

# Single-cell full-length spatial transcriptomics reveal the unexplored isoform diversity of the myocardium post-MI

Etienne Boileau<sup>1,2,3</sup>, Xue Li<sup>2,3</sup>, Christian Becker<sup>4</sup>, Ramona Casper<sup>4</sup>, Janine Altmüller<sup>5</sup>, Florian Leuschner<sup>2,3</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>Cologne Center for Genomics (CCG), University of Cologne, Weyertal 115b, 50931 Cologne, Germany

<sup>5</sup>Berlin Institute of Health at Charité-Universitätsmedizin Berlin, Max Delbrück Center for Molecular Medicine

Correspondence\*:

Christoph Dieterich

christoph.dieterich@uni-heidelberg.de

## 2 ABSTRACT

We introduce Single-cell Nanopore Spatial Transcriptomics (scNAST), a software suite 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 constructed a *de novo* transcriptome using long-read data, and successfully assigned 19,794 transcript isoforms in total, including clinically-relevant, but yet uncharacterized modes of transcription, such as intron retention or antisense overlapping transcription. We showed a higher transcriptome complexity in the healthy regions, and identified intron retention as a mode of transcription associated with the infarct area. Our data revealed a clear regional isoform switching among differentially used transcripts for genes involved in cardiac muscle contraction and tissue morphogenesis. 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.

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

Recent development in spatial transcriptomics addresses this challenge, but few studies only have provided spatially resolved insights into the cardiac transcriptome. Techniques such as microdissection (Wu et al., 2016; Burkhard and Bakkers, 2018) or, in particular, spatially barcoded arrays and *in situ* capturing, have enabled to retain the spatial information while profiling the whole transcriptome at near single-cell resolution, allowing to shed light on localized tissue neighbourhoods (Asp et al., 2017, 2019), but a detailed characterization of cellular zones of injury, repair and/or remodeling is lacking. MI is a complex spatio-temporal heterogeneous disease involving the whole heart, and unbiased spatial transcriptomics holds a promise to add tissue context to molecular profiling in the search of novel therapeutics.

One of the most established spatial transcriptomics methods is now widely available as the Visium platform by 10x Genomics. After the tissue section is fixed on the spatial slide, stained, and imaged, it is permeabilized to release RNA to bind to capture probes for on-slide cDNA synthesis. Library preparation is performed off-slide, and spatial barcodes and tissue image are used to overlay transcriptomics data with tissue context. Although several new such methods have recently been published (Rodrigues et al., 2019; Liu et al., 2020), they are for the most part relying on short-read library preparation. Thus they fail to generate sufficient overlaps to reconstruct transcriptomes *de novo*.

Adequate gene and transcript models are instrumental towards relevant proof of concepts and investigational new drug development in translational cardiac research (Müller et al., 2021). Third-generation or long-read sequencing allow to reconstruct truthful assemblies with fewer gaps and to characterize complete transcript isoforms and chimeric transcripts. This has recently been taken to the single-cell level using either Pacific Biosciences (PacBio) or Oxford Nanopore technology (Gupta et al., 2018; Lebrigand et al., 2020; Volden and Vollmers, 2021; Wang et al., 2021; Joglekar et al., 2021).

Here, taking advantage of the conceptual similarity between spatial and cell barcodes, we introduce Single-cell Nanopore Spatial Transcriptomics (SCNAST), a set of tools to facilitate whole transcriptome spatial profiling of full-length transcripts, based on a previously published Bayesian approach for cell barcode assignment (Wang et al., 2021). Our method relies on the commercially available Visium platform by 10x Genomics and Nanopore long-read sequencing, although it can in principle be extended to other technologies, as long as a hybrid sequencing approach is used for spatial barcode assignment.

In this short report, we demonstrate how SCNAST can be used to characterize the spatial transcriptional landscape of the heart post-MI. Using our workflow, we were able to assign a total of 7,616 spatial spots across four short axis sections of the heart, corresponding to distinct 19,794 transcript isoforms in total, encoded by 12,474 genes. We showed a higher transcriptome complexity in the healthy regions, and identified intron retention as a mode of transcription associated with the injured areas. Our results showed a

61 clear regional isoform switching among differentially used transcripts for genes involved in cardiac muscle  
62 contraction and tissue morphogenesis, many of them clinically relevant, opening new opportunities in  
63 translational cardiac research.

## MATERIAL AND METHODS

### 64 Experimental model of myocardial infarction

65 A C57BL/6 mouse (female, 8 weeks old, Janvier Labs) was exposed to 5% isoflurane for anesthesia. An  
66 intubation cannula was inserted orally into the trachea. The mouse was fixed on a heating plate at 37°C  
67 and maintained under anesthesia with 2% isoflurane. An incision was made from the left sternum to the  
68 midclavicular line. Skins and muscle layers were stretched with forceps and sutures. Another incision was  
69 made between the third and fourth intercostal space. The heart was exposed and subjected to permanent  
70 myocardial infarction with ligation of the left anterior descending (LAD) coronary artery. When the ribs  
71 and skins were fully closed, the isoflurane supply was cut off. Oxygen was then supplied until normal  
72 breathing was resumed.

### 73 Heart extraction and cryosection

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

### 83 10X Genomics Visium experiments

84 Four short axis sections of the heart were processed according to the manufacturer's protocol. Libraries  
85 were prepared individually, one per heart section. The Visium Spatial Tissue Optimization Slide & Reagent  
86 kit (10x Genomics) was used to optimize permeabilization conditions. Tissue permeabilization was  
87 performed for 24 min. Spatially barcoded full-length cDNA was generated using the Visium Spatial Gene  
88 Expression Slide & Reagent kit (10x Genomics). A fraction of each cDNA library was used for Nanopore  
89 sequencing. cDNA amplification was then conducted for 20 cycles of PCR (identified by qPCR), and  
90 400µl were used in the 10xGenomics Visium library preparation (100µl per section). The libraries were  
91 sequenced on a NovaSeq6000 (Illumina), with 29 bases from read 1 and 90 bases from read 2, at a depth  
92 of 160M reads per section (640M reads in total). The raw sequencing data was processed with the 10x  
93 Genomics Space Ranger v1.1.2 and mapped to the mm10 genome assembly (mm10-2020-A).

### 94 Oxford Nanopore sequencing libraries

95 Libraries for Nanopore sequencing were prepared according to the manufacturer's protocol for direct  
96 cDNA Sequencing (SQK-DCS109 Oxford Nanopore Technologies) with the following minor modifications:  
97 The protocol was started with 200ng cDNA with the End-prep step. Final library elution was performed in  
98 15µl to have material left for TapeStation analysis and Qubit BR measurement. Four GridIon flow cells

99 (FLO-MIN106) were loaded with 12 $\mu$ L libraries (74-115ng) by using a Flow Cell Priming kit EXP-FLP002.  
100 Base calling was done with Guppy v5.0.12. The High accuracy (HAC) model was selected for base calling  
101 (Q-Score cut-off > 9).

