

# ScNaST - a computational workflow for spatial long Nanopore read transcriptomics

Etienne Boileau<sup>1,2,3</sup>, Xue Li<sup>4</sup>, Ramona Casper<sup>5</sup>, Janine Altmüller<sup>6</sup>, Florian Leuschner<sup>4,7</sup>, Christoph Dieterich<sup>1,2,3</sup>

<sup>1</sup>Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Im Neuenheimer Feld 669, 69120 Heidelberg, Germany

<sup>2</sup>Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, Im Neuenheimer Feld 669, 69120 Heidelberg, Germany

<sup>3</sup>DZHK (German Centre for Cardiovascular Research) Partner Site Heidelberg/Mannheim

<sup>4</sup>Department of Cardiology, University Hospital Heidelberg, Im Neuenheimer Feld 410, 69120 Heidelberg, Germany

<sup>5</sup>Cologne Center for Genomics (CCG), University of Cologne, Weyertal 115b, 50931 Cologne, Germany

<sup>6</sup>Berlin Institute of Health at Charité-Universitätsmedizin Berlin, Max Delbrück Center for Molecular Medicine

<sup>7</sup>Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Simches Research Building, 185 Cambridge Street, Boston, Massachusetts 02114, USA

Correspondence\*:

Christoph Dieterich

christoph.dieterich@uni-heidelberg.de

## 2 ABSTRACT

We introduce Single-cell Nanopore Spatial Transcriptomics (scNAST), a set of tools to facilitate the analysis of spatial gene expression from second- and third-generation sequencing, allowing to generate a full-length single-cell transcriptional landscape of the tissue microenvironment. Taking advantage of the Visium Spatial platform, we adapted a strategy recently developed to assign barcodes to long-read single-cell sequencing data for spatial capture technology. Here, we demonstrate our workflow using four short axis sections of the mouse heart following myocardial infarction. We show that results ... Molecular signatures involved in cardiac remodeling integrated with morphological context may support the development of new therapeutics towards the treatment of heart failure and the reduction of cardiac complications.

Keywords: Spatial transcriptomics, Single-cell RNA sequencing, Oxford Nanopore Technology, Myocardial Infarction

## INTRODUCTION

Cell type heterogeneity has recently emerged as a major aspect in redrawing the cellular picture of the mammalian heart (Wang et al., 2020; Tucker et al., 2020; Litviňuková et al., 2020). Single-cell RNA-seq

15 (scRNA-seq) technology has enabled to explore crosstalk of different cardiac cell populations to identify  
16 response signatures involved in remodeling after myocardial infarction (MI) and ischemic injury (Cui  
17 et al., 2020; Forte et al., 2020; Ruiz-Villalba et al., 2020; Vafadarnejad et al., 2020; Molenaar et al., 2021;  
18 Gladka et al., 2021; Tombor et al., 2021; Heinrichs et al., 2021). These and other data provide a valuable  
19 compendium of information to better understand transcriptional changes occurring in cardiomyocyte and  
20 non-cardiomyocyte sub-populations in healthy, injured, and regenerating hearts. However, in all these  
21 studies, the original tissue architecture is destroyed and, in general, the morphological context is lost,  
22 including the relationship of cells to infarct, border and remote zones (van Duijvenboden et al., 2019).

23 Recent development in spatial transcriptomics addresses this challenge, but few studies only have provided  
24 spatially resolved insights into the cardiac transcriptome. Techniques such as microdissection (??) or, in  
25 particular, spatially barcoded arrays and *in situ* capturing, have enabled to retain the spatial information  
26 while profiling the whole transcriptome at near single-cell resolution, allowing to shed light on localized  
27 tissue neighbourhoods (??), but a detailed characterization of cellular zones of injury, repair and/or  
28 remodeling is lacking. MI is a complex spatio-temporal heterogeneous disease involving the whole heart,  
29 and unbiased spatial transcriptomics holds a promise to add tissue context to molecular profiling in the  
30 search of novel therapeutics.

