between ischemic and non-ischemic heart failure. To trace genetic signals to active regions and affected genes in the cell types and cell subtypes, we utilized enhancer-gene links based on correlations between enhancer activity and gene expression across different tissues and cell types and links based on correlations across individuals from heart tissue in GTEx and enhancing GTEx (eGTEx). Our findings demonstrate the importance of high-resolution subtype mapping and genetic partitioning in identifying candidate mechanisms underlying human cardiovascular disease.

Results

The first expression map from the live human heart

We obtained live (*ex vivo*) cardiac tissue biopsies from 13 donors collected during open-heart surgery from the right atrial appendage of 5 control, 4 ischemic heart disease, 2 ischemic heart failure and 2 non-ischemic heart failure patients (i.e., pressure-overload heart failure caused by aortic stenosis) (**Fig. 1a**), with each sample containing all layers of the heart wall, namely, epicardium, myocardium and endocardium. We collected nuclei successfully from all samples and profiled them using snRNA-seq by 10X Genomics, resulting in 76,087 nuclei and 84 cell types/states, grouped in 12 main cell types (**Fig. 1b**) well represented across all samples (**Fig. 1c**). To annotate the main cell types, we used available marker genes from previous single-cell studies from embryonic, postnatal and adult hearts[13–16] and enrichment of Gene Ontology (GO) biological processes (**Fig. S1**).

In contrast to previous studies, endocardial endothelial cells and epicardial mesothelial cells were among the most abundant cell types comprising 23% and 14% of the cells, respectively, compared to 6.8% and 0.5% in [13]. The distribution of other cell types corresponded to expected ratios in the cardiac tissue with cardiomyocytes comprising 27%, fibroblasts 17%, vascular endothelial cells 10%, macrophages 3%, pericytes 1.9%, lymphocytes 1%, Schwann cells 1%, smooth muscle cells 0.6%, adipocytes 0.6%, and neurons 0.2% of the cells. Partially, the capture of all these cell types stems from the sample collection procedure, which prioritized the collection of all three layers of the heart wall in each sample (**Fig. 1a** and **Fig. S2**). In addition, optimization of the data processing increased the quality of the captured signal significantly (**Fig. S3**).

Furthermore, samples collected from live tissue may preserve cell types better compared to tissue collected after death (*postmortem)*, as live tissue shows better cell enrichment, significantly stronger expression signals, and better separation of all cell types and cell subtypes than tissue collected after brain or cardiovascular death (**Fig. 1d-e** and **S4**). The most affected populations in the *postmortem* tissue were endocardial and vascular endothelial cells, which form the inner layers of the heart and blood vessels, respectively, and epicardial mesothelial cells, which form the outer layer of the heart (**Fig. 1e-f**), suggesting that the results from these cell populations are most compromised in *postmortem* tissue.

The marker genes for the main cell types are presented in **Table S1**. The cells expressed several previously identified markers (**Fig. S1a**), in addition to novel genes that could aid in the experimental dissection of the cell signaling and improve targeting of the cells *in vitro* and *in vivo* (**Fig S5-12**). Endocardial endothelial cells

expressed high levels of *PCDH7, PCDH15, NPR3, LEPR,* and *POSTN* in addition to genes that were also expressed in the vascular endothelium, such as *PECAM1, EMCN, ERG, VWF*, and *PTPRB* (**Fig. 1g**). For vascular endothelial cells, the most specific marker genes were *GRB10, ADGRL4, PREX2,* and *CYYR1.* Pericytes enriched for *RGS5, ABCC9, DACH1, SEMA5A, LHFPL6*, and *CARMN*, and epicardial mesothelial cells for *BNC2, ITLN1, C3, SLC4A4, SULF1, GPM6A, EZR*, and *PRG4*.

Vascular subtypes drive disease progression

In addition to ample endocardial and mesothelial cell populations, we identified unprecedentedly rich populations of vascular cell subtypes, altogether 12 populations (18% venous, 12% arteriole, 13% arterial capillary 1, 10% coronary, 8% arterial capillary 2, 8% disease enriched arterial capillary 3, 6% venous capillary 4, 3% lymphatic and 0.7% inflammatory endothelial cells, in addition to 11% pericytes 1, 5% of arterial pericytes 2, and 5% smooth muscle cells, **Fig. S13a**) based on a combination of known and novel marker genes (**Fig. 2a-b and Fig. S9**).

Of note, only a subset of previously identified markers[17] was found useful in the living human heart tissue (**Fig. S13b**), which may be due to differences between species, *ex vivo* vs. *postmortem* tissue, and sample processing times and procedures, which may affect the transcriptome. The enriched marker genes for the vascular subtypes are presented in **Fig. S9 and Table S2**. For coronary endothelial cells, the strongest markers were *GJA5, EFNA5, BMX, DKK2,* and *VEGFC*, in addition to general arterial markers, *EPAS1*, *EFNB2*, *SOX13*, *PCSK5*, and *DLL4*. The best markers for Capillary 1 were *RGCC* and *CA4*, for Capillary 2 they were *PRDM1*, *COLEC12* and *PDGFD*, and for Capillary 4 *RGCC* and *AQP1*. Capillary 3 lacked clear unique marker genes, but expressed all capillary markers. Lymphatic endothelial cells expressed *FLT4*, and inflammatory endothelial cells *ICAM1*, *CCL2*, *CXCL2*, *CXCL1*, and *CXCL10* (**Fig 2b**).

Comparison of the gene expression changes between the patient groups confirmed that disease-relevant changes can also be observed in the cells of the right atrium (**Fig. 2c-f,** and **Tables S3-S11**). Overall, the changes detected in the right atrial vasculature in ischemic heart disease suggest increased apoptosis/necrosis of the cells and decreased angiogenesis (**Fig. 2c**). The main pathways involved in these functions were apoptosis, and Hippo, Wnt/b-catenin, and b-adrenergic signaling (**Fig. 2c**), which may collectively be contributing to a dysfunctional vascular wall and increased ischemia of the surrounding tissue. The main driver cell populations that were expressing atherosclerotic disease signaling were pericytes, coronary artery endothelial cells, arterioles, and capillary 3 population. All these cell subtypes showed increased expression of pro-apoptotic genes and downregulation of anti-apoptotic genes, with strongest signal in coronary endothelium (**Fig. 2e**). In addition, capillary 3 profile indicatd increased endothelin 1 production and induction of fibrogenesis, whereas arterioles showed increased senescence, inflammatory (CXCR4 and chemokine signaling), and endothelin signaling. Further analysis of these pathways documented a reduced expression of the transcription factor KLF2 in the coronary and arteriolar endothelial cells in ischemic heart disease compared to control (**Fig. S14**), strongly indicating a dysfunctional state in the vascular endothelium of the ischemic heart disease patients[18].

In ischemic heart failure, pro-apoptotic signaling was replaced with increased viability and survival in all

populations, except smooth muscle cells and coronary artery and capillary 4 endothelial cells (**Fig. 2c**). All populations showed signatures of increased angiogenesis and/or vascular development and remodeling. However, the driver cell types, namely pericytes and capillary endothelial cells, were also promoting atherogenesis through senescence, inflammation, platelet adhesion/aggregation and thrombosis (**Fig. 2d,e**), thereby further contributing to vascular dysfunction and unstable neovascularization.

The changes observed in non-ischemic heart failure resembled ischemic heart failure but were more intense and evident across all cell populations. The strongest driver effects were measured in the pericyte 2 population and the weakest in coronary artery endothelial cells (**Fig. 2c,d**), which, according to their gene expression signature, were transitioning from endothelial to a mesenchymal phenotype (**Fig. 2e**). Overall, signaling promoting cell movement, invasion, inflammation, vascular development and remodeling was especially pronounced. Furthermore, endothelial or epithelial to mesenchymal transition, redox stress response, and tumorigenic signaling were strong across all vascular cell populations. Importantly, we identified pericytes as strong drivers of the disease-related changes, demonstrating anti-angiogenic signaling through VEGFA, EDN1, EGF, HIF1A suppression for ischemic heart disease, pro-inflammatory signaling through NFkb, TNF, IRF, IL1B activation for ischemic heart failure, and tumorigenic network of increased HIF1A, TGFB, RHOA, and YAP signaling for non-ischemic heart failure (**Fig. 2f**).

Subpopulations capture disease signals

For cardiomyocytes, endocardial endothelial cells, fibroblasts, and mesothelial cells, we analyzed differential gene expression at the level of all cells, subpopulations, and disease-enriched clusters for those populations that had clear shift between control and disease cell locations in the UMAP embedding (**Fig. 3** and **S15-17**). Full results of the differential gene expression and disease/function enrichments are presented in **Tables S18-S42**.

Enriched signals in “all cells” analysis, which compared gene expression between groups in all cells of a given cell type (e.g., cardiomyocytes, **Fig. 3b**), mostly suggested developmental anomalies, cell death and cancer across all cell types. To some extent, this may be due to the differences in gene expression between the subtypes, many of which showed opposing effects between cell subpopulations for the specific diseases and biological functions (e.g., arrhythmia, **Fig. 3c,d**), which would be expected to cancel each other out if all cells were analyzed together.