## 102 Spatial barcode assignment

103 To account for source-specific quality differences, each heart section (Illumina libraries) was processed  
104 separately using Scanpy v1.7.2 (Wolf et al., 2018), keeping only spatial barcodes with approximately 150  
105 < counts < 18000, 250 < genes < 5000, detected in at least 2 spots, and with less than 40% mitochondrial  
106 counts. The resulting datasets were concatenated, normalized, and the union of highly variable genes (per  
107 section) were kept for final analysis. Batch balanced KNN (BBKNN) (Polański et al., 2020) with ridge  
108 regression (Park et al., 2020) was used for integration and batch correction, starting from a coarse clustering  
109 obtained from a BBKNN-corrected graph.

110 For the Nanopore libraries, samples were demultiplexed and processed with ScNapBar v1.1.0 (<https://github.com/dieterich-lab/single-cell-nanopore>) using a naive Bayes model to  
111 assign spatial barcodes (Wang et al., 2021). Briefly, for each heart section, the Space Ranger results  
112 (Illumina libraries) were used to parametrize a model of barcode alignment features to discriminate correct  
113 versus false barcode assignment in the Nanopore data. FASTQ files were mapped using minimap2 v2.21 (Li,  
114 2018). For transcript isoform quantification, a *de novo* transcriptome annotation was generated. Alignment  
115 files were processed by StringTie v2.1.5 in long read mode with the reference annotation to guide the  
116 assembly process (Kovaka et al., 2019). The annotations were merge into a non-redundant set of transcripts  
117 and compared to the reference using GffCompare v0.12.2 (Pertea and Pertea, 2020), after removing single-  
118 exon transcripts. To generate feature-spatial barcode matrices, alignment files were split into multiple files,  
119 one per spatial barcode, based on the barcode assignments, converted to FASTQ, and aligned to the *de*  
120 *novo* transcriptome with minimap2. Abundances were quantified with Salmon v1.5.2 in alignment-based  
121 mode using a long read error model (Patro et al., 2017). Each section was processed separately using  
122 Scanpy v1.7.2 and integrated with BBKNN, as described above for the Illumina data. Spatial barcodes  
123 were filtered for counts (approximately 50 < counts < 4000), transcripts (approximately 50 < transcripts  
124 < 2000, detected in at least 2 spots), and ribosomal genes (< 40%). Neighbors enrichment and cluster  
125 co-occurrence analyses were performed using Squidpy (Palla et al., 2021).

## 127 Identification of anatomical regions

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

133 Labels were then assigned to Nanopore spatial barcodes based on the set of matching Illumina barcodes.  
134 However, not all assigned barcodes were labeled due to quality control filtering criteria that were different  
135 between Illumina and Nanopore datasets. To assign labels to the remaining Nanopore barcodes, seed  
136 labelling was performed with scANVI using the set of assigned labels as groundtruth (Xu et al., 2021). The  
137 top expressed transcripts were then identified as described above for the Illumina data.

**138 Spatial spot deconvolution**

139 Spatial spots were deconvoluted using stereoscope (scvi-tools v0.15.0) (Andersson et al., 2020). Heart  
140 data (Smart-Seq2 and 10x Genomics) from the Tabula Muris (Schaum et al., 2018) were used as reference  
141 dataset and highly variable genes were identified. For the Smart-Seq2 data, gene length normalization was  
142 applied. The model was trained on the single cell reference dataset on the intersection of genes found in  
143 the spatial (Illumina) data, and proportions were inferred for each Visium spot for each cell type in the  
144 reference dataset. Labels were then assigned to Nanopore spots are described above.

**145 Spatial gene expression (Illumina)**

146 For each heart section, genes with spatial expression patterns were obtained with SPARK (FDR <  
147 0.05) (Sun et al., 2020). Overrepresented biological processes in each region were identified using spatially  
148 variable markers that were previously identified ( $\log FC < 0.2$ , p-value < 0.05), using Enrichr with the  
149 Python package GSEApY.

**150 Differential transcript usage (Nanopore)**

151 To identify changes in relative usage of transcripts/isoforms within genes, differential transcript usage  
152 (DTU) tests were performed using satuRn (Gillis et al., 2021). Only multi-exon transcripts and genes  
153 with more than one isoform were kept for the analyses. The transcript count matrix was further filtered  
154 to keep transcripts expressed in a worthwhile number of spots, determined by the design (but greater  
155 than at least 10% of the smallest group size), with a CPM count above a threshold of 1(median library  
156 size) $^{-1}$ . In addition, transcripts were kept only if they had a minimum count of 1 across all spots. A quasi-  
157 binomial generalized linear model was fit using a design comparing each anatomical region with another,  
158 or each anatomical region with the rest of all regions. A two-stage testing procedure was performed using  
159 stageR (Van den Berge and Clement, 2021), with an OFDR of 0.05. Results were reported using a student's  
160 t-test statistic, computed with estimated log-odds ratios. To identify isoform-switching genes, significant  
161 genes were identified for each contrast with at least two transcripts showing a switching expression pattern  
162 between regions of interest.

**163 Transcript classes**

164 Transcript classes were assigned in scNAST using Gffcompare (Pertea and Pertea, 2020). For the  
165 identification of isoforms and transcript classes between regions, markers previously identified were used to  
166 select transcripts in each region with a  $\log FC > 0.1$  (p-value < 0.05). Enrichment of transcripts of a certain  
167 class (e.g. intron retention) was calculated using a Fisher exact test, using only non-equality classes. Motif  
168 enrichment was performed using Simple Enrichment Analysis (SEA) from the MEME suite, using retained  
169 intron sequences vs. annotated (reference transcript) sequences (Bailey et al., 2015).

## RESULTS

**170 scNAST enables the demarcation of spatially distinct regions of the myocardium post-MI**

171 Fresh-frozen tissue samples were stained, imaged and fixed on Visium Spatial Gene Expression Slides  
172 (10X Genomics) for permeabilization and *in situ* RNA capture. Full-length cDNA libraries were split for  
173 the preparation of 3' Illumina short-read and direct long-read Nanopore sequencing libraries. Short-read  
174 data were used for the assignment of spatial barcodes to Nanopore reads using the scSNAPBAR workflow,  
175 and subsequently used to define anatomical regions within the tissue organization (Fig. 1A). Long-read

176 data were used for transcriptome assembly and transcript abundance quantification, and layered onto the  
177 stained images to reveal the spatial organization of isoform expression. The Nanopore data comprises  
178 of four heart slices (or samples) with a total of 25,5 million reads, reaching a relatively high sequencing  
179 saturation (Fig. 1B), and providing a significant gain in coverage along full-length transcripts (Fig. 1C).  
180 After spatial barcode assignment, libraries had a median of 2.8 million reads per sample, with over 70%  
181 assigned reads (Figs. 1D and Supplementary Figure S1A). Despite variations across samples, we observed  
182 a good per spot correlation between Illumina and Nanopore libraries (Fig. Supplementary Figure S1B,C).  
183 Per spatial spot ( $55\mu\text{m}$ ), after quality filtering, we observed a median read count varying between 593 and  
184 2021 corresponding to a median of 311 (respectively 1000) distincts isoforms (Fig. Supplementary Figure  
185 S2, see also Figs. Supplementary Figure S3 Supplementary Figure S4).