31 One of the most established spatial transcriptomics methods is now widely available as the Visium  
32 platform by 10x Genomics. After the tissue section is fixed on the spatial slide, stained, and imaged, it is  
33 permeabilized to release RNA to bind to capture probes for on-slide cDNA synthesis. Library preparation  
34 is performed off-slide, and spatial barcodes and tissue image are used to overlay transcriptomics data with  
35 tissue context. Although several new such methods have recently been published (??), they are for the  
36 most part relying on short-read library preparation. Thus they are subject to amplification biases and fail to  
37 generate sufficient overlaps to reconstruct transcriptomes *de novo*.

38 Adequate gene and transcript models are instrumental towards relevant proof of concepts and  
39 investigational new drug development in translational cardiac research (Müller et al., 2021). Third-  
40 generation or long-read sequencing allow to reconstruct truthful assemblies with fewer gaps and  
41 to characterize complete transcript isoforms and chimeric transcripts. This has recently been taken  
42 to the single-cell level using either Pacific Biosciences (PacBio) or Oxford Nanopore technology  
43 (Nanopore) (?????).

44 Here, taking advantage of the conceptual similarity between spatial and cell barcodes, we introduce  
45 Single-cell Nanopore Spatial Transcriptomics (SCNaST), a set of tools to facilitate whole transcriptome  
46 spatial profiling of full-length transcripts, based on a previously published Bayesian approach for cell  
47 barcode assignment (?). Our method relies on the commercially available Visium platform by 10x Genomics  
48 and Nanopore long-read sequencing, although it can in principle be extended to other technologies, as long  
49 as a hybrid sequencing approach is used for spatial barcode assignment.

50 We demonstrate ??

## MATERIAL AND METHODS

### 51 Experimental model of myocardial infarction

52 A C57BL/6 mouse (female, 8 weeks old, Janvier Labs) was exposed to 5% isoflurane for anesthesia. An  
53 intubation cannula was inserted orally into the trachea. The mouse was fixed on a heating plate at 37°C  
54 and maintained under anesthesia with 2% isoflurane. An incision was made from the left sternum to the

midclavicular line. Skins and muscle layers were stretched with forceps and sutures. Another incision was made between the third and fourth intercostal space. The heart was exposed and subjected to permanent myocardial infarction with ligation of the left anterior descending (LAD) coronary artery. When the ribs and skins were fully closed, the isoflurane supply was cut off. Oxygen was then supplied until normal breathing was resumed. 24 hours after surgery,  $100\mu\text{l}$  of blood was collected from the retro-orbital sinus. After incubation with heparin for one hour at room temperature, tubes containing blood were centrifuged for 15 min (12,000g at room temperature).  $10\mu\text{l}$  of the supernatant plasma was diluted with  $390\mu\text{l}$  PBS for Cardiac Troponin T (cTnT) analysis. The cTnT level is a reference for cardiac infarction size.

### 63 Heart extraction and cryosection

Three days after permanent LAD ligation, the mouse was sacrificed for organ harvest. After washing with cold PBS several times to remove the blood, the heart was transferred into a bath of isopentane (Millipore Sigma) frozen by liquid nitrogen. The freshly obtained heart was kept fully submerged in isopentane for 5 min until fully frozen. Pre-cooled Cryomold on the dry ice was filled with chilled TissueTek OCT compound without introducing bubbles. The frozen tissue was then transferred into the OCT with pre-cooled forceps and placed on the dry ice until the OCT was completely frozen. OCT-embedded tissue blocks were removed from the Cryomold and mounted on the specimen stage.  $10\mu\text{m}$  sections were cut in a cryostat at  $-10^\circ\text{C}$  and placed within a Capture Area on the pre-equilibrated Visium Spatial Slides (10x Genomics). The slides were later sealed in individual 50 ml Falcon at  $-80^\circ\text{C}$  ready for further processing.

