



OPEN

DATA DESCRIPTOR

The alternative splicing landscape of infarcted mouse heart identifies isoform level therapeutic targets

Binbin Xia^{1,2,7}, Jianghua Shen^{2,3,4,5,7}, Hao Zhang^{2,3,4,5}, Siqi Chen^{2,3,4,5}, Xuan Zhang^{1,6}, Moshi Song^{2,3,4,5} & Jun Wang^{1,2}

Alternative splicing is an important process that contributes to highly diverse transcripts and protein products, which can affect the development of disease in various organisms. Cardiovascular disease (CVD) represents one of the greatest global threats to humans, particularly acute myocardial infarction (MI) and subsequent ischemic reperfusion (IR) injury, which involve complex transcriptomic changes in heart tissues associated with metabolic reshaping and immunological response. In this study, we used a newly developed ONT full-length transcriptomic approach and performed transcript-resolved differential expression profiling in murine models of MI and IR. We built an analytical pipeline to reliably identify and quantify alternative splicing products (isoforms), expanding on the currently available catalog of isoforms described in mice. The updated alternative splicing landscape included transcripts, genes, and pathways that were differentially regulated during IR and MI. Our study establishes a pipeline to profile highly diverse isoforms using state-of-the-art long-read sequencing, builds a landscape of alternative splicing in the mouse heart during MI and IR.

Background & Summary

Cardiovascular disease (CVD) is among the greatest threats to human health globally, with acute and chronic forms of CVD accounting for the highest global morbidity and mortality among all known diseases¹. In particular, acute myocardial infarction (MI) affects three million people worldwide, with percutaneous coronary intervention considered as an effective strategy to reduce mortality^{2,3}. However, ischemic reperfusion (IR) injury, which can account for 50% of the final infarction size, remains a serious complication that results from a mixture of metabolic disorder, inflammation, oxidative stress, and microvascular obstruction⁴. IR significantly affects the prognosis of MI patients, and thus has been extensively studied from multiple angles, including metabolic and immune responses of heart tissues largely based on transcriptome-centered analysis⁵. The implementation of bulk and single-cell RNA sequencing, and more recently spatial transcriptomics, has enabled the identification of important pathways contributing to IR, providing potential targets for intervention^{6–10}. At present, mainstream RNA-seq has limitations associated with the platform used, in particular in the case of short-read data (usually 150–250 bp paired reads) produced by Illumina platforms.

In eukaryotes, the central dogma states that genes encoded in genomic DNA are transcribed to RNA and translated into proteins. This process has complexities, such as alternative splicing (AS) of introns that leads to multiple transcript isoforms of the same gene, potentially carrying out distinct functions. Previous research has confirmed that multiple AS types occur, including exon skipping, alternative 5'-splice site, alternative 3'-splice site, mutually exclusive exons, and intron retention¹¹. In humans, previous studies have indicated that the sixty thousand genes of the human genome encompass >234,000 isoforms across different tissues¹². Moreover, isoform-level functional differences have been reported for a plethora of genes in the brain, testis, muscle, immune cells, and potentially more¹³. To date, investigations on the variety of isoforms present

¹CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China. ²University of Chinese Academy of Sciences, Beijing, 100049, China. ³State Key Laboratory of Membrane Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, China. ⁴Key Laboratory of Organ Regeneration and Reconstruction, Chinese Academy of Sciences, Beijing, 100101, China. ⁵Beijing Institute for Stem Cell and Regenerative Medicine, Beijing, 100101, China. ⁶Faculty of Biological Science and Technology, Baotou Teacher's College, Baotou, 014030, China. ⁷These authors contributed equally: Binbin Xia, Jianghua Shen. ✉e-mail: songmoshi@ioz.ac.cn; junwang@im.ac.cn

in different tissues are largely limited by the inability of short reads to identify and quantify new isoforms. In contrast, long-read sequencing technologies, such as Oxford Nanopore Technology (ONT) and PacBio, can produce reads in the range of 10kbs, drastically improving the ability to identify new gene isoforms and providing increasing evidence of isoform-level differentiation of gene functions¹⁴. This technological improvement may be particularly important for relevant diseases such as MI and IR.