186 Clustering based on short-read gene expression defined four broad morphological regions: two remote  
187 zones that stemmed from differences in sequencing depth between heart slices, a border zone, and the infarct.  
188 Remote zones and, to a lesser extent the border zone, were largely associated with cardiomyocyte markers,  
189 while the border and infarct zones were characterized by a higher expression of endothelial, myofibroblast,  
190 and immune marker genes, as well as with markers of fibrosis and inflammation (Fig. 2A). The region  
191 classification was then transferred to the Nanopore data (Fig. 2B), and markers of each region were identified  
192 (Fig. Supplementary Figure S5). We also investigated, using the Illumina libraries, whether our spatial  
193 transcriptomics data reflected known biological processes of myocardial infarction. We identified spatially  
194 variable genes across samples and characterized each region using biological processes (Fig. Supplementary  
195 Figure S6, and supporting information **Supplementary Table S1**). Remote zones were generally associated  
196 with cardiac muscle contraction linked to an overrepresentation of healthy cardiomyocytes, and included  
197 genes such as *Strt1* (SERCA regulator DWORF), cardiac troponin I (*Tnni3*), myosin-binding protein C  
198 (*Mybpc3*), or ventricular myosin light chain-2 (*Myl2*). The border zone was characterized by genes such  
199 as *Nppb* or *Ankrd1*, both of which were reported to be upregulated in the border zone after MI (Hama  
200 et al., 1995; Mikhailov and Torrado, 2008), and other genes associated with the complement system and  
201 myogenesis. Overrepresented processes were however similar to those found in the infarct area. Top  
202 markers of the infarct area included collagens and genes associated with TGF- $\beta$  and p53 signalling, with  
203 hypoxia, coagulation, epithelial to mesenchymal transition, and the extracellular matrix. Overrepresented  
204 processes included neutrophil degranulation, and gene expression associated with the innate immune  
205 system. Remote, border, and infarct zones had a clear distinct spatial organization: the two remote zones  
206 co-occurred at short distances with one another, but did not show any neighborhood enrichment with the  
207 border and infarct zones (Fig. 2C,D). We observed a slight co-enrichment of the infarct and the border  
208 zone at medium distances (Fig. 2D), revealing how spatial resolution may, in general, affect morphological  
209 classification.

210 Following this strategy, we successfully assigned 7,616 spatial spots, corresponding to distinct 19,794  
211 transcript isoforms in total, encoded by 12,474 genes. Among all genes, we observed 8,131 (67,3%) that  
212 expressed a single isoform and 3,953 (32.7%) that expressed 2 or more isoforms (Fig. 2E). Although  
213 predicted by our assembly, genes with many isoforms were expressed at a lower threshold and were  
214 not included in our final analyses. We observed variations in the number of isoforms per gene across  
215 morphological regions of each heart slice, with significant differences between either of the remote zones  
216 and the border and the infarct areas, suggesting the existence of a higher transcriptome diversity in the  
217 healthy regions (Fig. 2F).

**218 scNAST reveals the spatial isoform diversity of the myocardium post-MI**

219 Transcript isoforms were largely associated with exact matches to the reference annotation, multi-exon  
220 transcripts with at least one junction match to the reference (*e.g.* exon skipping and exon extension),  
221 transcripts longer than the reference (containment of reference), completely novel transcripts (intergenic),  
222 transcripts with exonic overlap, intronic transcripts, or retained introns (Fig. 3A). Among these classes,  
223 we were particularly interested in novel transcripts associated with retained introns. Intron retention (IR)  
224 occurs when an intron is transcribed into pre-mRNA and remains in the final mRNA. Only recently has  
225 IR become of interest due to its associations with complex diseases (Zhang et al., 2020). Interestingly,  
226 among differentially expressed transcript markers, we found across the infarct and border zones 65 distinct  
227 IR transcripts, compared to 1,612 in the remote zones, corresponding to an odds ratio of 1.73 (p-value  
228 0.008). When considering the infarct area only, the odds were even greater (odds ratio 2, p-value 0.002).  
229 Introns that were retained across the infarct and border zone were enriched in motifs associated with  
230 poly(C)-binding protein 1 and 3 (Pcbp1/3), KH RNA-binding domains (Khdrbs1), and Rbm38, a homolog  
231 of Rbm24, a pivotal cardiac splice factor (Weeland et al., 2015) (supporting information **Supplementary**  
232 **Table S2**). The genes harboring these transcripts were enriched in RNA metabolic/catabolic processes,  
233 RNA binding, and the unfolded protein response (Fig. 3B). Among these, we found cardiac muscle alpha  
234 actin (Actc1), the major protein of the thin filament responsible for cardiac contraction (Fig. 3C). Our data  
235 revealed a relatively high expression of the IR transcript isoform across all regions, but a comparatively  
236 greater contribution to the infarct area (Fig. 3D). At the Actc1 locus, we also identified two novel antisense  
237 overlapping transcripts with a good coverage across all regions (Fig. 3C). There is now growing evidence  
238 that antisense-mediated gene regulation is involved in different pathophysiological contexts, including  
239 heart disease (Luther, 2005; Zinad et al., 2017; Celik et al., 2019). Recently, a long non-coding RNA  
240 transcribed antisense near the Actc1 gene was described for its role in cardiomyocyte proliferation and  
241 cardiac repair (Ponnusamy et al., 2019). Using the Tabula Muris (Schaum et al., 2018) as a reference  
242 dataset to perform deconvolution, we identified cardiomyocytes, and to a lesser extent, endothelial and  
243 myofibroblasts as the predominant origin of these novel antisense isoforms, with a correlation pattern that  
244 matched that of the largest Actc1 annotated isoform (Fig. 3E).

**245 scNAST allows to study regional isoform switching with spatial context**

246 Among the multi-isoform genes, we looked for those showing a splicing pattern variation across  
247 regions. We identified 109 significant regional isoform switching genes across all comparisons between  
248 the remote, border, and infarct areas (OFDR 0.05 at both gene and transcript level, supporting  
249 information **Supplementary Table S3**). Our results showed a clear regional isoform switching among  
250 differentially used transcripts for genes involved in cardiac muscle contraction and tissue morphogenesis,  
251 many of them clinically relevant, such as Ankrd1 (Mikhailov and Torrado, 2008), Sparc (McCurdy et al.,  
252 2011), or Clu (Turkieh et al., 2019) (Fig. 4A). The largest number of isoform-switching genes were found  
253 between the infarct and the remote zones (Fig. 4B). Among them, we found PDZ and LIM domain protein  
254 5 (Pdlim5), a gene encoding for a protein that localizes to the Z-disk by binding to  $\alpha$ -actinin, and which has  
255 been implicated in dilated cardiomyopathy (Verdonschot et al., 2020) and in heart failure with preserved  
256 ejection fraction (Soetkamp et al., 2021). Pdlim5 undergoes extensive splicing in both human and mouse.  
257 Our data showed a significant change in isoform usage between two annotated protein-coding isoforms  
258 between the infarct and the border or the remote zones (Fig. 4C,D,E). The shorter isoform was mostly  
259 found in the border and remote zones, while the longer isoform was significantly, albeit at lower levels,  
260 expressed in the infarct area (Fig. 4C,E).