### 73 10X Genomics Visium experiments

Four short axis sections of the heart were processed according to the manufacturer's protocol. The Visium Spatial Tissue Optimization Slide & Reagent kit (10x Genomics) was used to optimize permeabilization conditions. Tissue permeabilization was performed for 24 min. Spatially barcoded full-length cDNA was generated using the Visium Spatial Gene Expression Slide & Reagent kit (10x Genomics). A fraction of each cDNA library was used for nanopore sequencing. cDNA amplification was then conducted for 20 cycles of PCR (identified by qPCR), and  $400\mu\text{l}$  were used in the 10xGenomics Visium library preparation ( $100\mu\text{l}$  per section). The libraries were sequenced on a NovaSeq6000 (Illumina), with 29 bases from read 1 and 90 bases from read 2, at a depth of 160M reads per section (640M reads in total). The raw sequencing data was processed with the 10x Genomics Space Ranger 1.1.2 and mapped to the mm10 genome assembly (mm10-2020-A).

### 84 Oxford Nanopore sequencing libraries

Libraries for Nanopore sequencing were prepared according to the manufacturer's protocol for direct sequencing of native RNAs (SQK-DCS109 Oxford Nanopore Technologies) with the following minor modifications: 2 ml tubes were used for the HulaMixer, pellets were re-suspended with ABB (room temperature) and centrifugated 2 times, elution was performed in  $15\mu\text{l}$  to have material for TapeStation and Qubit BR, and Flow Cell Priming kit EXP-FLP002 (new version) was used. Starting with 200ng cDNA each, four GridIon flow cells (FLO-MIN106) were loaded with 12ml libraries, with a final cDNA concentration determined by Qubit BR. Base calling was done with Guppy v5.0.12. The High accuracy (HAC) model was selected for base calling (Q-Score cut-off  $>9$ ).

### 93 Spatial barcode assignment

To account for source-specific quality differences, each heart section (Illumina libraries) was processed separately using Scanpy v1.7.2 (?), keeping only spatial barcodes with approximately  $150 < \text{counts} <$

96 18000, 250 < genes < 5000, detected in at least 2 spots, and with less than 40% mitochondrial counts. The  
97 resulting datasets were concatenated, normalized, and the union of highly variable genes (per section) were  
98 kept for final analysis. Batch balanced KNN (BBKNN) (?) with ridge regression (Park et al., 2020) was used  
99 for integration and batch correction, starting from a coarse clustering obtained from a BBKNN-corrected  
100 graph.

101 For the Oxford Nanopore libraries, samples were demultiplexed and processed with ScNapBar v1.1.0  
102 using a Naive Bayes model to assign spatial barcodes (?). Briefly, for each heart section, the Space  
103 Ranger results (Illumina libraries) were used to parametrize a model of barcode alignment features to  
104 discriminate correct versus false barcode assignment in the Nanopore data. FASTQ files were mapped  
105 using minimap2 v2.21 (?). For transcript isoform quantification, a de novo transcriptome annotation was  
106 generated. Alignment files were processed by StringTie v2.1.5 in long read mode with the reference  
107 annotation to guide the assembly process (?). The annotations were merge into a non-redundant set of  
108 transcripts and compared to the reference using GffCompare v0.12.2 (?), after removing single-exon  
109 transcripts. To generate feature-spatial barcode matrices, alignment files were split into multiple files,  
110 one per spatial barcode, based on the barcode assignments, converted to FASTQ, and aligned to the de  
111 novo transcriptome with minimap2. Abundances were quantified with Salmon v1.5.2 in alignment-based  
112 mode using a long read error model (?). Each section was processed separately using Scanpy v1.7.2 and  
113 integrated with BBKNN, as described above for the Illumina data. Spatial barcodes were filtered for counts  
114 (approximately 50 < counts < 4000), transcripts (approximately 50 < transcripts < 2000, detected in at  
115 least 2 spots), and ribosomal genes ( $\geq$  40%).

## 116 Identification of anatomical regions

117 For the Illumina libraries, the neighborhood graph was computed using BBKNN. Spots were clustered  
118 with a low resolution (0.3) to identify anatomical regions such as infarct, border and remote zones. Marker  
119 genes were identified using a Wilcoxon rank sum test with Benjamini–Hochberg correction, by comparing  
120 the expression of each gene in a given cluster against the rest of the spots. The final clusters were manually  
121 annotated.

122 Labels were then assigned to Nanopore spatial barcodes based on the set of matching Illumina barcodes.  
123 However, not all assigned barcodes were labeled due to quality control filtering criteria that were different  
124 between Illumina and Nanopore datasets. To assign labels to the remaining Nanopore barcodes, seed  
125 labelling was performed with scANVI using the set of assigned labels as groundtruth (?). The top expressed  
126 transcripts were then identified as described above for the Illumina data.