The most interesting results were obtained when disease-enriched clusters were compared to control clusters (**Fig. 3e,f**). In ischemic heart disease, disease-enriched clusters in cardiomyocytes were enriched for increased engulfment of cells, contractility, glucose tolerance and metabolism, uptake and oxidation of lipids, and modification and metabolism of reactive oxygen species and hydrogen peroxide, compared to control cells, whereas other cells in these ischemic heart disease-enriched clusters were enriched for increased cardiac contractile function, cardiogenesis, angiogenesis and development of vasculature, all signals that were missing from the ischemic cardiomyocytes (**Fig. 3f**). Among the main upstream regulators in ischemic heart disease were ANGPT2, IL-15, PPARG, PPARGC1A, and PPARA each of which have been implicated in changes in substrate utilization not only in cardiomyocytes but also in other cell types (**Fig. 3g**). In ischemic heart failure,

even in the right atrium, there was a substantial enrichment of the pathways implicated in cardiac hypertrophy, inflammation, and vascular dysfunction, including pathways such as IL-15, IL-1, and renin-angiotensin signaling, which increase oxidative stress, in addition to systemic cytokines, chemokines and adhesion molecules, which promote glycolytic metabolism (**Fig. 3h**). In non-ischemic heart failure, cardiomyocytes were enriched for processes which underlie primary forms of cardiomyocyte dysfunction including cell engulfment, transformation, formation of stress fibers, glycolysis and premature senescence. In both ischemic and non-ischemic heart failure, a general enrichment of pathways which drive cardiomyocytes toward a more mesenchymal fate, such as rearrangement and formation of muscle, increased contractility, and more 'migratory' behaviors (e.g., neuritogenesis or angiogenesis), was further associated not only with metabolic switching but also with changes in ion channels and junctional proteins, leading to both partial dedifferentiation of the cells as well as vulnerability to contractile and arrhythmogenic phenotypes when combined with any additional stressors (**Fig. 3d,e, f**). The main upstream regulators included VEGFA, EDN1, SREBF1, MYC and NFE2L2, each implicated in components of myocardial growth phenotypes, but not previously integrated in human heart failure (**Fig. 3i**).

In fibroblasts (**Fig S15**), ischemic heart disease caused changes that promoted degeneration of cells, lipolysis and formation of white adipocytes, fibrosis (fibrogenesis, shape change of fibroblasts, interstitial fibrosis connective tissue disorder), cardiac remodeling (formation of filaments and actin stress fibers, remodeling of cytoskeleton, dysmorphogenesis), Inflammation (quantity of interleukin, quantity of leukocytes and lymphocytes, immune response of tumor cells, Th1 immune response, influx of neutrophils), and cardiovascular disease. Activated regulators of these functions include miR-124-3p that regulates fibroblast conversion to neurons in human, GFI1 that converts adult fibroblasts to hematopoietic stem cells, RAF1 that promotes fibroblast migration and proliferation, and ADIPOQ that increases lipid accumulation, whereas SMARCA4 - a part of the chromatin remodelling complex SWI/SNF that is essential for heart fibrosis, NKX2-3 that inhibits cell proliferation, TP53, IFNG, and Insulin signaling were among the inhibited regulators. In ischemic heart failure (**Fig S15**), fibroblasts promoted cell death (apoptosis of fibroblasts and motor neurons), proliferation of fibroblasts, stem cells and epithelial cells, fibrogenesis, transformation and myofibroblast formation, metabolic changes, inflammation, atherogenesis (vascular permeability, uptake of fat, synthesis of phospholipid, lesion formation, migration of smooth muscle cells, aggregation of cells, thrombosis, infarction), migration and differentiation of epithelial cells, fibroblasts, smooth muscle and muscle cells, and cancer-like growth (neoplasia of fibroblasts, neoplastic cell transformation of fibroblasts), thereby following the route laid in ischemic heart disease but taking them step further. Among activated regulators, we found HGF, TP63, IL5, TFE3, TGFB1, and calmodulin, whereas PRKAA1, STEAP3, MAP4K4, and NUPR1 were among the inhibited regulators. Compared to ischemic heart failure, non-ischemic heart failure appeared more hypertrophic in nature with a fine balance between regenerative and disease signaling (**Fig S15**). More specifically, fibroblasts were enriched for signals enhancing cardiac remodelling, senescence, cell death, cardiovascular disease and fibrosis, but also signaling enhancing the function of the 'cardiac brain' (growth of neurites, development, proliferation, outgrowth and morphogenesis of neurons, axonogenesis, cell movement of Schwann cells),

increasing angiogenesis, clearance of dead cells, and proliferation of cardiomyocytes. The activated regulators included TGFB1, FGF2, SP1, AGT, CCR2, TP53, and IGF1, whereas miR-29-3p, let-7, miR-155-5p, and CTBP1 were inhibited.

In epicardial mesothelial cells (**Fig S16**), ischemic heart disease signaling suggested increased cell death and degeneration (muscle atrophy, neurodegeneration, neuromuscular disease, and apoptosis of neurons, stem cells, epithelial cells, cardiomyocytes, and heart cells in general), occurence of electrophysiological disturbances (arrhythmia, seizures), loss and dysfunction of myocardium, fibrosis, senescence, metabolic disturbances (disorder of lipid metabolism, accumulation of lipid, triacylglycerol metabolism), risk of myocardial infarction, deteriorated cardiac function and higher mortality (cardiac lesion, sphingolipid and ceramide synthesis, infarction). In ischemic heart failure, the profile changed from neuromuscular changes found in ischemic heart disease towards inflammation, cell movement, cardiac remodeling, oxidative stress, and vascular damage with some tendency to regeneration, all of which have been suggested to happen in epicardial cells, and are now shown in human tissue cells. In non-ischemic heart failure, mesothelial cells were enriched for increased differentiation of muscle cells, cell transformation to mesenchymal phenotype, cell movement and homing, contractility, adipogenesis, capillary vessel formation, proliferation and cell cycle progression, but also showed signs of inflammation, fibrosis, oxidative stress, atherogenesis, and occlusion of blood vessel. A closer look on the mesothelial subpopulation ME1 (**Fig S16**) showed pathways that were shared between the heart failure groups (e.g., Type II diabetes, iNOS, and TGF-β signaling pathways), decreased in ischemic heart disease and heart failure and increased in non-ischemic heart failure (e.g., WNT signaling, ferroptosis, and death receptor signaling pathways), downregulated in ischemic heart disease and upregulated in both ischemic and non-ischemic heart failure (e.g., NRF2-mediated oxidative stress response, senescence, tumor microenvironment, IL-17, HIF1α, cardiac hypertrophy, EMT, and thrombin signaling pathways), and distinct pathways across the disease groups, such as antioxidant action of vitamin C, RhoGDI, PPAR, PTEN, and sirtuin signaling for ischemic heart disease, aldosterone signaling in epithelial cells, and IP_3 biosynthesis in ischemic heart failure, and MYC-mediated apoptosis signaling, intrinsic prothrombin activation pathway, hepatic fibrosis signaling pathway, IL-23, interferon, FGF, ceramide, and ErbB2-ErbB3 signaling in non-ischemic heart failure. List of upstream regulators (**Fig S16**) for ischemic heart disease was laden with microRNAs, such as miR-124-3p, miR-155-5p, miR-30c-5p, and let-7a-5p which have all been linked to phenotypic transitioning and development of cells in addition to miR-1, which is strongly linked to development and function of the myocardium as well as cardiovascular disease, especially hypertrophy, ischemia and arrhythmias. Among the downregulated regulators were TCF7L2, VEGF, FOXO1, MYC, and TGFB1. For ischemic heart failure, the list of main inducers included CEBPA, CSF1, TAZ, TGFB1, HNF4A, whereas for non-ischemic heart failure the list included PRL, CSF2, RAC1, TNF, ANGPT2, HOXA3, ROCK1, ERBB2, WNT11, NRG1, and BMP10.

Endocardial endothelial cells (**Fig S17**) responded to ischemic heart disease by increasing turnover, development and activation of fibroblasts and connective tissue, response, function and formation of muscle

cells, glucose and lipid metabolism, cell transformation, senescence, oxidative and ER stress, apoptosis of cardiomyocytes and endothelial cells, and endothelial activation. Across endocardial subpopulations, hierarchical clustering at the pathway level divided the populations into two groups with EEC1, EEC4, EEC5, and EEC10 in one group with an overall trend of upregulation, and the rest in the other with more down-regulatory profile. The first cluster displayed most similarity in upregulation of pathways, such as cardiac hypertrophy, HIF1α, regulation of endothelial-to-mesenchymal transition, fibrosis, and white adipose tissue browning, across the subgroups. For the second group, among the suppressed pathways were regulation of the endothelial-to-mesenchymal transition, eNOS, PDGF, HMGB1, protein kinase A signaling, and senescence pathway. Leukocyte extravasation, IL-8, and inositol phosphate signaling were upregulated across most endocardial subpopulations. The top activated upstream regulators for each subpopulation were NUPR1, IKZF1, and MYC for **EEC0**, PKD1, HOXA10 and FOXO1 for **EEC1**, tretinoin, IL4 and beta-estradiol for **EEC2**, MITF, ZBED6 and ROCK2 for **EEC3**, IFNG, EGF and HIF1A for **EEC4**, KITLG, EPAS1 and CD38 for **EEC5**, miR-34a-5p, PGR and HIF1A for **EEC6**, MITF, TNF and KMT2D for **EEC7**, KLF2, ERBB2 and E2F1 for **EEC8**, beta-estradiol, hydrogen peroxide and insulin for **EEC9**, ESR2, MAPK1 and beta-estradiol for **EEC10**, and TGFB1, ERBB2 and CSF1 for **EEC11**. In ischemic heart failure (**Fig S17**), endocardial cells signalled glucose metabolism disorder, increased immune cell activation, cell migration, contraction of cells and cardiac muscle, angiogenesis, vascular disease, cell death of heart cells, formation of reactive oxygen species, phospholipid metabolism, lipid oxidation, and blood clot formation. We observed a wealthier pathway enrichment and activity in subgroups EEC0-8, and considerably weaker in groups EEC9-11. In terms of cardiac hypertrophy, EEC0, EEC5 and EEC6 presented strongest downregulation, and EEC2, EEC8 and EEC11 showed strongest upregulation of the pathway signaling, whereas cardiac regeneration was significantly downregulated in EEC2 and EEC5-8 and upregulated in EEC1, EEC3-4 and EEC11. The top activated regulators in ischemic heart failure were IFNG, GATA1 and NUPR1 for **EEC0**, NR3C1, E2F1 and F2 for **EEC1**, CXCL12, ESR2 and XBP1 for **EEC2**, LDB1, LMO2 and GPS2 for **EEC3**, ERG, PRL and IFNARs for **EEC4**, miR-199a-5p, IL1RN and CDKN2A for **EEC5**, ERBB2, miR-1-3p and PRL for **EEC6**, PDGFs, MYC and (9E,11Z)-octadeca-9,11-dienoic acid that regulates e.g. PPARG, IL6, SLC2A4, CXCL8, FASN, SREBF1, ADIPOQ, ATF3 and IL1B for **EEC7**, TP53, PML and miR-16-5p for **EEC8**, SMARCA4, TWIST and STAT6 for **EEC9**, CX3CR1, miR-155-5p and PRKAA2 for **EEC10**, and dihydrotestosterone, ESR2 and STK11 for **EEC11**. In non-ischemic heart failure (**Fig S17**), endocardial cells presented enhanced formation of microvasculature, muscle contractility, production of hydrogen peroxide, morphogenesis of cardiovascular system, turnover of intercellular connections, adipogenesis, structural changes, tubulation and rearrangement of cells. Compared to ischemic heart failure, non-ischemic heart failure showed enhanced pathway activity across all endocardial subclusters with EEC9 and EEC10 showing exceptionally strong signaling activity. Pathways with increased activity across most subclusters included thrombin, IL-8, actin cytoskeleton, integrin, cardiac hypertrophy, hepatic fibrosis, CXCR4, synaptogenesis, and inositol phosphate signaling. EEC1 and EEC9 presented increased activity across many pathways with only weak or no activity in other subclusters. For example, FGF, Th1, IL-7, Type I diabetes, IL-23, and apelin cardiac fibroblast signaling were upregulated in EEC1 and melatonin, IL-1, and CCR3 signaling in eosinophils for EEC9. The main upstream regulators for each subcluster included LBD1, LMO2 and miR-210