## DISCUSSION

261 SCNaST expands the spatial transcriptomics toolbox to third-generation long-read sequencing. SCNaST is  
262 based on commercially available platforms, and can be used to investigate the isoform landscape of complex  
263 tissue. We presented here a transcriptome-wide approach to explore annotated and *de novo* isoform diversity  
264 with morphological context in the mouse heart after myocardial infarction (MI). Our approach opens-up  
265 new opportunities to understand the spatial and molecular organization of the heart following injury. The  
266 infarcted heart has usually been divided into three zones: the infarct, an ischemic border zone, roughly  
267 defined as the area immediately adjacent to the infarct, and a remote zone. Although previous studies  
268 showed transcriptional differences between proximal and distal areas to the infarct (van Duijvenboden  
269 et al., 2019a), observations were limited to conventional short-read sequencing approaches or were limited  
270 in spatial resolution. Our results not only reflected known biological processes of myocardial infarction,  
271 but also revealed the spatial organization of gene expression and transcriptome diversity consistent with  
272 the underlying biological condition. We identified novel isoforms in clinically relevant loci and regional  
273 isoform switches that could play a role in the myocardium post-MI. We envision that our work will serve  
274 as a reference for future developments and studies integrating long-read sequencing with spatial gene  
275 expression data.

## DATA AVAILABILITY

276 The datasets generated for this study can be found in the [to be deposited at SRA]. Additional  
277 information, supporting data and scripts to generate the figures are available at <https://github.com/dieterich-lab/ScNaST>.  
278

## CODE AVAILABILITY

279 SCNAPBAR is publicly available at:  
280 <https://github.com/dieterich-lab/single-cell-nanopore>. SCNaST is publicly  
281 available at:  
282 <https://github.com/dieterich-lab/ScNaST>.

## AUTHOR CONTRIBUTIONS

283 CD, FL supervised the research and acquired funding. JA planned and supervised Visium experiments  
284 and library preparation. XL performed MI experiments and tissue extraction. CB, RC performed Visium  
285 experiments and library preparation. EB was in charge of software development, analyzed the data and  
286 wrote the manuscript.

## FUNDING

287 EB and CD acknowledge support by the Klaus Tschira Stiftung gGmbH [00.219.2013] and by Informatics  
288 for Life funded by the Klaus Tschira Foundation. CD acknowledges the DZHK (German Centre for  
289 Cardiovascular Research) Partner Site Heidelberg/Mannheim.

## ACKNOWLEDGMENTS

290 EB and CD would like to thank Harald Wilhelm for maintaining the high-performance compute cluster of  
291 the Dieterich Lab.

## CONFLICT OF INTEREST

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

## SUPPLEMENTAL DATA

293 **Supplementary Figure S1**

294 scNAST methodology.

295 **Supplementary Figure S2**

296 Quality control (Nanopore).

297 **Supplementary Figure S3**

298 Quality control (Illumina).

299 **Supplementary Figure S4**

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

301 **Supplementary Figure S5**

302 Top markers of each region.

303 **Supplementary Figure S6**

304 Enriched processes among spatially variable marker genes.

305 **Supplementary Table S1**

306 Spatially variable genes.

307 **Supplementary Table S2**

308 Motif enriched in IR transcripts.

309 **Supplementary Table S3**

310 Differential transcript usage (isoform-switching genes).

## REFERENCES

- 311 Andersson, A., Bergenstråhle, J., Asp, M., Bergenstråhle, L., Jurek, A., Fernández Navarro, J., et al.  
312 (2020). Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography.  
313 *Communications Biology* 3, 565. doi:10.1038/s42003-020-01247-y. PMID: 33037292
- 314 Asp, M., Giacomello, S., Larsson, L., Wu, C., Fürth, D., Qian, X., et al. (2019). A spatiotemporal  
315 organ-wide gene expression and cell atlas of the developing human heart. *Cell* 179, 1647–1660.e19.  
316 doi:<https://doi.org/10.1016/j.cell.2019.11.025>. PMID: 31835037