## 127 Spatial spot deconvolution

128 Spatial spots were deconvoluted using stereoscope (scvi-tools v0.15.0) (?). Heart data (Smart-Seq2 and  
129 10x Genomics) from the Tabula Muris (Schaum et al., 2018) were used as reference dataset and highly  
130 variable genes were identified. For the Smart-Seq2 data, gene length normalization was applied. The model  
131 was trained on the single cell reference dataset on the intersection of genes found in the spatial (Illumina)  
132 data, and proportions were inferred for each Visium spot for each cell type in the reference dataset. Labels  
133 were then assigned to Nanopore spots are described above.

---

**134 Spatial gene expression (Illumina)****135 Differential transcript usage (Nanopore)**

136 To identify changes in relative usage of transcripts/isoforms within genes, differential transcript usage  
137 (DTU) tests were performed using satuRn (?). Only multi-exon transcripts and genes with more than one  
138 isoform were kept for the analyses. The transcript count matrix was further filtered to keep transcripts  
139 expressed in a worthwhile number of spots, determined by the design (but greater than at least 10% of  
140 the smallest group size), with a CPM count above a threshold of  $1(\text{median library size})^{-1}$ . In addition,  
141 transcripts were kept only if they had a minimum count of 1 across all spots. A quasi-binomial generalized  
142 linear model was fit using a design comparing each anatomical region with another, or each anatomical  
143 region with the rest of all regions. A two-stage testing procedure was performed using stageR (?), with an  
144 OFDR of 0.05. Results were reported using a student's t-test statistic, computed with estimated log-odds  
145 ratios.

## RESULTS

146 Fresh-frozen tissue samples were stained, imaged and fixed on Visium Spatial Gene Expression Slides  
147 (10X Genomics) for permeabilization and *in situ* RNA capture. Full-length cDNA libraries were split for  
148 the preparation of 3' Illumina short-read and direct long-read Nanopore sequencing libraries. Short-read  
149 data were used for the assignment of spatial barcodes to Nanopore reads using the SCNAPBAR workflow,  
150 and subsequently used to define anatomical regions within the tissue organization (Fig. 1A). Long-read  
151 data were used for transcriptome assembly and transcript abundance quantification, and layered onto the  
152 stained images to reveal the spatial organization of isoform expression. The Nanopore data comprises  
153 of four heart slices (or samples) with a total of 25,5 million reads, reaching a relatively high sequencing  
154 saturation (Fig. 1B), and providing **we expect here to have a more uniform coverage... see Figure 1 C, this**  
155 **is a little unexpected. I am checking this...** After spatial barcode assignment, libraries had a median of  
156 2.8 million reads per sample, with over 70% assigned reads (Figs. 1D and Supplementary Figure S1A).  
157 Despite variations in transcriptome alignments between samples, we observed a good correlation between  
158 spots across all samples between Illumina and Nanopore libraries (Fig. Supplementary Figure S1B,C, see  
159 also Figs. Supplementary Figure S2 to Supplementary Figure S4).

160 Clustering based on short-read gene expression defined four broad morphological regions: two remote  
161 zones that stem from differences in sequencing depth between heart slices, a border zone, and the infarct.  
162 The region classification was then transferred to the Nanopore data (Fig. 2A,B) Remote zones and, to a  
163 lesser extent the border zone, are largely associated with cardiomyocyte markers, while the border and  
164 infarct zones are characterized by a higher expression of endothelial, myofibroblast, and immune marker  
165 genes, as well as with markers of fibrosis and inflammation (Fig. 2C).