Indeed, previous studies of transcripts, especially novel transcripts, have focused on the analysis of variable splicing events, such as Skipped exon (SE), Retained Intron (RI), and so on. However, these studies were limited mainly to differential analysis, rendering it difficult to obtain definitive expression products. In this study, we employed a newly developed ONT full-length transcriptome approach and differential expression profiling. We established an analytical pipeline to reliably identify and quantify alternative spliced products (isoforms) with important functional impact on specific diseases. Applying this analytical framework to myocardial ischemia-reperfusion provides a way to integrate the joint use of short-read-length NGS and ONT-based full-length transcriptome sequencing data, providing a paradigm for the use of this data and the study of related diseases.

Methods

Animal care and ischemic reperfusion model. The C57BL/6J mice utilized in this study were procured from the animal center at the Institute of Zoology, Chinese Academy of Sciences (Beijing, China). All experimental procedures involving animals were ethically conducted, having received approval from the Institutional Care and Ethical Committee of the Institute of Zoology, Chinese Academy of Sciences (animal permit number: IOZ20190057). The study adhered to the principles outlined in the Guide for the Care and Use of Laboratory Animals (NIH publication 8th Edition, 2011).

The IR model involved mice being anesthetized with isoflurane, followed by a left thoracotomy to expose the heart, and a transient ligation of the left main descending coronary artery (LCA) for 40 min, succeeded by reperfusion for 72 hr. Conversely, the MI model encompassed the permanent ligation of the left anterior descending coronary artery and ended at 72 hr post-surgery. The Sham group underwent a similar surgical procedure but without LCA ligation. Subsequent to surgical interventions, mice were accommodated in well-ventilated cages, provided with autoclaved diets, and maintained in a specific pathogen-free facility with a 12-hr light cycle. Euthanasia was executed through isoflurane inhalation followed by cervical dislocation.

Library preparation and sequencing. Total RNA was isolated from the samples using Trizol reagents (Invitrogen, 15596018CN), where infarct core and peripheral regions were used in the MI and MISHam groups, and infarct core tissue was used in the IR and IRSham groups. We assessed the RNA integrity of all samples using Agilent 2100 to exclude possible errors caused by their degradation. All samples showed good RNA quality, with RIN (RNA Integrity Number) of 8.5 or higher, with most samples greater than 9 (Supplementary Figure S2). The SQK-PCB109 (Oxford Nanopore Technologies) kit was used to prepare a full-length transcriptome sequencing library. In detail, 2 μ M VN Primer (VNP) and 10 mM dNTPs were added and incubated at 65 °C for 5 min. Then snap cooled the samples and add 5xRT Buffer, RNaseOUT, and 10 μ M Strand-Switching Primer (SSP) at 42 °C for 2 min. Next, 1 μ l of Maxima H Minus Reverse Transcriptase was added, which allowed the full-length transcript to be reversely transcribed into cDNA *in vitro* for further PCR. Sequencing uses PromethION (FLO-PRO002) to generate fast5 files and the sequencing was scheduled to run for 72 hr or stop early when there are less than 10 active pores. Reads with quality score (Q score) < 7 were filtered out to ensure the quality of sequencing. Next-generation sequencing (Illumina)-based RNAseq was applied to the same samples. In total, 16 samples were sequenced for four groups (MI, IR, MI-Sham, and IR-Sham) with four samples in each group.

Basecalling and full-length transcript identification. Fast5 files were basecalled by Guppy (version 5.0.11) to generate fastq files. Pychopper (version 2.5.0) was used to identify full-length sequences from the filtered reads and NanoFilt (version 2.8.0)¹⁵ was used to remove low-quality (Q < 7) and relatively short sequences (L < 500) to ensure the accuracy of novel transcripts. After these filters, the retained reads are called High-confidence high-quality reads (HQ-FL, 50% of all FL data in terms of bp were retained, Supplementary Table S3). The sequence quality and length are visualized using NanoPlot (version 1.38.1)¹⁵. GENCODE M26 (GRCm39) transcript reference and genome reference (primary assembly) with the main annotation were used in sequence alignment and quantification¹².