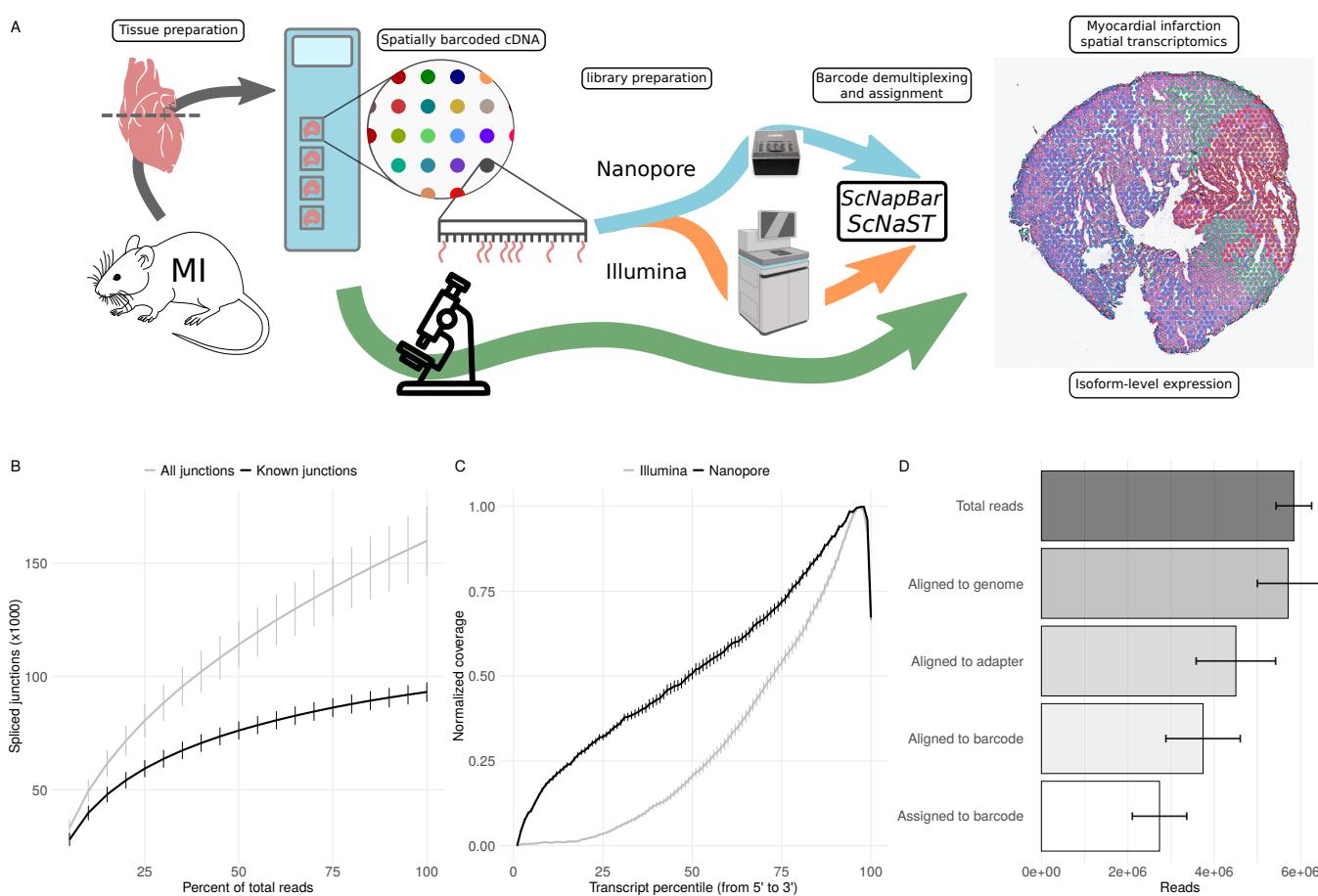
- 317 Asp, M., Salmén, F., Ståhl, P. L., Vickovic, S., Felldin, U., Löfing, M., et al. (2017). Spatial detection  
318 of fetal marker genes expressed at low level in adult human heart tissue. *Scientific Reports* 7, 12941.  
319 doi:10.1038/s41598-017-13462-5. PMID: 29021611
- 320 Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The meme suite. *Nucleic acids research*  
321 43, W39–W49. doi:10.1093/nar/gkv416. 25953851[pmid]
- 322 Burkhard, S. B. and Bakkers, J. (2018). Spatially resolved rna-sequencing of the embryonic heart  
323 identifies a role for wnt/b-catenin signaling in autonomic control of heart rate. *eLife* 7, e31515.  
324 doi:10.7554/eLife.31515. PMID: 29400650
- 325 Celik, S., Sadegh, M. K., Morley, M., Roselli, C., Ellinor, P. T., Cappola, T., et al. (2019). Antisense  
326 regulation of atrial natriuretic peptide expression. *JCI Insight* 4. doi:10.1172/jci.insight.130978. PMID:  
327 31503546
- 328 Cui, M., Wang, Z., Chen, K., Shah, A. M., Tan, W., Duan, L., et al. (2020). Dynamic transcriptional  
329 responses to injury of regenerative and non-regenerative cardiomyocytes revealed by single-nucleus  
330 rna sequencing. *Developmental Cell* 53, 102–116.e8. doi:https://doi.org/10.1016/j.devcel.2020.02.019.  
331 PMID: 32220304
- 332 Forte, E., Skelly, D. A., Chen, M., Daigle, S., Morelli, K. A., Hon, O., et al. (2020). Dynamic interstitial  
333 cell response during myocardial infarction predicts resilience to rupture in genetically diverse mice. *Cell  
334 Reports* 30, 3149–3163.e6. doi:https://doi.org/10.1016/j.celrep.2020.02.008
- 335 Gilis, J., Vitting-Seerup, K., Van den Berge, K., and Clement, L. (2021). saturn: scalable analysis of  
336 differential transcript usage for bulk and single-cell rna-sequencing applications [version 1; peer review:  
337 2 approved with reservations]. *F1000Research* 10. doi:10.12688/f1000research.51749.1
- 338 Gladka, M. M., Kohela, A., Molenaar, B., Versteeg, D., Kooijman, L., Monshouwer-Kloots, J., et al. (2021).  
339 Cardiomyocytes stimulate angiogenesis after ischemic injury in a zeb2-dependent manner. *Nature  
340 Communications* 12, 84. doi:10.1038/s41467-020-20361-3. PMID: 33398012
- 341 Gupta, I., Collier, P. G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., et al. (2018). Single-cell  
342 isoform rna sequencing characterizes isoforms in thousands of cerebellar cells. *Nature Biotechnology*  
343 36, 1197–1202. doi:10.1038/nbt.4259. PMID: 30320766
- 344 Hama, N., Itoh, H., Shirakami, G., Nakagawa, O., ichi Suga, S., Ogawa, Y., et al. (1995). Rapid ventricular  
345 induction of brain natriuretic peptide gene expression in experimental acute myocardial infarction.  
346 *Circulation* 92, 1558–1564. doi:10.1161/01.CIR.92.6.1558. PMID: 7664440
- 347 Heinrichs, M., Ashour, D., Siegel, J., Büchner, L., Wedekind, G., Heinze, K. G., et al. (2021). The healing  
348 myocardium mobilizes a distinct B-cell subset through a CXCL13-CXCR5-dependent mechanism.  
349 *Cardiovascular Research* 117, 2664–2676. doi:10.1093/cvr/cvab181. PMID: 34048536
- 350 Joglekar, A., Prjibelski, A., Mahfouz, A., Collier, P., Lin, S., Schlusche, A. K., et al. (2021). A  
351 spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nature  
352 Communications* 12, 463. doi:10.1038/s41467-020-20343-5. PMID: 33469025
- 353 Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome  
354 assembly from long-read rna-seq alignments with stringtie2. *Genome Biology* 20, 278. doi:10.1186/  
355 s13059-019-1910-1. PMID: 31842956
- 356 Lebrigand, K., Magnone, V., Barbry, P., and Waldmann, R. (2020). High throughput error corrected  
357 nanopore single cell transcriptome sequencing. *Nature Communications* 11, 4025. doi:10.1038/  
358 s41467-020-17800-6. PMID: 32788667
- 359 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100.  
360 doi:10.1093/bioinformatics/bty191. PMID: 29750242

- 361 Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C. L., Lindberg, E. L., et al.  