166 We successfully assigned 7616 spatial spots, corresponding to distinct 19794 transcript isoforms in total,  
167 encoded by 12474 genes. Transcript isoforms were largely associated with exact matches to the reference  
168 annotation, multi-exon transcripts with at least one junction match to the reference (*e.g.* exon skipping  
169 and exon extension), transcripts longer than the reference (containment of reference), completely novel  
170 transcripts (intergenic), transcripts with exonic overlap, or intronic transcripts (Fig. 2D). Among all genes,  
171 we observed 8131 (67,3%) that expressed a single isoform and 3953 (32.7%) that expressed 2 or more  
172 isoforms (Fig. 2E). Although predicted by our assembly, genes with many isoforms were expressed at a  
173 lower threshold and were not included in our final analyses. We also noticed variations in the number of  
174 isoforms per gene across morphological regions of each heart slice, with significant differences between

175 either of the remote zones and the border and the infarct areas, suggesting a higher transcriptome diversity  
176 in the healthy regions (Fig. 2F).

### 177 0.1 Regional isoform switching ...

178 A few interesting cases: Actc1, Crip2, Tnni3, Tmsb4x, Myh7, Sparc, Clu, etc. Need to add deconvolution  
179 results: can we identify cellular origin of some of the observed results? We need a way to visualize the  
180 different isoforms to see if they all make sense! Snapshots from IGV, etc. Add biological context: MI

## DISCUSSION

### DATA AVAILABILITY

181 The datasets [GENERATED/ANALYZED] for this study can be found in the [NAME OF REPOSITORY]  
182 [LINK].

### CODE AVAILABILITY

### AUTHOR CONTRIBUTIONS

183 TBD

### FUNDING

184 TBD

### ACKNOWLEDGMENTS

185 TBD

### CONFLICT OF INTEREST

186 The authors declare that they have no conflict of interest.

### SUPPLEMENTAL DATA

187 **Supplementary Figure S1**

188 SCNaST methodology.

189 **Supplementary Figure S2**

190 Quality control (Illumina).

191 **Supplementary Figure S3**

192 Quality control (Nanopore).

193 **Supplementary Figure S4**

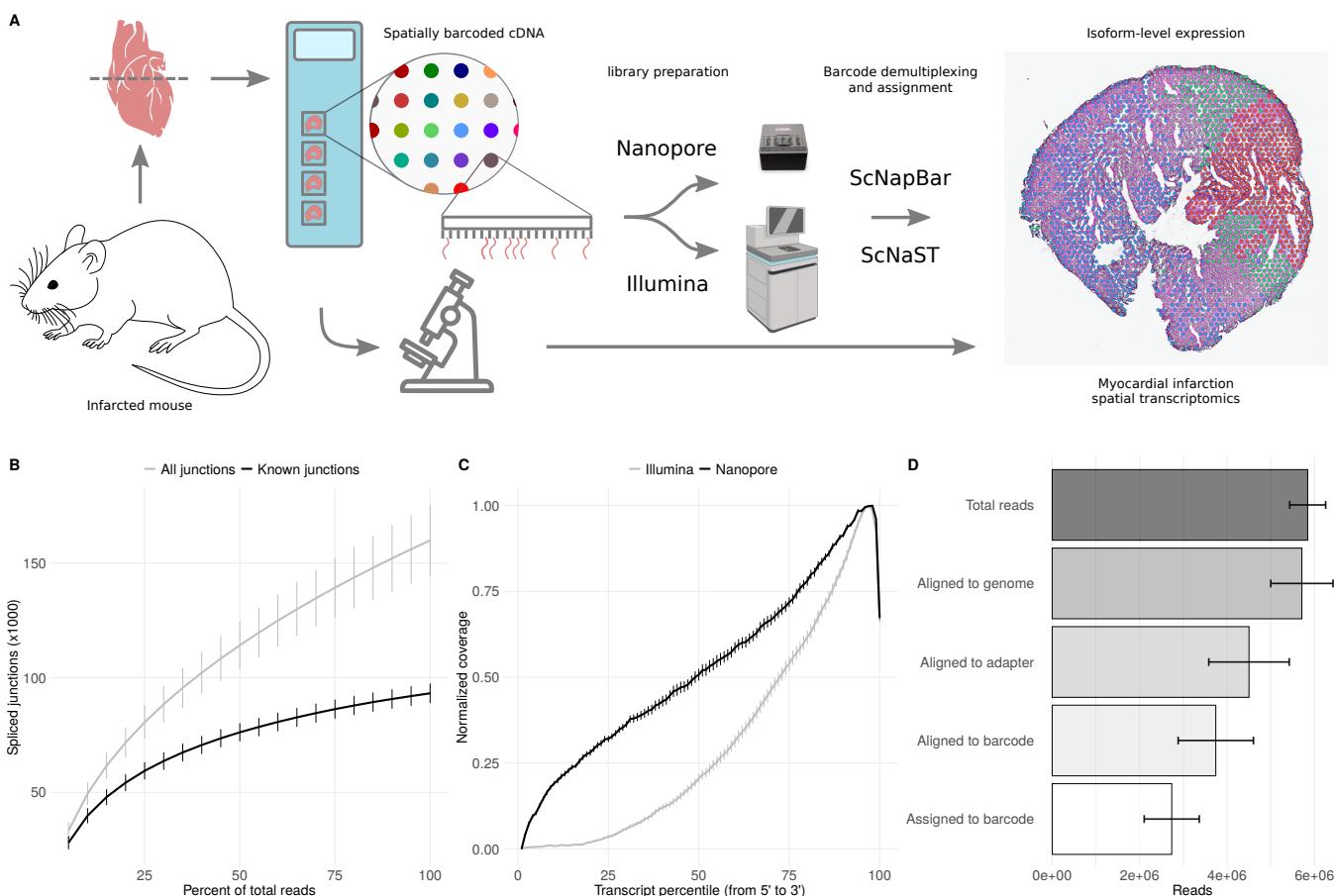
194 Spatial distribution of UMIs or counts and genes or transcripts.