Novel transcript identification. To achieve full-length-transcript identification of transcribed genes, we established a bioinformatic workflow (Fig. 1A–C) that utilizes full-length transcript sequences from ONT to obtain all isoforms (Fig. 1B); we additionally performed Illumina sequencing for benchmarking and correction of sequencing errors in ONT reads. The discovery of novel transcripts is achieved through alignment and collapse with FLAIR (version 1.5)¹⁶ and SQANTI3 (version 4.2)¹⁷. Firstly, the clean fastq files from the Nanopore RNA-Sequencing were merged and aligned to genomic reference by Minimapp2 (version 2.21)¹⁸ with the parameters “-ax splice -uf -k14-secondary=no” to genomic reference. Secondly, the alignment result is further processed to correct the splice junction via FLAIR using the given annotation files. Then we obtained reliable and high-quality transcripts after collapsing redundant sequences. Finally, we used SQANTI3 to identify the ORF region of transcripts, polyA tails, and other information.

We then performed stringent length filtering (minimum of 500 bp) and aligned the reads to the reference genome sequence (GRCm39) to obtain all junctions. We further filtered the reads using FLAIR to remove sequence redundancy and intrinsic errors in ONT reads, in particular filtering out low-abundance, low-quality regions (Fig. 1C). After annotating the transcripts via SQANTI3, a total of 33,335 high-quality non-redundant transcripts (isoforms) were obtained. The transcripts were mapped to a total of 21,702 non-overlapping