for **EEC0**, glucose, BRCA1 and FLI1 for **EEC1**, PRL, IL2 and fatty acid for **EEC2**, FGF2, LDLR and hydrogen peroxide for **EEC3**, TNF, TGFB1 and hydrogen peroxide for **EEC4**, beta-estradiol, UCP1 and TGFB1 for **EEC5**, FOS, PRKAR1A and CNR1 for **EEC6**, RAF1, REL and IKBKB for **EEC7**, IKBKB, NOTCH and SMARCB1 for **EEC8**, GDF2, NR3C1 and CX3CR1 for **EEC9**, PGR, ERG2 and PRKD1 for **EEC10**, and MYCN, IL15 and IKBKB for **EEC11**.

Disease-associated genetic variations affect disease-relevant pathways

Ischemic heart disease is a consequence of coronary artery disease and coronary microvascular dysfunction. To identify enhancers that mediate effects of disease-associated genetic variants to genes in ischemic heart disease, we linked enhancers to their target genes and mapped coronary artery disease associated single nucleotide polymorphisms (SNPs) to the enhancers in each cell/subtype using two parallel approaches. The first strategy identified enhancers from 100 eGTEx left ventricle samples overlapping with strong GWAS variants, and established their links to target genes based on the shared genetic regulation between enhancer activity and gene expression. The second approach utilized the EpiMap analysis[11], which further partitioned multifactorial traits into single-tissue and multi-tissue loci to suggest context-specific mechanisms. We used these approaches to map GWAS-linked genes in each main cell type (**Fig. S18** and **Table S43-S44**) and in the vascular subtypes (**Fig. 4a**, **S19**, and **Table S45-S46**). Generally, these GWAS-linked genes were enriched in common pathways that mediate disease-relevant changes in cell function context-specifically (**Fig. S18-S19** and **Table S47-S48**). For example, in ischemic heart disease, the coronary endothelial GWAS-linked genes enriched for increased muscle function, formation and extension of cellular protrusions, microtubule dynamics, differentiation of epithelial cells, angiogenesis, immune response, and development of connective tissue cells, and vasculature, whereas in ischemic heart failure, they enriched for increased vascular occlusion, differentiation of epithelial tissue, cell death of connective tissue cells, lipid concentration, congenital heart disease, vascular lesion, and nitric oxide synthesis (**Fig. 4a**). In vascular subtypes, pericyte 2 and capillary 4 populations enriched for most GWAS-linked genes, in the main cell types, the top two were cardiomyocytes and pericardial mesothelial cells.

The top differentially expressed genes ($p < 0.05$) in coronary endothelial cells for ischemic heart disease and ischemic heart failure are shown in **Fig. 4b**. Interestingly, among the top upregulated genes are both *JCAD* (ischemic heart disease) and *SVIL* (ischemic heart failure), which arise from the same locus on chromosome 10 and link to the same regulatory element bearing several coronary artery disease GWAS SNPs (**Fig. 4b-c**). Of note, *JCAD* was downregulated in heart failure and *SVIL* was not differentially expressed in ischemic heart disease compared to control in coronary endothelial cells (**Fig. 4d**), which is the only endothelial subtype differentially expressing these genes. Overall, *JCAD* was most abundant in endothelial cells, whereas *SVIL* was expressed primarily in cardiomyocytes, fibroblasts and smooth muscle cells (**Fig. S20**). This is in line with the data from human aorta (**Fig. 4e**)[19], where *JCAD* was most abundant in endothelial cells and *SVIL* in smooth muscle cells.

The locus of JCAD/SVIL has been shown to be significantly associated with coronary artery disease, and several of the top associated variants are located in the regulatory elements active in the heart and coronary

artery (**Fig 4f and S21**). The top candidate suggested by the EpiMap data was rs9337951 (**Fig. 4c and S21**). To further identify individual causal variants, we tested each regulatory variant for its effect on reporter gene expression in human aortic endothelial cells and in primary human aortic smooth muscle cells subjected to pro-atherogenic stimuli (IL1beta and cholesterol-loading) (**Fig. S22** and **Table S49**). Altogether, six regions of 200 bp were studied representing haplotypes that encompassed nine common variants and seven rare variants in the European population. Of these, two common variants, namely rs9337951 and rs1342150 and four rare variants including rs531337994, rs148641196, rs113622617 and rs193042870 exhibited allele specific enhancer activity (**Fig. 4g** and **Table S49**). Motif analysis suggested that the rare variant, rs148641196, with highest allele specific activity restored a functional IRF binding site of an ETS binding site that could potentially promote inflammatory signaling during disease development (**Fig. 4h**). Of note, the expression of ETS2 is downregulated in coronary endothelial cells both in ischemic heart disease and ischemic heart failure, whereas IRF7 is upregulated in both and IRF1 is upregulated in heart failure (**Fig. 4i**). These transcription factor expression patterns could potentially further promote the changes observed in the regulatory element activity between the GWAS variants and reference sequences.

Taken together, we have shown how the expression maps of the live human heart obtained in this study can be utilized to dissect the cell types and cell subtypes where coronary artery disease-linked genes function, decipher pathways and biological functions affected by these genes, and filter putative causal variants in a GWAS-highlighted locus to find candidates for further functional testing.

Discussion

In this study, we present the first expression map from the live human heart. In addition, we dissect the functional changes in the diverse cell types and cell subtypes in ischemic heart disease and heart failure. We map disease-associated genetic variations to regulatory regions and link them to genes to map the disease-linked genes to the cell types in which they manifest, and further explore their combined effects in the disease context, thereby demonstrating the importance of high-resolution subtype mapping and genetic partitioning in identifying key mechanisms in human cardiovascular disease. We focus on the right atrium as the orthogonal information source, where direct consequences of local tissue damage are less prominent and the observed changes are linked to maintained disease signaling that changes the tissue function, eventually leading to organ failure. Taken together, the study confirms that, although the right atrium is not the primary site of injury causing ischemic heart disease, it is a descriptor of the rich source of information integrated at the organ level that affects heart function, and of the cell types orchestrating the adaptive/maladaptive responses to sustain homeostasis and/or drive the progression to organ failure.

In order to dissect human cardiovascular disease and improve the disease diagnostics and treatment, a comprehensive map of all disease-relevant tissues is imperative. A direly needed cell atlas for healthy human heart was recently published from 14 human hearts and 6 cardiac regions, collected after brain or cardiovascular death [13]. We compared the published data from the right atrium to our data from the living tissue to determine the possible caveats to consider between dead and live tissue. Overall, live tissue yielded greater numbers of cells, better separation of distinct cell types and more defined cell subpopulations.

Epicardial mesothelial, endocardial endothelial cells and vascular endothelial cells, which form the outer layer of the heart, inner layer of the heart and innermost layer of the blood vessels, respectively, were the three main cell types that were most under-represented in the samples collected after death, although samples collected from the *postmortem* tissue would be expected to contain these cell types in equal numbers compared to live tissue. Moreover, we were able to identify neurons from the live tissue, which were absent from the *postmortem* data. The signal from the samples collected after brain death was closer to living tissue compared to samples collected after cardiovascular death, which may result from shorter processing delay after circulatory arrest. Results from the pericyte population suggest that even when the cell population is more enriched in the *postmortem* tissue compared to live tissue, the signal collected from the cells is weaker also for the marker genes, which may make the cell identification more challenging. However, it should be noted that the *postmortem* samples were processed using both V2 and V3 chemistry of the 10X Genomics, which may have decreased the signal strength overall and, therefore, future studies from the *postmortem* tissue may perform better, although the analysis gave mixed results and could not conclude that based on the present samples. Taken together, our study confirms the cell types identified in the previously published atlas of the healthy human heart, adding more cell types and better precision due to improved data quality, thereby allowing for further subtype characterization, which is crucial for deciphering disease mechanics.