## REFERENCES

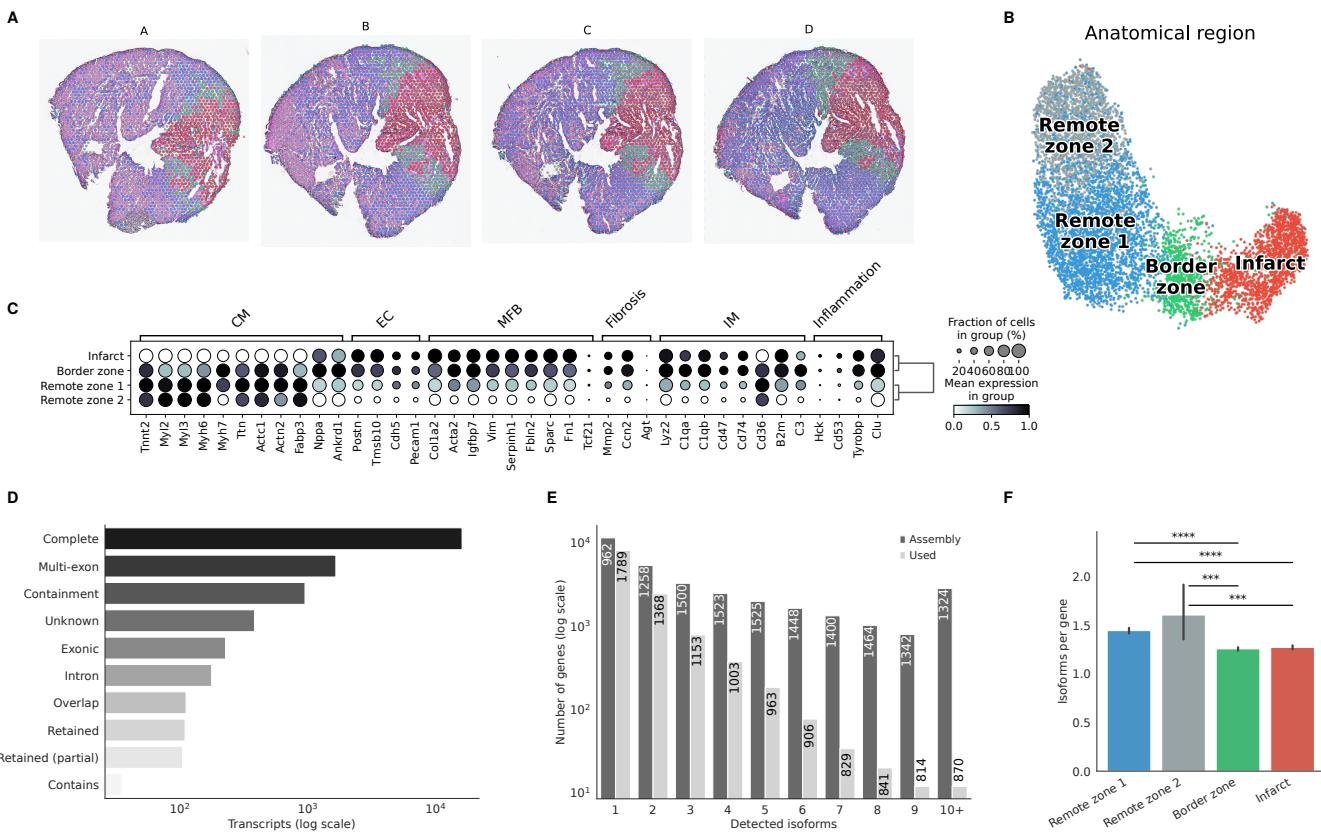
- 195 Cui, M., Wang, Z., Chen, K., Shah, A. M., Tan, W., Duan, L., et al. (2020). Dynamic transcriptional  
196 responses to injury of regenerative and non-regenerative cardiomyocytes revealed by single-nucleus  
197 rna sequencing. *Developmental Cell* 53, 102–116.e8. doi:<https://doi.org/10.1016/j.devcel.2020.02.019>.  
198 PMID: 32220304
- 199 Forte, E., Skelly, D. A., Chen, M., Daigle, S., Morelli, K. A., Hon, O., et al. (2020). Dynamic interstitial  
200 cell response during myocardial infarction predicts resilience to rupture in genetically diverse mice. *Cell  
201 Reports* 30, 3149–3163.e6. doi:<https://doi.org/10.1016/j.celrep.2020.02.008>
- 202 Gladka, M. M., Kohela, A., Molenaar, B., Versteeg, D., Kooijman, L., Monshouwer-Kloots, J., et al. (2021).  
203 Cardiomyocytes stimulate angiogenesis after ischemic injury in a zeb2-dependent manner. *Nature  
204 Communications* 12, 84. doi:10.1038/s41467-020-20361-3. PMID: 33398012
- 205 Heinrichs, M., Ashour, D., Siegel, J., Büchner, L., Wedekind, G., Heinze, K. G., et al. (2021). The healing  
206 myocardium mobilizes a distinct B-cell subset through a CXCL13-CXCR5-dependent mechanism.  
207 *Cardiovascular Research* 117, 2664–2676. doi:10.1093/cvr/cvab181. PMID: 34048536
- 208 Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C. L., Lindberg, E. L., et al.  
209 (2020). Cells of the adult human heart. *Nature* 588, 466–472. doi:10.1038/s41586-020-2797-4. PMID:  
210 32971526
- 211 Molenaar, B., Timmer, L. T., Droog, M., Perini, I., Versteeg, D., Kooijman, L., et al. (2021). Single-cell  
212 transcriptomics following ischemic injury identifies a role for b2m in cardiac repair. *Communications  
213 Biology* 4, 146. doi:10.1038/s42003-020-01636-3. PMID: 33514846
- 214 Müller, T., Boileau, E., Talyan, S., Kehr, D., Varadi, K., Busch, M., et al. (2021). Updated and enhanced  
215 pig cardiac transcriptome based on long-read rna sequencing and proteomics. *Journal of Molecular and  
216 Cellular Cardiology* 150, 23–31. doi:<https://doi.org/10.1016/j.yjmcc.2020.10.005>. PMID: 33049256