362 (2020). Cells of the adult human heart. *Nature* 588, 466–472. doi:10.1038/s41586-020-2797-4. PMID:  
363 32971526
- 364 Liu, Y., Yang, M., Deng, Y., Su, G., Enninfu, A., Guo, C. C., et al. (2020). High-spatial-resolution  
365 multi-omics sequencing via deterministic barcoding in tissue. *Cell* 183, 1665–1681.e18. doi:<https://doi.org/10.1016/j.cell.2020.10.026>. PMID: 33188776
- 367 Luther, H. P. (2005). Role of endogenous antisense rna in cardiac gene regulation. *Journal of Molecular  
368 Medicine* 83, 26–32. doi:10.1007/s00109-004-0613-5. PMID: 15592803
- 369 McCurdy, S. M., Dai, Q., Zhang, J., Zamilpa, R., Ramirez, T. A., Dayah, T., et al. (2011). Sparc mediates  
370 early extracellular matrix remodeling following myocardial infarction. *American journal of physiology.  
371 Heart and circulatory physiology* 301, H497–H505. doi:10.1152/ajpheart.01070.2010. 21602472[pmid]
- 372 Mikhailov, A. and Torrado, M. (2008). The enigmatic role of the ankyrin repeat domain 1 gene in heart  
373 development and disease. *Int J Dev Biol* 52, 811–821. doi:10.1387/ijdb.082655am. PMID: 18956313
- 374 Molenaar, B., Timmer, L. T., Droog, M., Perini, I., Versteeg, D., Kooijman, L., et al. (2021). Single-cell  
375 transcriptomics following ischemic injury identifies a role for b2m in cardiac repair. *Communications  
376 Biology* 4, 146. doi:10.1038/s42003-020-01636-3. PMID: 33514846
- 377 Müller, T., Boileau, E., Talyan, S., Kehr, D., Varadi, K., Busch, M., et al. (2021). Updated and enhanced  
378 pig cardiac transcriptome based on long-read rna sequencing and proteomics. *Journal of Molecular and  
379 Cellular Cardiology* 150, 23–31. doi:<https://doi.org/10.1016/j.yjmcc.2020.10.005>. PMID: 33049256
- 380 Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., et al. (2021). Squidpy: a  
381 scalable framework for spatial single cell analysis. *bioRxiv* doi:10.1101/2021.02.19.431994
- 382 Park, J.-E., Botting, R. A., Conde, C. D., Popescu, D.-M., Lavaert, M., Kunz, D. J., et al. (2020). A  
383 cell atlas of human thymic development defines t cell repertoire formation. *Science* 367, eaay3224.  
384 doi:10.1126/science.aay3224
- 385 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-  
386 aware quantification of transcript expression. *Nature Methods* 14, 417–419. doi:10.1038/nmeth.4197.  
387 PMID: 28263959
- 388 Pertea, G. and Pertea, M. (2020). Gff utilities: Gffread and gffcompare [version 1; peer review: 3 approved].  
389 *F1000Research* 2020 9, 304. doi:10.12688/f1000research.23297.1. PMID: 32489650
- 390 Polański, K., Young, M. D., Miao, Z., Meyer, K. B., Teichmann, S. A., and Park, J.-E. (2020). Bbknn: Fast  
391 batch alignment of single cell transcriptomes. *Bioinformatics* 36, 964–965. doi:10.1093/bioinformatics/  
392 btz625. PMID: 31400197
- 393 Ponnusamy, M., Liu, F., Zhang, Y.-H., Li, R.-B., Zhai, M., Liu, F., et al. (2019). Long noncoding rna  
394 cpr (cardiomyocyte proliferation regulator) regulates cardiomyocyte proliferation and cardiac repair.  
395 *Circulation* 139, 2668–2684. doi:10.1161/CIRCULATIONAHA.118.035832
- 396 Rodrigues, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., et al. (2019).  
397 Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.  
398 *Science* 363, 1463–1467. doi:10.1126/science.aaw1219. PMID: 30923225
- 399 Ruiz-Villalba, A., Romero, J. P., Hernández, S. C., Vilas-Zornoza, A., Fortelny, N., Castro-Labrador,  
400 L., et al. (2020). Single-cell rna sequencing analysis reveals a crucial role for cthrc1 (collagen triple  
401 helix repeat containing 1) cardiac fibroblasts after myocardial infarction. *Circulation* 142, 1831–1847.  
402 doi:10.1161/CIRCULATIONAHA.119.044557. PMID: 32972203
- 403 Schaum, N., Karkanias, J., Neff, N. F., May, A. P., Quake, S. R., Wyss-Coray, T., et al. (2018). Single-  
404 cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature* 562, 367–372. doi:10.1038/  
405 s41586-018-0590-4. PMID: 30283141