Upon injury, such as myocardial infarction, the cells closest to the injury site respond quickly and impact via paracrine signaling the function of the whole heart. These acute changes lead over time to more chronic changes and malfunction as the injured heart fails to regenerate and heal properly. Unlike the human heart, adult zebrafish possess the capacity to regenerate damaged tissue through efficient cardiomyocyte proliferation, endocardial activation, reactivation of epicardial fetal gene expression, and efficient resolution of fibrosis[20]. Successful functional regeneration of an injured human heart would be expected to follow a similar pattern, involving clearance of dead tissue, restoration of lost muscle, revascularization, electrical coupling of new cardiomyocytes, and resolution of inflammation and collagen/fibrin [21]. Although cardiomyocytes in the disease-enriched clusters in this study signaled increased clearance and survival of cells, as well as transition to dedifferentiated and, thus, potentially more proliferative state across all disease groups in the atrial tissue, they were also enriched for pathogenic changes and were distinctly absent of signals for cardiogenesis, angiogenesis and vascular development and enriched for fibrosis. Cardiac fibroblasts are generally in a low proliferative state, but are activated upon injury through cytokine stimulation, thereby promoting angiogenesis and fibrosis, which are important for healing cardiac tissue injury on the site of the injury in the acute phase, but can later on contribute to impaired cardiac function and, over time, arrhythmias[22]. Here, while fibroblast gene expression in non-ischemic heart failure indicated support for cardiac regeneration and function, it also suggested sustained activity that would contribute to cardiovascular disease and excess fibrosis. In ischemic heart disease and heart failure inflammation and pro-atherogenic changes were especially pronounced. Moreover, in both cardiomyocyte and fibroblast populations, there were clear differences between cell subpopulations within each disease group, indicating that although the ‘sicker’ cells tended to favor malfunction, the ‘less sick’ cells were still trying to balance the organ function through processes supporting

regeneration and normal organ function.

In the developing heart, epicardium is the source of nonmyocyte cardiac lineages, such as vascular smooth muscle cells, pericytes, and fibroblasts, as well as paracrine cues that are essential for fetal cardiac growth, and coronary vessel development. Interestingly, cardiac injury has been shown to reactivate the developmental gene programs that stimulate epithelial-to-mesenchymal transition, but the role epicardium plays in the adult human cardiovascular disease progression or cardiac tissue repair remains unclear[20]. Here, epicardial mesothelial cells showed strong signals indicating differentiation, transition and transformation of the cells, but also disease-driving changes, such as degeneration, electrophysiological disturbances, loss and dysfunction of myocardium, fibrosis, inflammation and vascular damage, placing epicardium in the center of disease-associated signaling. After heart injury, such as myocardial infarction or sustained ischemia, epicardium has been previously reported to activate in an organ-wide manner, expanding the message of the local damage to unaffected sites. This activation has been further shown to amplify the subsequent inflammatory response and contribute to extracellular components and growth factors that affect heart tissue structure and its electrophysiological properties[20].

During cardiac development and injury, endocardium has the potential to differentiate into multiple lineages. It is the source of coronary vessels and a resource for angiogenesis and gives rise to cushion mesenchymal cells that can expand, migrate and contribute to pericyte, smooth muscle cell and fibroblast populations. In addition, endocardial cells can contribute to intramyocardial adipocytes, hematopoietic cells, and a subset of cardiomyocytes[23]. Currently, the functional study of the adult endocardium, especially under various disease conditions, is hindered by lack of genetic tools that target adult endocardial cells specifically. This study provides several new, specific markers for adult human endocardial endothelial cells that can now be used to dissect endocardial function and its molecular mechanisms in cardiovascular disease more specifically. These results will help to investigate if the endocardium could be harnessed and reprogrammed to regenerate cells or tissue for the treatment of cardiovascular disease as recent studies have suggested[22].

Vascular endothelial cells are among the most abundant cell types in the heart and play an active role in cardiac physiology and pathology. For hypertrophic or regenerating heart, coronary angiogenesis is critical for maintaining cardiac vascularization and perfusion. Furthermore, both clinical and experimental data suggests that reduction of capillary density contributes to the development of contractile dysfunction and heart failure[24]. Although some heart cell subtypes in this study display signatures for the generation of angiogenic signals and for the contribution to vessel formation and neovascularization of the ischemic or failing heart, the overall signals were mixed with both pro- and anti-angiogenic signals being presented by the various cell types and cell subtypes. The vasculature itself showed anti-angiogenic signals in ischemic heart disease, which turned into more angiogenic when the disease advanced to ischemic heart failure. Overall, vascular dysfunction was evident in all disease groups. Interestingly, one of the central players in the disease progression were pericytes that presented strong disease-associated signatures in all three disease groups suggesting that this cell-type may constitute an under-appreciated driver of heart disease. Previous studies have documented pericytes as critical for the maintenance of the blood brain barrier and homeostasis in the central nervous system[25,26], as

well as for the production of paracrine factors targeting capillary endothelial cells to promote an anti-thrombotic endothelial phenotype[27].

For each GWAS locus, the lead SNP is in a linkage disequilibrium with multiple variants that span through the locus, any of them having the potential of being the causal variant or one of the variants responsible for the disease association. The variants further act upon one or more genes locally or at a distance to mediate the effect on disease risk. The set of genes may vary between cell types, but also between cell states (e.g., inflammation vs. basal state) or disease states (e.g., early vs. late disease). Furthermore, the genes may be enriched in specific disease-associated pathways and functions potentially heightening the subtle effect that a given variant/gene in isolation would have on cell function. The inadequate dissection of cell-type-specific gene expression in the human heart is a major obstacle for understanding the molecular mechanisms of cardiovascular disease. Moreover, the missing maps have made it especially difficult to decipher the causal variants behind GWAS signals and their cell types of action. Here, we combined our recent efforts to link disease-associated genetic variations to putative target genes across cell types and tissues[11] and the expression maps created in this study to trace down the potential cell-type-specific regulatory networks affected by coronary artery disease-associated genetic variations. The enrichment of the linked genes across all cell types and vascular subtypes revealed several disease-associated key functions and pathways affected cell-type-specifically by the GWAS-linked genes, emphasizing not only the importance of cell-type and cell-subtype level investigation of the GWAS effects but also the necessity of creating expression maps of the human heart in various cardiovascular conditions.

Activating the repair mechanisms in the adult heart requires directing the tissue to promote neovascularization over scarring, and regeneration over pathological remodeling. However, the picture on the tissue level is not simple. Injury of one part of the heart resonates through the whole organ leading to mixed changes. It is important to acknowledge that not all cells of the same cell type signal or act the same way and, thus, activation of developmental programs that might be desirable from the regenerative point of view may lead to uncertain outcomes when activated during injury. Inflammation and oxidative stress change responses and lead to activation of signaling cascades that feed fibrosis over resolution, cardiomyocyte remodeling over proliferation, arrhythmia over sinus rhythm, and metabolic changes that further increase reactive oxygen species production, lipid accumulation, and formation of oxidized lipids and as a result, promote vascular dysfunction.

In conclusion, these data indicates that to better dissect the pathogenesis of human cardiovascular disease, we need to map the cell-type-specific markers of human tissue cells, use these markers to further define the cell locations and microenvironments in the tissue milieu and decipher the cell-subtype-specific changes in gene expression across various cardiovascular traits to enable cell-type-specific interrogation of the human disease. Ultimately, the integrated information will be used to better understand the complex cell-cell interactions and may enable the identification of critical drivers of cardiovascular disease and the development of novel therapeutic strategies for ischemic heart disease.

Methods

Sample selection

Human right atrial appendage tissue biopsies were harvested as a part of the ongoing prospective CAREBANK study (ClinicalTrials.gov Identifier: NCT03444259). Ethical Committee of the Hospital District of Southwest Finland approved the protocol and study complies with the Declaration of Helsinki as revised in 2002. CAREBANK has enrolled patients undergoing open-heart cardiac surgery (coronary bypass surgery, operations for valvular heart disease and ascending aorta) since February 2016 at Turku University Hospital. All patients gave their written informed consent prior to surgery. Samples were snap-frozen immediately upon collection and stored in -70C. Clinical data collection was performed prospectively in an electronic case report form by a trained research nurse with structured questionnaires. Patients with ischemic heart disease had an indication for coronary revascularization as appropriate by the current guidelines. Ischemic heart failure was defined as reduced left ventricular ejection fraction below 45% in echocardiography in addition to above criteria for revascularization. Non-ischemic heart failure was defined as reduced left ventricular ejection fraction below 45% without coronary artery disease in preoperative coronary angiogram. Control patients had no coronary artery disease in preoperative coronary angiogram and they underwent cardiac surgery for valvular heart disease.

Histochemistry

Tissues were embedded in Optimal cutting temperature (OCT) compound and stored in -80ºC. OCT-embedded tissue blocks were sectioned at a thickness of 10 microns. Frozen tissue sections were air-dried at room temperature and then fixed in 4% paraformaldehyde in PBS solution for 10 minutes. Fixed tissue sections were stained with Hematoxylin (Vector Laboratories, Burlingame, CA) for 10 mins. After several washes in distilled water, sections were then stained with 0.5% Eosin (Sigma-Aldrich, St. Louis, MO) for 10 seconds followed by several washes in distilled water. Sections were then dehydrated by sequentially dipping 10 times in 50%, 70%, 95% and 100% ethanol respectively. Slides were completely air-dried and then cleared by dipping in xylene for 10 seconds. Stained tissue sections were mounted using EcoMount mounting medium (Biocare Medical, Pacheco, CA) and images were captured with a Nikon DX-DB digital camera using a Nikon Digital Sight DS-U1 microscope.