- 217 Park, J.-E., Botting, R. A., Conde, C. D., Popescu, D.-M., Lavaert, M., Kunz, D. J., et al. (2020). A  
218 cell atlas of human thymic development defines t cell repertoire formation. *Science* 367, eaay3224.  
219 doi:10.1126/science.aay3224
- 220 Ruiz-Villalba, A., Romero, J. P., Hernández, S. C., Vilas-Zornoza, A., Fortelny, N., Castro-Labrador,  
221 L., et al. (2020). Single-cell rna sequencing analysis reveals a crucial role for cthrc1 (collagen triple  
222 helix repeat containing 1) cardiac fibroblasts after myocardial infarction. *Circulation* 142, 1831–1847.  
223 doi:10.1161/CIRCULATIONAHA.119.044557. PMID: 32972203
- 224 Schaum, N., Karkanias, J., Neff, N. F., May, A. P., Quake, S. R., Wyss-Coray, T., et al. (2018). Single-  
225 cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature* 562, 367–372. doi:10.1038/  
226 s41586-018-0590-4. PMID: 30283141
- 227 Tombor, L. S., John, D., Glaser, S. F., Luxán, G., Forte, E., Furtado, M., et al. (2021). Single cell sequencing  
228 reveals endothelial plasticity with transient mesenchymal activation after myocardial infarction. *Nature  
229 Communications* 12, 681. doi:10.1038/s41467-021-20905-1. PMID: 33514719
- 230 Tucker, N. R., Chaffin, M., Fleming, S. J., Hall, A. W., Parsons, V. A., Bedi, K. C., et al. (2020).  
231 Transcriptional and cellular diversity of the human heart. *Circulation* 142, 466–482. doi:10.1161/  
232 CIRCULATIONAHA.119.045401. PMID: 32403949
- 233 Vafadarnejad, E., Rizzo, G., Krampert, L., Arampatzi, P., Arias-Loza, A.-P., Nazzal, Y., et al. (2020).  
234 Dynamics of cardiac neutrophil diversity in murine myocardial infarction. *Circulation Research* 127,  
235 e232–e249. doi:10.1161/CIRCRESAHA.120.317200. PMID: 32811295
- 236 van Duijvenboden, K., de Bakker, D. E., Man, J. C., Janssen, R., Günthel, M., Hill, M. C., et al.  
237 (2019). Conserved  $\beta$ -nppb $\beta$ -i $\beta$ + border zone switches from mef2- to ap-1&#x2013;driven gene program.  
238 *Circulation* 140, 864–879. doi:10.1161/CIRCULATIONAHA.118.038944. PMID: 31259610