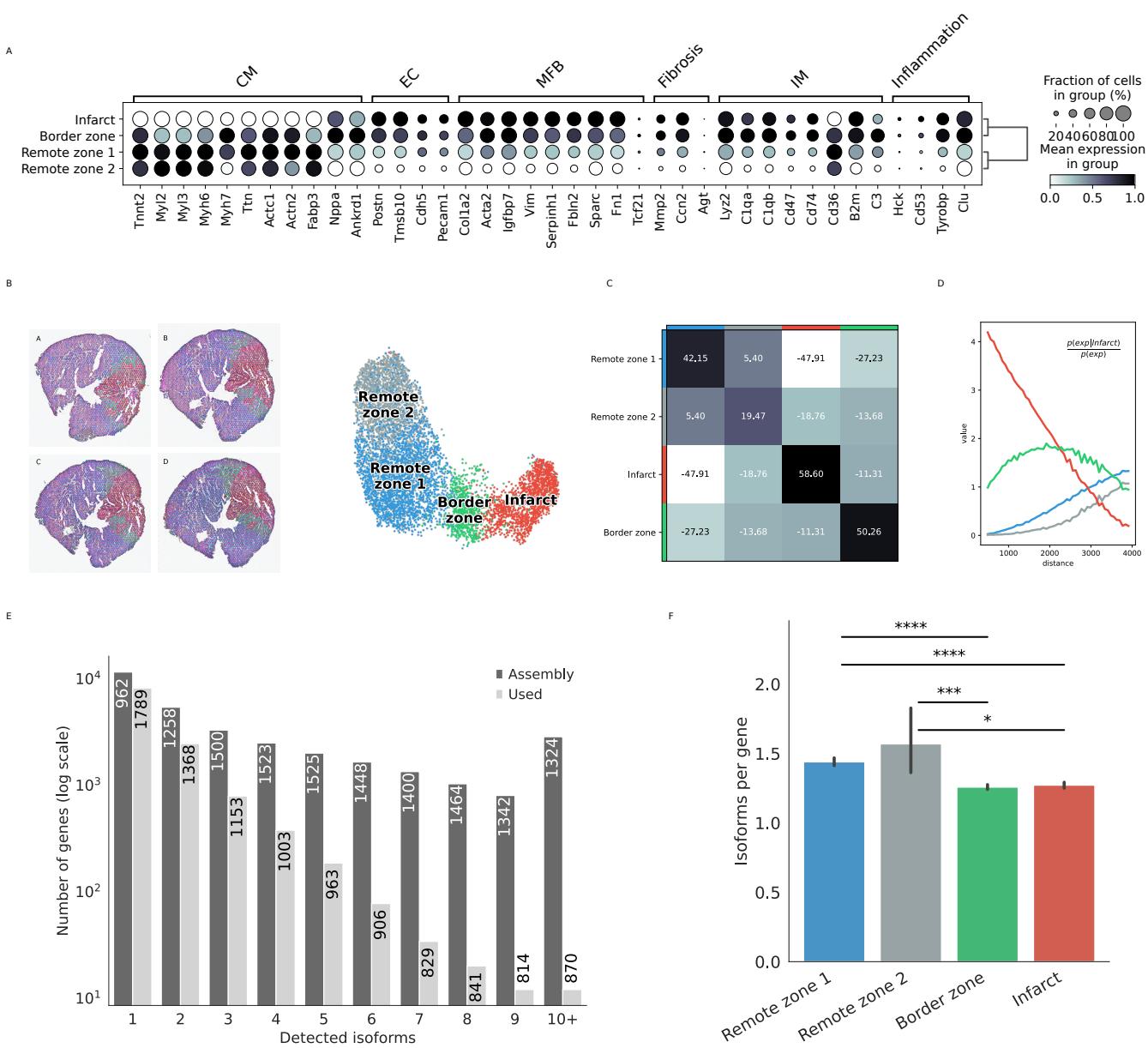
- 406 Soetkamp, D., Gallet, R., Parker, S. J., Holewinski, R., Venkatraman, V., Peck, K., et al. (2021).  
407 Myofilament phosphorylation in stem cell treated diastolic heart failure. *Circulation Research* 129,  
408 1125–1140. doi:10.1161/CIRCRESAHA.119.316311. PMID: 34641704
- 409 Sun, S., Zhu, J., and Zhou, X. (2020). Statistical analysis of spatial expression patterns for spatially  
410 resolved transcriptomic studies. *Nature Methods* 17, 193–200. doi:10.1038/s41592-019-0701-7
- 411 Tombor, L. S., John, D., Glaser, S. F., Luxán, G., Forte, E., Furtado, M., et al. (2021). Single cell sequencing  
412 reveals endothelial plasticity with transient mesenchymal activation after myocardial infarction. *Nature  
413 Communications* 12, 681. doi:10.1038/s41467-021-20905-1. PMID: 33514719
- 414 Tucker, N. R., Chaffin, M., Fleming, S. J., Hall, A. W., Parsons, V. A., Bedi, K. C., et al. (2020).  
415 Transcriptional and cellular diversity of the human heart. *Circulation* 142, 466–482. doi:10.1161/  
416 CIRCULATIONAHA.119.045401. PMID: 32403949
- 417 Turkieh, A., Porouchani, S., Beseme, O., Chwastytiak, M., Amouyel, P., Lamblin, N., et al. (2019).  
418 Increased clusterin levels after myocardial infarction is due to a defect in protein degradation systems  
419 activity. *Cell Death & Disease* 10, 608. doi:10.1038/s41419-019-1857-x. PMID: 31406108
- 420 Vafadarnejad, E., Rizzo, G., Krampert, L., Arampatzi, P., Arias-Loza, A.-P., Nazzal, Y., et al. (2020).  
421 Dynamics of cardiac neutrophil diversity in murine myocardial infarction. *Circulation Research* 127,  
422 e232–e249. doi:10.1161/CIRCRESAHA.120.317200. PMID: 32811295
- 423 Van den Berge, K. and Clement, L. (2021). *stageR: stage-wise analysis of high throughput gene expression  
424 data in R*. R package version 1.16.0
- 425 van Duijvenboden, K., de Bakker, D. E., Man, J. C., Janssen, R., Günthel, M., Hill, M. C., et al.  
426 (2019). Conserved  $\beta$ -nppb $\beta$ -border zone switches from mef2- to ap-1-driven gene program.  
427 *Circulation* 140, 864–879. doi:10.1161/CIRCULATIONAHA.118.038944. PMID: 31259610
- 428 Verdonschot, J. A. J., Robinson, E. L., James, K. N., Mohamed, M. W., Claes, G. R. F., Casas, K.,  
429 et al. (2020). Mutations in pdlim5 are rare in dilated cardiomyopathy but are emerging as potential  
430 disease modifiers. *Molecular genetics & genomic medicine* 8, e1049–e1049. doi:10.1002/mgg3.1049.  
431 31880413[pmid]
- 432 Volden, R. and Vollmers, C. (2021). Highly multiplexed single-cell full-length cdna sequencing of human  
433 immune cells with 10x genomics and r2c2. *bioRxiv* doi:10.1101/2020.01.10.902361
- 434 Wang, L., Yu, P., Zhou, B., Song, J., Li, Z., Zhang, M., et al. (2020). Single-cell reconstruction of the adult  
435 human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function.  
436 *Nature Cell Biology* 22, 108–119. doi:10.1038/s41556-019-0446-7. PMID: 31915373
- 437 Wang, Q., Boenigk, S., Boehm, V., Gehring, N. H., Altmueller, J., and Dieterich, C. (2021). Single cell  
438 transcriptome sequencing on the nanopore platform with scsnapbar. *RNA (New York, N.Y.)* 27, 763–770.  
439 doi:10.1261/rna.078154.120. PMID: 33906975
- 440 Weeland, C. J., van den Hoogenhof, M. M., Beqqali, A., and Creemers, E. E. (2015). Insights into  
441 alternative splicing of sarcomeric genes in the heart. *Journal of Molecular and Cellular Cardiology* 81,  
442 107–113. doi:https://doi.org/10.1016/j.yjmcc.2015.02.008. PMID: 25683494
- 443 Wolf, F. A., Angerer, P., and Theis, F. J. (2018). Scanpy: large-scale single-cell gene expression data  
444 analysis. *Genome Biology* 19, 15. doi:10.1186/s13059-017-1382-0. PMID: 29409532
- 445 Wu, C.-C., Kruse, F., Vasudevarao, M. D., Junker, J. P., Zebrowski, D. C., Fischer, K., et al. (2016).  
446 Spatially resolved genome-wide transcriptional profiling identifies bmp signaling as essential regulator  
447 of zebrafish cardiomyocyte regeneration. *Developmental Cell* 36, 36–49. doi:https://doi.org/10.1016/j.  
448 devcel.2015.12.010. PMID: 26748692