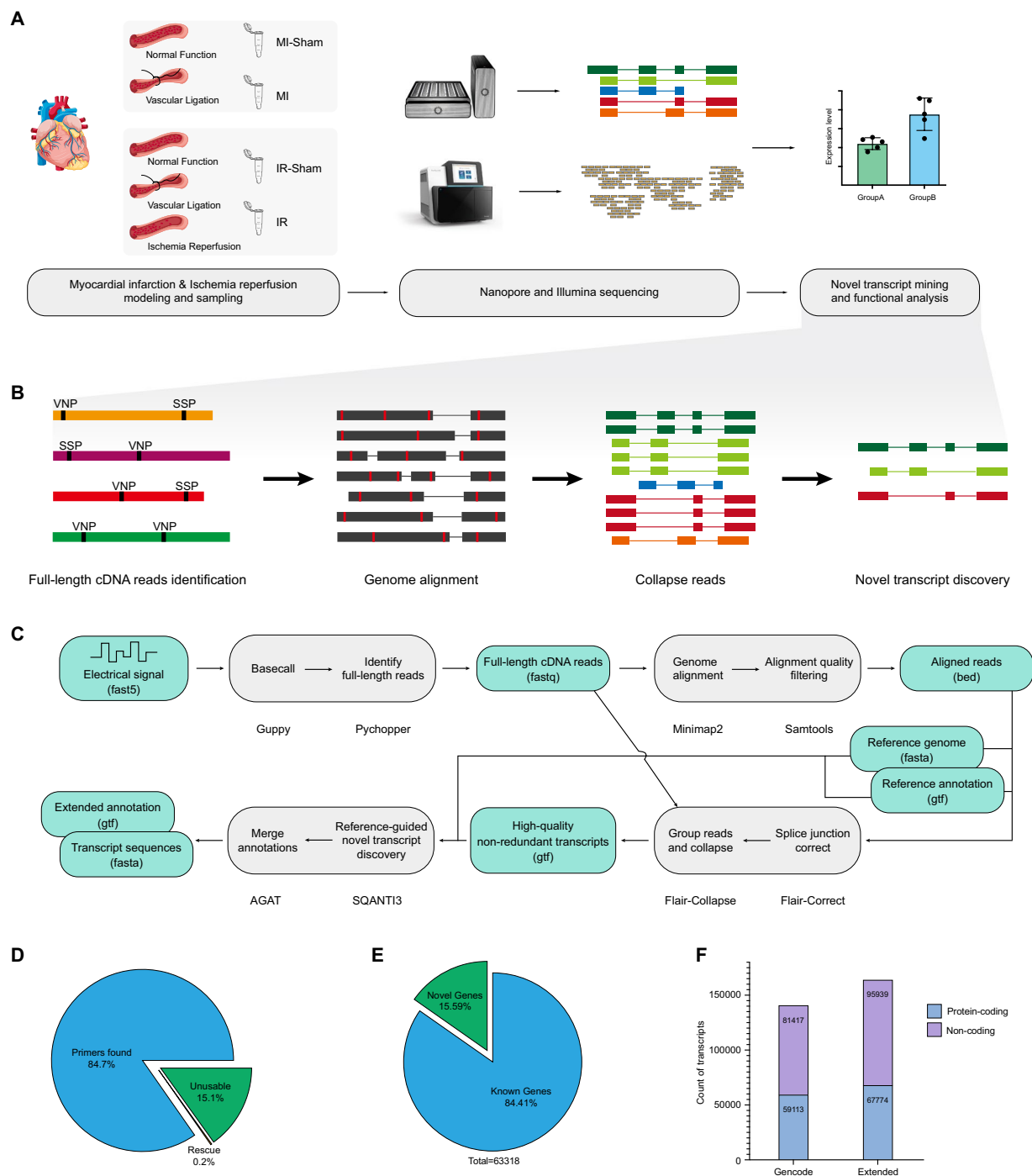


Fig. 1 Overview of our studies and sequencing results. **(A)** The overall framework of our pipeline can be roughly divided into four parts, in which the identification of novel transcripts can be further subdivided in **(B)**, and the operations and software involved in these steps are described in **(C)**. **(D)** Percentage of sequences identified to VPNs and SSPs (Primers found) and chimeric sequences recovered by the software (Rescue) as well as removed sequences (Unusable) in the Oxford nanopore data. **(E)** Percentage of novel genes versus known genes resulting from the identification of novel transcripts. **(F)** Number of extended reference genomes after fusion of novel transcripts with known transcripts in GENCODE.

chromosomal regions, of which 9,874 (45.5%) do not contain annotated exons of genes, but belong to either intergenic regions (8,235, 37.95%), anti-sense regions (1,637, 7.54%), or introns of annotated genes (2, 0.009%), which are hereafter referred to as 'novel genes' (novelGene). Our discovery of novel transcripts expands the number of known transcriptomes in mice by 16.5%, from the original 140,530 transcripts in GENCODE M26 to a total of 163,713 (Fig. 1F), in which 9,874 of the 63,318 genes (15.59%) identified were novelGene (Fig. 1E).

We next focused particularly on the proportion of AS categories containing protein-coding transcripts since proteins function as effector molecules and factors responding to environmental changes. A total of 16,853 (53.35%) transcripts were detected by ONT sequencing encoded proteins (Fig. 2D) and were classified into seven categories: NIC, ISM, Intergenic, Antisense, Genic, Fusion and NNC. The number of novel transcripts coding for proteins identified in each category was significantly higher in the experimental group (MI and IR) compared to the control group (Sham) (Mann-Whitney test $p < 0.05$ for intergroup comparisons across categories of AS, as denoted by an asterisk in Fig. 2L). The formation of mature mRNA in vertebrates involves the cleavage and polyadenylation of precursor mRNA, and the resulting polyadenylation requires the recognition of an A-rich hexamer called the polyA signal (PAS)¹⁹, which tends to be located at an uneven distance of 10–30 bp upstream of the polyA. Among the main similar features of the transcripts was the median length, which was similar across categories (Supplementary Figure S1C). However, there was a significant difference in the PAS-to-polyA distance for NIC and Intergenic compared to known transcripts (FSM) (FSM = -19.89 ± 7.71 bp, NIC = -19.32 ± 8.07 bp, t-test p -value < 0.0001 ; and Intergenic = -19.39 ± 12.92 bp, t-test p value = 0.0041, negative sign represents PAS is located upstream of polyA) (Supplementary Figure S1D).

Merging of original annotation and novel transcript information. The position information of the identified novel transcripts was added to the transcript annotation file acquired from GENCODE M26 by AGAT²⁰ suite (version 0.8.1) to get an updated transcripts annotation for transcript quantification. We used the script “agat_sp_merge_annotations.pl” from the AGAT suite to merge two transcript annotations, and then manually filled in the missing annotation field and removed the redundant transcripts.