239 Wang, L., Yu, P., Zhou, B., Song, J., Li, Z., Zhang, M., et al. (2020). Single-cell reconstruction of the adult  
 240 human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function.  
 241 *Nature Cell Biology* 22, 108–119. doi:10.1038/s41556-019-0446-7. PMID: 31915373

## FIGURE CAPTIONS



**Figure 1.** SCNaST methodology. A Schematic of the SCNaST workflow using a hybrid sequencing approach on Nanopore and Illumina platforms to assign spatial barcodes to long-read sequencing. B Nanopore sequencing saturation showing the number of splice sites detected at various levels of subsampling. A curve that reaches a plateau before getting to 100% data suggest that all known junctions in the library have been detected. The curve shows the mean  $\pm$  SE of four samples. C Normalized transcript coverage for Nanopore and Illumina. The curves show the mean  $\pm$  SE of four samples. D Reads assigned by SCNAPBAR at each step of the workflow. The bars show the mean  $\pm$  SE of four samples.



**Figure 2.** Defining morphological regions after MI. **A** Annotation of mouse heart regions after MI via short-read clustering, transferred to the Nanopore data. **B** UMAP representation of the Nanopore data using the region annotation from short-read clustering. **C** Dot plot showing the expression of selected markers associated with the expression of CM=cardiomyocytes, EC=endothelial cells, MFB=myofibroblasts, IM=immune cells, or with fibrosis and inflammation. **D** Barplot showing how full-length transcripts obtained with scNAST compare to the existing mouse annotation. Labels Complete (=), Multi-exon (j), Containment (k), Unknown (u), Exonic (x), Intron(i), Overlap (o), Retained (m, n), and Contains (y) are explained in <https://ccb.jhu.edu/software/stringtie/gffcompare.shtml>. **E** Barplot showing the frequency distribution of the number of isoforms per gene, either stemming from the assembly, or found in the final data after quality control filtering. The median length of transcripts is indicated in each bar for each category. **F** Average number of isoform per gene detected for each morphological region. Significance was measured using a Mann-Witney U-test (\*\* = <0.001, \*\*\*\* = <0.0001)