Isolation of nuclei from fresh-frozen and frozen OCT-embedded ex vivo cardiac tissue

For OCT samples, surrounding OCT was removed with a scalpel and the tissue was washed three times with ice-cold 0.04% BSA in PBS. For all samples, the tissue was transferred into a tube containing 1 ml of ice-cold lysis buffer (0.32 M sucrose, 5 mM $CaCl_2$, 3mM MgAc, 2.0 mM EDTA, 0.5mM EGTA, 10 mM Tris-HCl (pH 8.0), 1 mM DDT, EDTA-free protease inhibitor cocktail (Roche, Complete tablet), 1 U/µl RNase inhibitors (TaKaRa)), cut into 2-3 mm pieces with scissors, transferred to an ice-cold glass dounce tissue grinder and after 5 min of incubation on ice, it was stroked with a “Loose” and a “Tight” pestle,15 times each. The solution was passed through a 40-micron strainer and centrifuged at 1000 x g for 8 min at 4°C on a swing bucket rotor. The supernatant was removed and the nuclear pellet was resuspended in 1 ml of nuclei suspension buffer (0.04% BSA in PBS with RNase inhibitors) and centrifuged at 1000 x g for 8 min at 4°C on a swing bucket

rotor. After removal of the supernatant the nuclear pellet was resuspended in the remaining nuclei suspension buffer. The quality and concentration of the nuclear suspension was estimated under the microscope following staining with Trypan blue. Nuclei concentration was adjusted to 1000 nuclei / µl.

Droplet-based snRNA-seq

Immediately after nuclei isolation, we proceeded with the preparation of single-nuclei RNA (snRNA) libraries using the droplet-based RNA sequencing technology. Briefly, 5000-6000 nuclei were profiled per sample using the Chromium Single Cell 3' RNA reagent kit v3 according to the 10X Genomics protocol. The generated cDNA libraries were indexed, pooled, and sequenced using the NovaSeq 6000 S2 system and reagent kits (100 cycles) (Illumina, Inc).

Bioinformatics preprocessing

The samples were processed using cellranger v3.0 and the data analysed using Scanpy v1.7 on python v3.9. Post-mortem samples were taken from ref [13], and processed the same way. Cells that had less than 500 genes or more than 6000 were removed. Mitochondrial percentage was calculated for each cell (sum of counts from mitochondrial genes / sum of all counts for that cell). We calculated a mitochondrial filtering fraction (set at q = (1/sqrt(number of cells)) * 60), capped at 0.9 maximum. Cells with the higher mitochondrial content were filtered out so that (100 * q) percent of cells remained. Duplicates were removed using Scrublet[28]. Counts are normalized and logarithmized (ln(counts per 10000) + 1). The resulting normalized matrix was used for visualisation of gene expression.

Expression correction, sample heterogeneity correction, embedding and clustering

Genes that were expressed in less than 10 cells (across all samples) were filtered out. Variability filtering of genes was also applied (minimum mean of 0.005, maximum mean of 5 and minimum normalized dispersion of -0.25). Linear regression was used to correct for the number of counts per cell and mitochondrial percentage. Subsequent steps were standard scaling, Principal Component Analysis (PCA) and Harmony batch correction [29] at the sample level. Afterwards, a 10 nearest neighbor graph was constructed, based on the top 40 PCs, which was used for UMAP embedding. Clusters were found using the Leiden algorithm [30] at resolution 1.

Cell-type identification

Clusters were annotated through cell-type specific marker expression and correlation. Annotations were additionally confirmed through gene set enrichment using Enrichr [31] (from gseapy wrapper), from the first 500 significantly differentially expressed genes (wilcoxon ranked-sum, Benjamini Hochberg FDR).

Subtyping

Subclusters within cell-types were found through the reprocessing of a subset of cells, by regression, scaling and PCA steps. Batch correction was also performed again at the sample level, in addition with k nearest neighbours graph and UMAP plot. Clusters were discovered with Leiden clustering (resolution 1). In mesothelial cells, endocardial cells, cardiomyocytes and fibroblasts, small clusters were found to be enriched for markers specific for other cell types and were thus removed as doublets.

Differential expression analysis

Apart from basic cluster-specific marker identification, differential gene expression analysis was performed using the Nebula R package. Mitochondrial percentage and number of counts were used as additive confounding covariate in the model. LN method was used when the number of cells exceeded 1000, and the HL method when not. Benjamini-Hochberg method was used for FDR correction. Phenotype density figures were constructed using the '*density'* Scanpy function.

Ingenuity Pathway Analysis

The affected 'Diseases and biological functions', 'Canonical pathways' and 'Upstream regulators' were studied using the 'Core analysis' and 'Comparison analysis' of the Ingenuity Pathway Analysis (IPA, Qiagen Redwood City, www.qiagen.com/ingenuity) based on the differentially expressed genes for each cell type / cell subtype.

GWAS linking

GWAS-linked enhancers were identified from eGTEx tissue by overlapping with the variants with strong GWAS signal (p-value < $1e^{-6}$) of heart traits, including heart rate, hypertension and coronary artery disease. GWAS-linked gene were further inferred by quantifying shared genetic regulation between GWAS-linked enhancer activity and target gene expression from eGTEx and GTEx data by colocalization analysis (R coloc package), Mendelian randomization (R MendelianRandomization package) and partial correlation. From Epimap[11] we further integrated a total of 774 genes linked with coronary artery disease and filtered all genes by mean expression (min 0.2 in one or more cell (sub)type).

Allelic activity reporter assay

STARR-Seq massively parallel reporter assay[32] was used to compare the transcriptional activity of SNP alleles. The reporter library has been described in [33]. Primary human aortic smooth muscle cells (Thermo Scientific) and teloHAEC human immortalized aortic endothelial cells (ATCC) were cultured according to the manufacturers' recommendations. The library was transfected into smooth muscle cells by Neon electroporation (Thermo Scientific) as described in [34]. For teloHAEC, $3x10^7$ cells per replicate were transfected by Neon electroporation using multiple reactions of the 100 µl tip containing 15 µg of library DNA and $5x10^6$ cells per reaction. Three 1350V pulses of 10 ms width were used. Twenty-four hours after electroporation, cells were treated with either 10 ng/ml IL-1β (Sino Biological) or mock-treated. After 24 h treatment, cells were processed as described in [34]. For both cell types, library preparation and data processing was carried out as described in [34].

Prediction of TF binding motif disruption

To predict disruption of TF binding due to a SNP, the candidate SNPs were analyzed using all 94 high-confidence binding models that were published by [35] and derived by training deltaSVM models on *in vitro* protein–DNA binding data. The original authors' recommended thresholds were used for determining sequence binding and allelic disruption. In addition, motif disruption was predicted based on [36].

Data availability

GEO accession for scRNA-seq data of human and mouse aorta is GSE131780. snRNA-seq data for human right atrial appendage will be made available upon the publication of the paper.

Code availability

Code will be made available upon the publication of the paper.

Acknowledgements

This work was supported by the Academy of Finland [to SLK, grant no: 342074, to MUK, grant no: 333021], Clinical Research Fund (EVO) of Turku University Hospital, Turku, Finland [to TK], European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [to MUK, grant no: 802825], Finnish Foundation for Cardiovascular Research [to SLK, MUK and TK], Finnish Medical Foundation [to TK], Sigrid Jusélius Foundation [to SLK and MUK], Orion Research Foundation [to SLK], and Yrjö Jahnsson Foundation [to SLK].