Characteristics of alternative splicing in the mouse heart. Furthermore, the alternative splicing landscape of myocardial infarction and myocardial ischemic reperfusion tissue was obtained by annotating non-redundant transcripts. Transcripts that could be mapped to annotated genes were divided into multiple categories using SQANTI3 (Fig. 2A). The largest category was Full Splice Match (FSM), which represents perfect matches to known transcripts (11,097 or 35.13%). This was followed by Intergenic, which refers to transcripts detected in the intergenic region (7,769 or 24.59%). Additionally, Novel in Catalog (NIC) was identified, which contains annotated splice junctions or new combinations formed by annotated exons (5,527, 17.50%). Incomplete Splice Match (ISM), which corresponds to transcripts matching a subsection of known transcripts (4,987 or 15.79%). Antisense, which does not overlap with reference genes but is antisense to annotated genes (1,537, 4.87%). In contrast, 629 (2.00%) genic transcripts were identified with a mixture of annotated introns and exons. Other categories were even rarer, including Fusion, Novel Not in Catalog (NNC), and Genic Intron. It is noteworthy that among the novel transcripts, a large proportion (67.09%) was multi-exon (Fig. 2E).

We then analyzed the AS patterns (composition of different classes of novel isoforms) between the two groups of mice and found that there were significant differences between the MI and the MI-Sham group, as well as between the IR and the IR-Sham group (Asterisks imply Mann-Whitney test $p < 0.05$; Fig. 2K,L). This suggests isoforms can be thought of as disease-specific markers. A significantly higher number of transcripts was found in the MI and IR groups compared to the two Sham groups (Fig. 2J, Mann-Whitney test $p = 0.0286$), including 19,258 unique transcripts in the MI group (of which 2,111 are novel isoforms); and 13,501 unique transcripts in the IR group (of which 1,667 are novel isoforms; Fig. 2G,H).

Gene and transcript quantification. For the alignment of short-read data, we applied Hisat2 (version 2.2.1)²¹ and Samtools (version 1.7)²² was used to convert sam files to bam files. Next, we used featureCounts (version 2.0.1)²³ to quantify gene expression (with parameter “-g gene_id”) and excluded rare transcripts with the sum of the reads below 10. Transcripts (either annotated or novel) were quantified using kallisto (version 0.50.1)²⁴.

Differential expression analysis of genes and transcripts. Taking the gene count matrix as input, differential expression analysis was performed using DESeq 2 (version 1.34.0)²⁵ to calculate the log2 fold change (log2FC) and adjusted p -value of each annotated gene and an adjusted p -value < 0.1 was considered statistically significant. In contrast, transcripts were analyzed using the Swish method (with fishpond version 2.4.1)²⁶, and a p -value < 0.05 was significantly different.

We then quantitatively analyzed differentially expressed genes (DEGs) between groups. Taking advantage of the extended reference, we quantified the transcripts that could be mapped to known exonic regions of genes in each group. A total of 11,378 genes had a significantly different expression in MI ($|FC| > 1.5$, adjusted p -value < 0.1) compared to MI-Sham, including 5,809 up-regulated genes (of which 492 were novelGene), and 5,569 down-regulated genes (of which 781 were novelGene; Fig. 3E). We also found that a total of 8,796 genes (of which 891 were novelGene) had a significantly different expression in the IR compared to IR-Sham ($|FC| > 1.5$, adjusted p -value < 0.1), including 4,402 up-regulated genes (of which 340 were novelGene) and 4,394 down-regulated genes (of which 551 were novelGene; Fig. 3H).

We then performed transcript-level analysis with a special focus on differentially expressed isoforms. We found that, compared to MI-Sham, MI had 1,735 transcripts with significantly down-regulated isoforms (of which 346 were novel transcripts,) and 2,447 transcripts with significantly up-regulated expression (of which 249 were novel transcripts; Fig. 4A). In the case of IR, the corresponding numbers were 696, 155, 835, and 82, respectively (Fig. 4C).

Data Records

Nanopore and NGS sequencing data involved in the experiments have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE275835²⁷.