- 449 Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M. I., and Yosef, N. (2021). Probabilistic harmonization  
 450 and annotation of single-cell transcriptomics data with deep generative models. *Molecular Systems  
 451 Biology* 17, e9620. doi:<https://doi.org/10.15252/msb.20209620>. PMID: 33491336
- 452 Zhang, D., Hu, Q., Liu, X., Ji, Y., Chao, H.-P., Liu, Y., et al. (2020). Intron retention is a hallmark and  
 453 spliceosome represents a therapeutic vulnerability in aggressive prostate cancer. *Nature Communications*  
 454 11, 2089. doi:[10.1038/s41467-020-15815-7](https://doi.org/10.1038/s41467-020-15815-7). PMID: 32350277
- 455 Zinad, H. S., Natasya, I., and Werner, A. (2017). Natural antisense transcripts at the interface between host  
 456 genome and mobile genetic elements. *Frontiers in Microbiology* 8. doi:[10.3389/fmicb.2017.02292](https://doi.org/10.3389/fmicb.2017.02292)

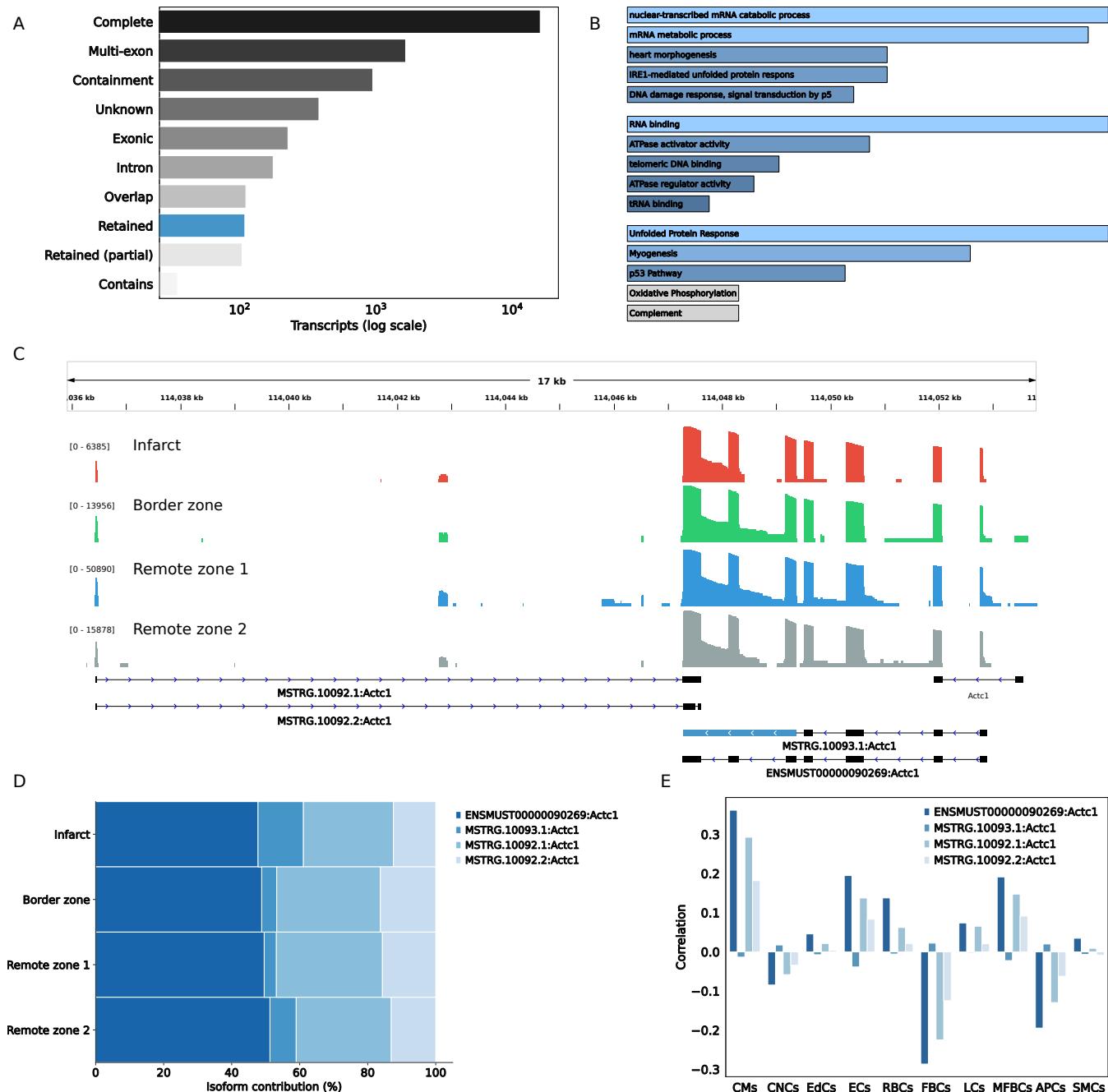
## FIGURE CAPTIONS



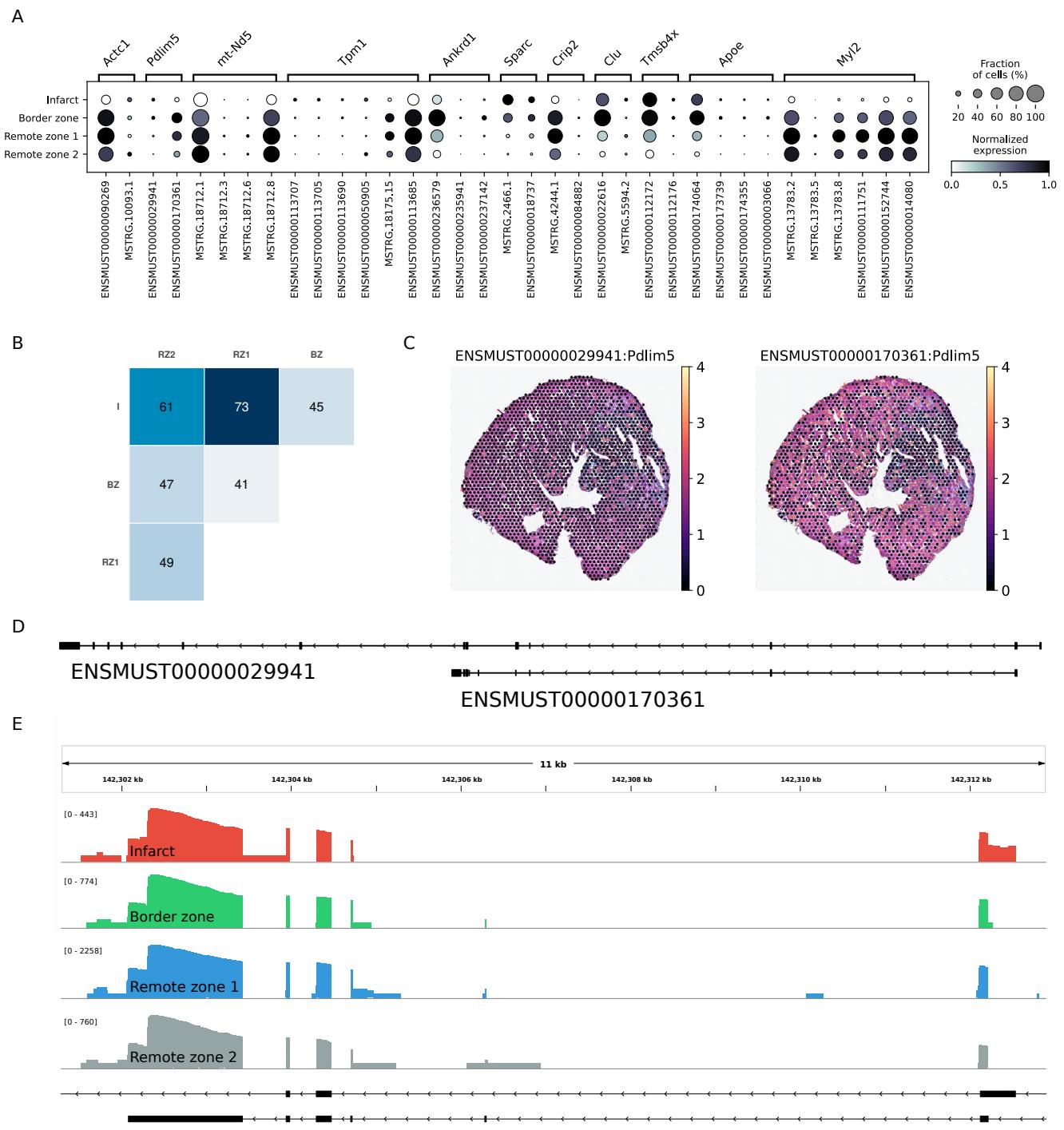
**Figure 1.** scNAST methodology. A Schematic of the scNAST workflow using a hybrid sequencing approach on Nanopore and Illumina platforms to assign spatial barcodes to long-read sequencing. B Nanopore sequencing saturation showing the number of splice sites detected at various levels of subsampling. 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 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, based on the short-read Illumina data. B Annotation of mouse heart regions after MI via short-read clustering, transferred to the Nanopore data. Scatter plot in spatial coordinates of the anatomical regions (left) and UMAP representation of the Nanopore data using the region annotation from short-read clustering (right). Colors in the spatial scatter plot are matching those of the UMAP. C Neighbors enrichment analysis in one heart axis section. The heatmap shows the enrichment score on spatial proximity between the different anatomical regions. Spots belonging to two different regions that are close together will have a high score, and vice-versa. D Cluster co-occurrence in spatial dimensions in one heart axis section. Line plot showing the the conditional probability of observing a given region conditioned on the presence of the infarct region, computed across increasing radii size around each spots. Distance units are given in pixels of the Visium source image. 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-Whitney U-test (\*\* = <0.001, \*\*\*\* = <0.0001)



**Figure 3.** Characterizing the isoform diversity after MI. A 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>. B Over-representation analysis of genes harboring novel transcripts with intron retention (IR). From top to bottom, biological processes, molecular function, and hallmark gene sets from the Molecular Signatures Database (MSigDB). C Exonic structure of the different *Actc1* isoforms, including novel isoforms identified by scNAST with intron retention and exonic antisense overlap. Coverage (log scale) is shown for each region with a different scale. D *Actc1* isoform contributions to total *Actc1* expression in the different heart regions. E Per spot correlation observed between spatial deconvolution of cell types and *Actc1* isoforms.



**Figure 4.** Regional isoform switching after MI. A Dot plot showing the top 5 most significant isoform-switching genes per contrast across all comparisons between the remote, border, and infarct areas, using a stage-wise testing procedure at an overall false discovery rate (OFDR) of 0.05. The top genes were restricted to the spatially variable genes. B Heatmap representing the number of isoform-switching genes identified between any two regions. C *Pdlim5* isoform expression in one heart axis section. D *Pdlim5* isoforms track. E Zoom of D to show coverage of both isoforms across regions.