Figures

Figure 1

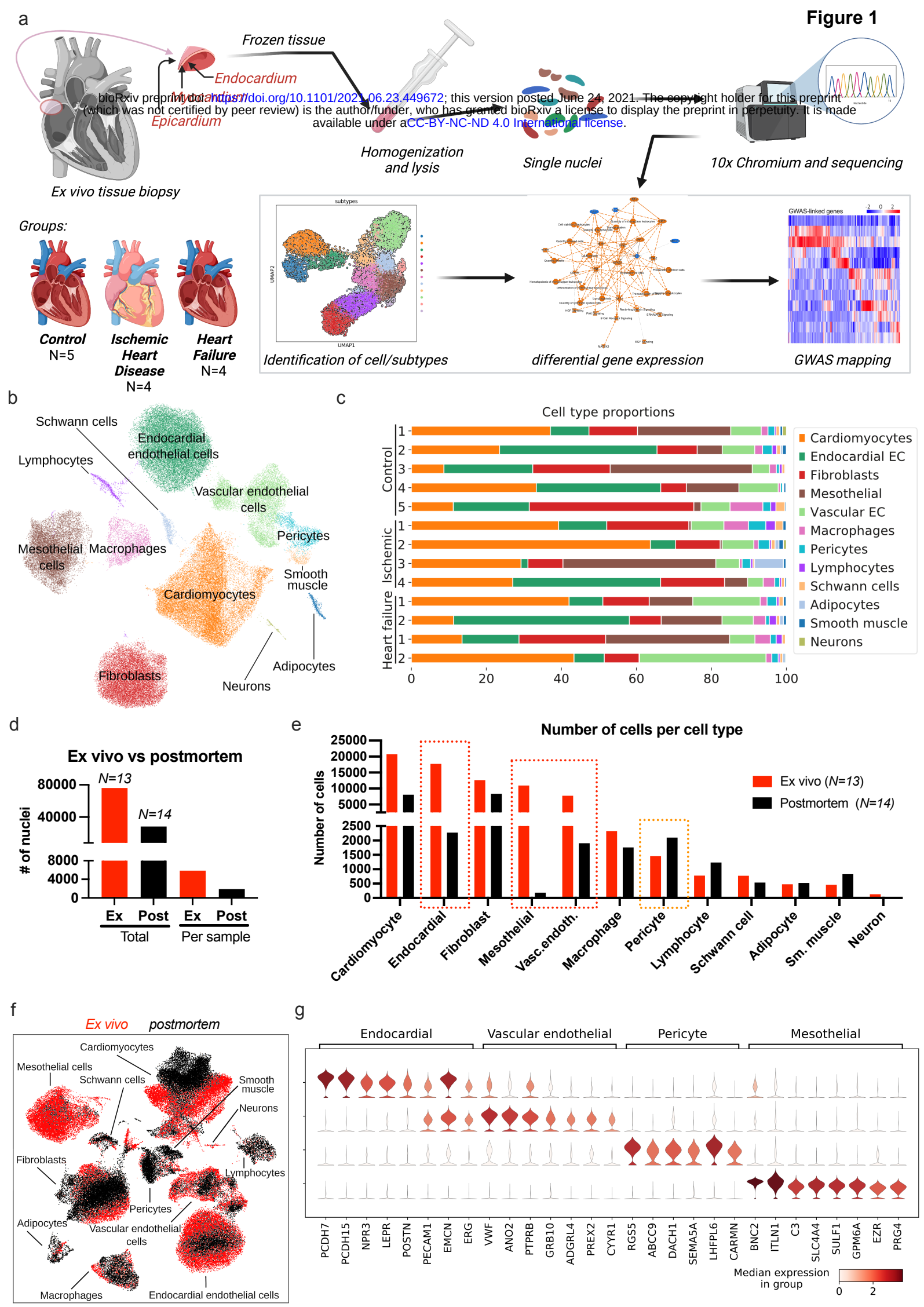


Figure 1. **Identification of marker genes and comparison of *ex vivo* and *postmortem* tissue. a.** *Ex vivo* samples were obtained from the right atrial appendage from 13 individuals. The samples were assigned to 3 groups: control (*N* = 5), ischemic heart disease (*N* = 4), and heart failure, which was further separated into Ischemic heart failure (*N* = 2) and non-ischemic heart failure (*N* = 2). Single-nuclei RNA-seq was processed using the 10X v3 Chromium chemistry. **b.** Uniform manifold approximation and projection (UMAP) embedding of the nuclei distinguishes populations that were identified as 12 different main cell types based on marker gene expression and GO term enrichments. **c.** Cell-type distribution for each sample is presented using the same colour code as in (B). **d.** Comparison of collected nuclei in *ex vivo* and *postmortem*[13] right atrial tissue by total number of samples and per sample. **e.** Number of cells per cell type presented for the *ex vivo* and *postmortem* samples. **f.** UMAP illustration of the differences between the two datasets among identified cell type populations. *Postmortem* cells are coloured in orange and *ex vivo* cells are shown in gray. **g.** Expression of cell-type-specific marker genes for endocardial endothelial cells, vascular endothelial cells, pericytes, and epicardial mesothelial cells are shown in addition to endothelial marker genes shared between endocardial and vascular endothelial cells.

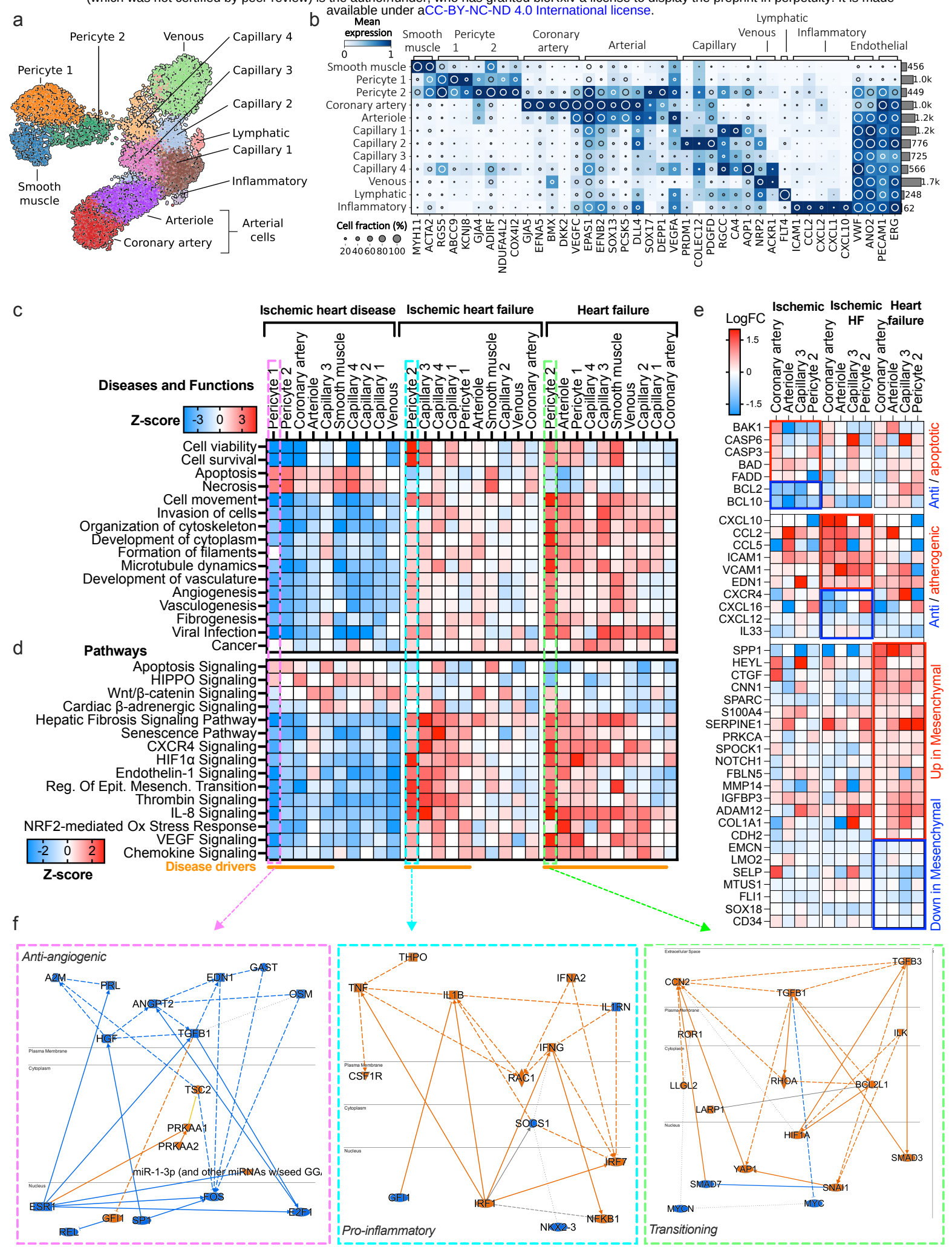


Figure 2. High-definition vascular subtyping and discovery of the subtype-specific disease signals. a. UMAP embedding of the 12 identified vascular subtypes. **b.** Specific marker gene expression in the vascular subtypes. Dot size represents the cell fraction expressing the gene, while colour represents min-max scaled mean expression within the subtype. **c-d.** Comparison Analysis of the Ingenuity Pathway Analysis (IPA) for Disease and Biological Function (C) and Pathway (D) enrichments based on differential expression analysis for the three phenotypes against the control samples. Analysis was performed for each vascular subtype independently. Disease driver subtypes are marked in orange. **e.** Heatmap of genes related to apoptosis, atherogenesis and endothelial-to-mesenchymal transition (EndMT) for coronary artery, arteriole, capillary 3, and pericyte 2 populations. Signal represents log fold change in the three phenotypes compared to control samples. **f.** IPA's Graphical Summary for the pericyte 2 population provides an overview of the major upstream regulators and diseases and biological functions in the IPA Core Analysis and illustrates their relation to one another.

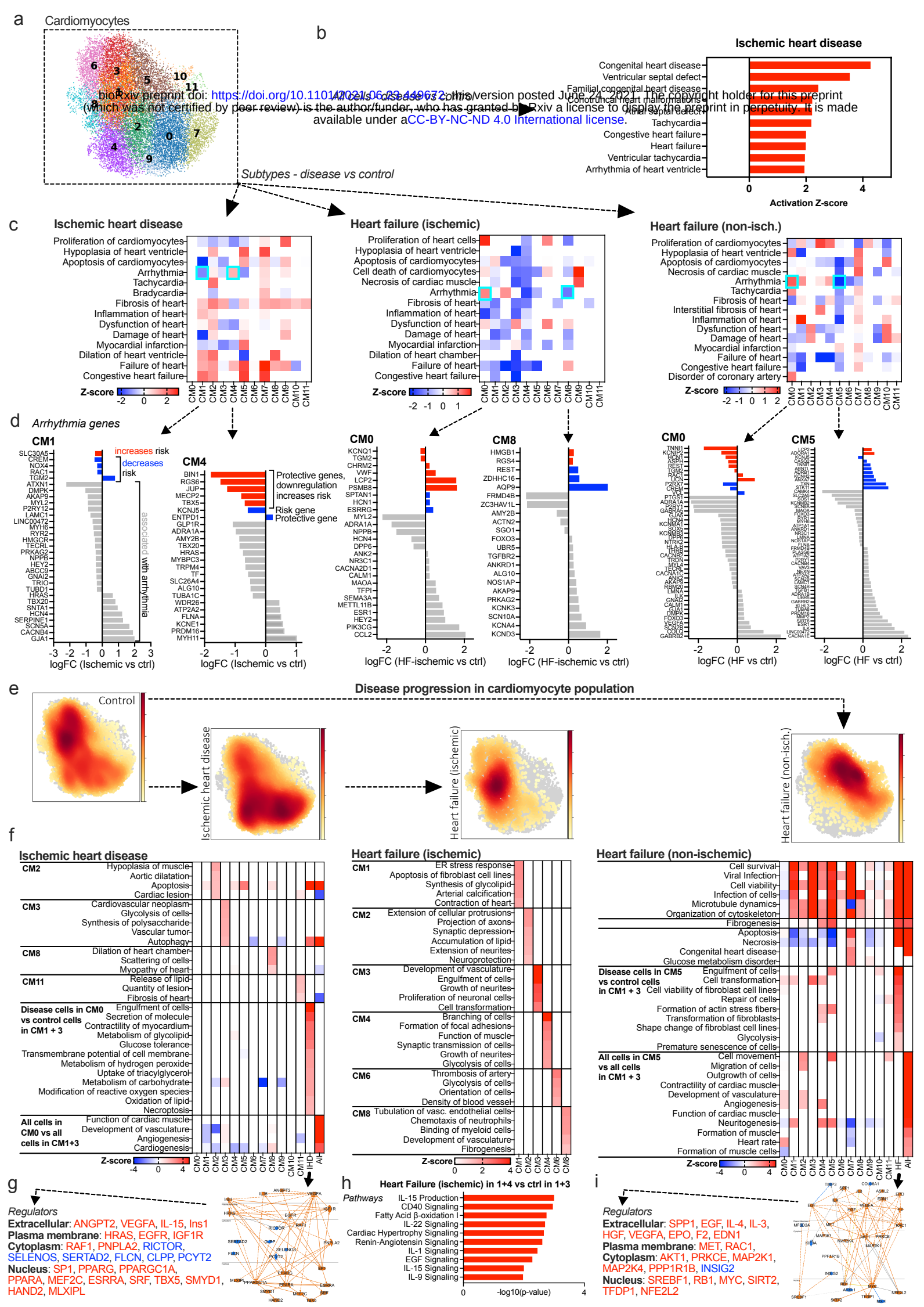
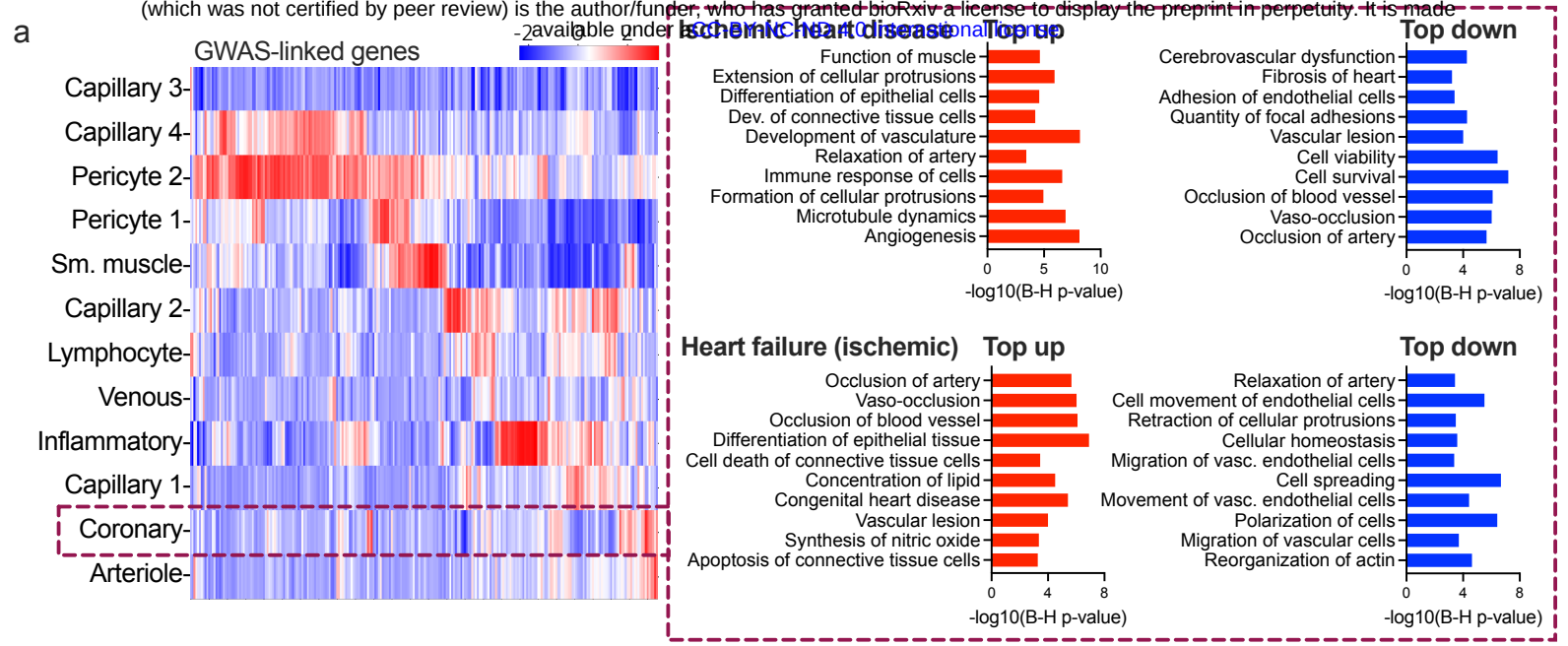
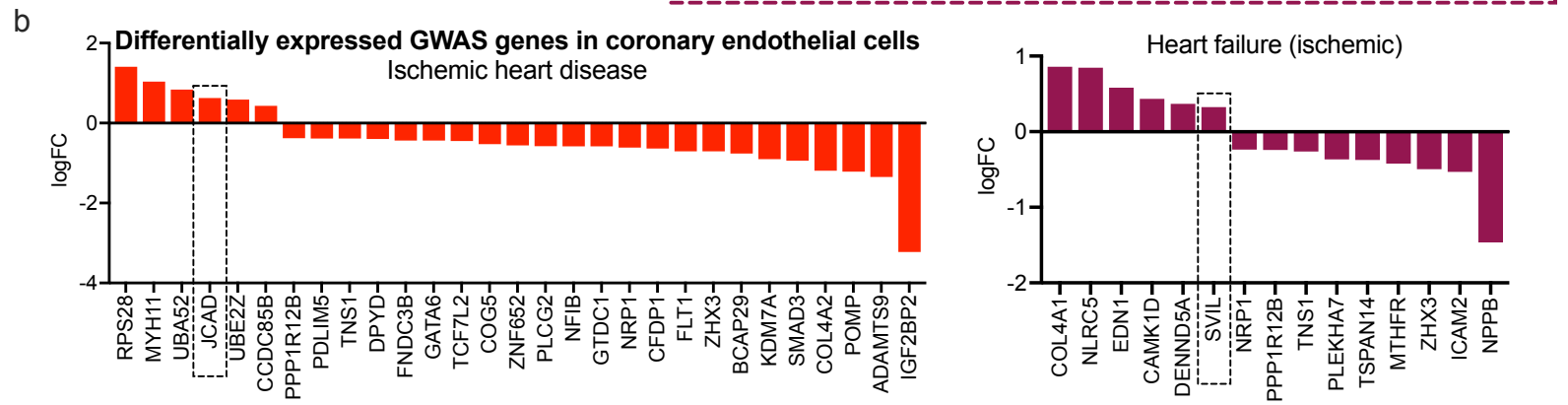
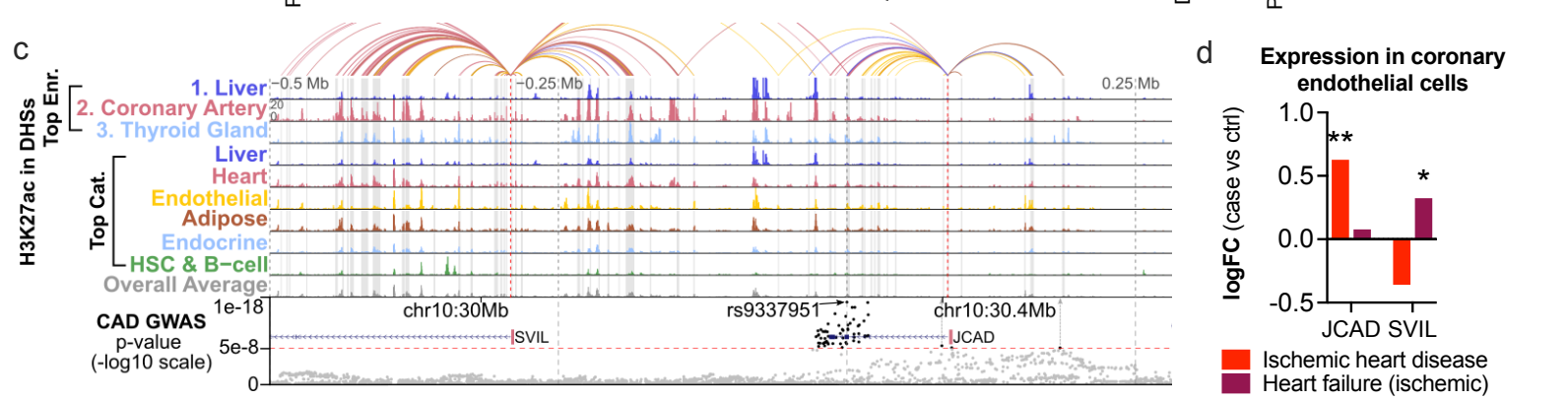


Figure 3. **Cardiomyocyte subpopulation-specific disease signatures. a.** UMAP embedding of the cardiomyocyte subclusters where the coloring is based on Leiden clusters. **b.** IPA enrichment of cardiotoxicity terms based on differential expression analysis of all cardiomyocytes (Ischemic heart disease against control). **c.** Cardiomyocyte subcluster-specific IPA enrichment of cardiotoxicity terms based on differential expression analysis for the three different disease phenotypes. **d.** Examples of the genes behind the arrhythmia signals in select subclusters in (C). Bars are colored based on whether the detected change in gene expression increases the risk of arrhythmia (red), decreases the risk (blue) or if the effect of the arrhythmia-related gene is not clear (orange) based on previous publications. **e.** UMAP embedding of cardiomyocyte subclusters highlighting the density of the cells for each phenotype, with yellow being the lowest and red the highest density. Background grey dots correspond to locations on the embedding that do not contain cells from that phenotype but of the other phenotypes. Arrows indicate the direction of the disease progression between the phenotypes. **f.** Heatmaps illustrate the IPA's comparison analysis for the enrichment of biological functions based on differential gene expression analysis in select subtypes, as well as in disease cells in disease-enriched subclusters against control cells in control enriched clusters, and all cells in these clusters against each other. **g.** IPA's Graphical Summary for the ischemic heart disease-enriched subclusters compared to control cells in control-enriched clusters. The upstream regulators suggested by the analysis are divided based on their cellular locations. **h.** Pathways enriched in the ischemic heart failure-enriched cell population compared to control-enriched clusters. The red color of the bars indicates increase in the pathway signaling. **i.** IPA's Graphical Summary for the non-ischemic heart failure-enriched subclusters compared to control cells in control-enriched clusters. The upstream regulators suggested by the analysis are divided based on their cellular locations.







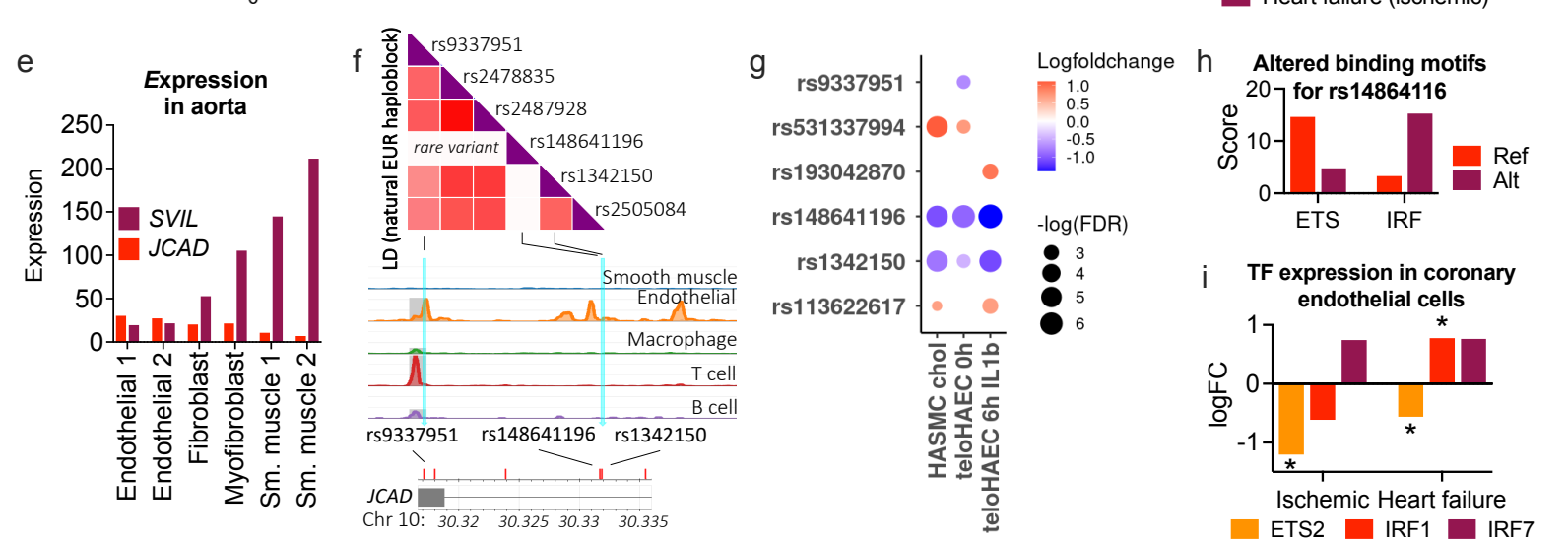


Figure 4. GWAS gene mapping and pathway enrichment. a. Heatmap showing the expression of genes linked with coronary artery disease GWAS loci in vascular cells. IPA's 'Diseases and Biological Functions' enrichments are shown as an example for coronary artery endothelial cells in ischemic heart disease and heart failure groups. 'Top up' shows top enrichments with positive Z-score (function increases based on differential gene expression) and 'Top down' top enrichments with negative Z-score (function decreases). All enrichments are statistically significant based on corrected p-values. **b.** Statistically significantly (p-value < 0.05) differentially expressed GWAS-linked genes in coronary artery endothelial cells for ischemic heart disease and ischemic heart failure. **c.** Epimap linking of the *JCAD/SVIL* locus. **d.** *JCAD* and *SVIL* expression in coronary endothelial cells in ischemic heart disease and ischemic heart failure patients compared to the control group. **e.** *JCAD* and *SVIL* expression in aortic cell (sub)types based on scRNA-seq [19]. **f.** Illustration of LD SNPs of the natural European haploblock combined with the scATAC-seq data from human aorta generated using LDlink[37]. **g.** Allele-specific enhancer activity measured with STARR-seq in teloHAECs and HASMCs in control conditions and after 6h IL1B (for teloHAECs) or cholesterol (for HASMCs) treatment. Color corresponds to the logFC in enhancer activity between the variant and the reference. Dot size represents statistical significance. Out of 16 SNPs tested, 6 induce significant changes in enhancer activity (FDR < 0.1). **h.** Transcription factor binding motifs altered by rs148641196. Position weight matrix scores for reference and alternate are shown. **i.** Changes in transcription factor (TF) *ETS2*, *IRF1*, and *IRF7* gene expression in coronary endothelial cells in ischemic heart disease and ischemic heart failure.

References

1. Musunuru, K. & Kathiresan, S. Genetics of Common, Complex Coronary Artery Disease. *Cell* **177**, 132–145 (2019).
2. Severino, P. *et al.* Ischemic Heart Disease and Heart Failure: Role of Coronary Ion Channels. *Int. J. Mol. Sci.* **21**, (2020).
3. Severino, P. *et al.* Ischemic Heart Disease Pathophysiology Paradigms Overview: From Plaque Activation to Microvascular Dysfunction. *Int. J. Mol. Sci.* **21**, (2020).
4. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
5. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
6. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
7. Ward, L. D. & Kellis, M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* **30**, 1095–1106 (2012).
8. Wang, X. *et al.* Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife* **5**, (2016).
9. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* **173**, 1573–1580 (2018).
10. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
11. Boix, C. A., James, B. T., Park, Y. P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300–307 (2021).
12. Pan, H. *et al.* Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. *Circulation* **142**, 2060–2075 (2020).
13. Litviňuková, M. *et al.* Cells of the adult human heart. *Nature* **588**, 466–472 (2020).
14. Suryawanshi, H. *et al.* Cell atlas of the foetal human heart and implications for autoimmune-mediated congenital heart block. *Cardiovasc. Res.* **116**, 1446–1457 (2020).
15. Tucker, N. R. *et al.* Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466–482 (2020).
16. Wang, L. *et al.* Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. *Nat. Cell Biol.* **22**, 108–119 (2020).
17. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764–779.e20 (2020).
18. Gimbrone, M. A., Jr & García-Cardeña, G. Endothelial Cell Dysfunction and the Pathobiology of Atherosclerosis. *Circ. Res.* **118**, 620–636 (2016).

19. Wirka, R. C. *et al.* Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat. Med.* **25**, 1280–1289 (2019).
20. Quijada, P., Trembley, M. A. & Small, E. M. The Role of the Epicardium During Heart Development and Repair. *Circ. Res.* **126**, 377–394 (2020).
21. Cao, J. & Poss, K. D. The epicardium as a hub for heart regeneration. *Nat. Rev. Cardiol.* **15**, 631–647 (2018).
22. Zhang, W., Liang, J. & Han, P. Cardiac cell type-specific responses to injury and contributions to heart regeneration. *Cell Regen* **10**, 4 (2021).
23. Zhang, H., Lui, K. O. & Zhou, B. Endocardial Cell Plasticity in Cardiac Development, Diseases and Regeneration. *Circ. Res.* **122**, 774–789 (2018).
24. Gogiraju, R., Bochenek, M. L. & Schäfer, K. Angiogenic Endothelial Cell Signaling in Cardiac Hypertrophy and Heart Failure. *Front Cardiovasc Med* **6**, 20 (2019).
25. Armulik, A. *et al.* Pericytes regulate the blood–brain barrier. *Nature* **468**, 557–561 (2010).
26. Török, O. *et al.* Pericytes regulate vascular immune homeostasis in the CNS. doi:10.1101/644120.
27. He, L. *et al.* Pericyte-specific vascular expression of SARS-CoV-2 receptor ACE2 – implications for microvascular inflammation and hypercoagulopathy in COVID-19. doi:10.1101/2020.05.11.088500.
28. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).
29. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
30. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
31. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).
32. Muerdter, F. *et al.* Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat. Methods* **15**, 141–149 (2018).
33. Selvarajan, I. *et al.* Integrative analysis of liver-specific non-coding regulatory SNPs associated with the risk of coronary artery disease. *Am. J. Hum. Genet.* **108**, 411–430 (2021).
34. Örd, T. *et al.* Single-Cell Epigenomics and Functional Fine-Mapping of Atherosclerosis GWAS Loci. *Circ. Res.* (2021) doi:10.1161/CIRCRESAHA.121.318971.
35. Yan, J. *et al.* Systematic analysis of binding of transcription factors to noncoding variants. *Nature* **591**, 147–151 (2021).
36. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **42**, 2976–2987 (2014).
37. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).