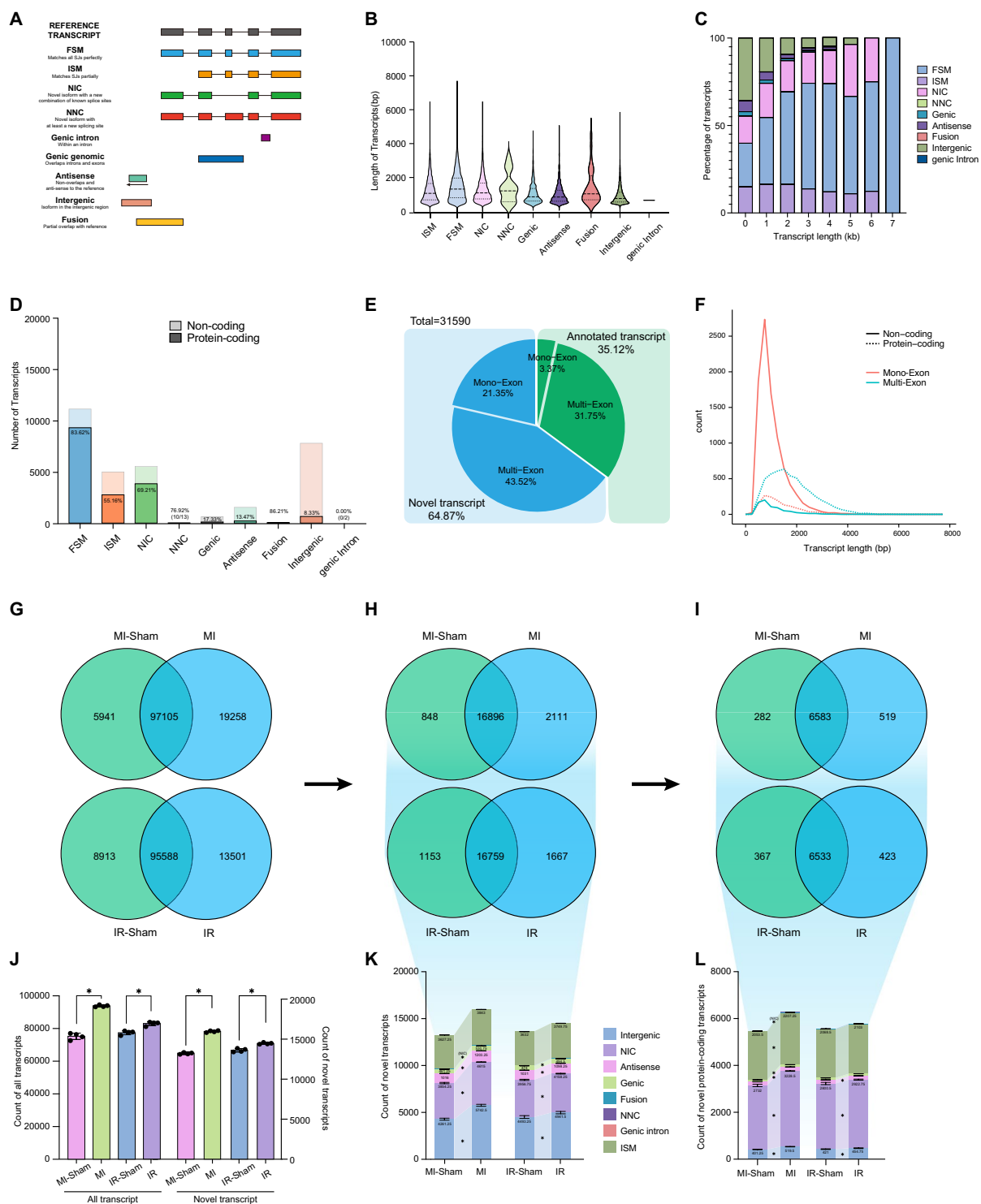


Fig. 2 Identification of novel transcripts and extension of reference annotations. **(A)** Schematic representation of the differences between different types of transcripts. **(B)** Sequence length distribution of different types of transcripts. **(C)** Percentage of each transcript type in different sequence length intervals. **(D)** Percentage of protein-coding and non-coding transcripts in different types. **(E)** Percentage of transcripts with only one exon (Mono-Exon) and multiple-exon transcripts (Multi-Exon). **(F)** Distribution of the number of transcripts of different lengths. The number of all transcripts **(G)**, novel transcripts **(H)**, and protein-coding novel transcripts **(I)** detected in MI/IR and IR-Sham/MI-Sham. **(J)** Number of transcripts detected in different groups of samples (Mann-Whitney test, asterisks indicate $p < 0.05$). **(K)** The number of different categories of novel transcripts. **(L)** The number of different categories of novel protein-coding transcripts (Mann-Whitney test, asterisks indicate $p < 0.05$).

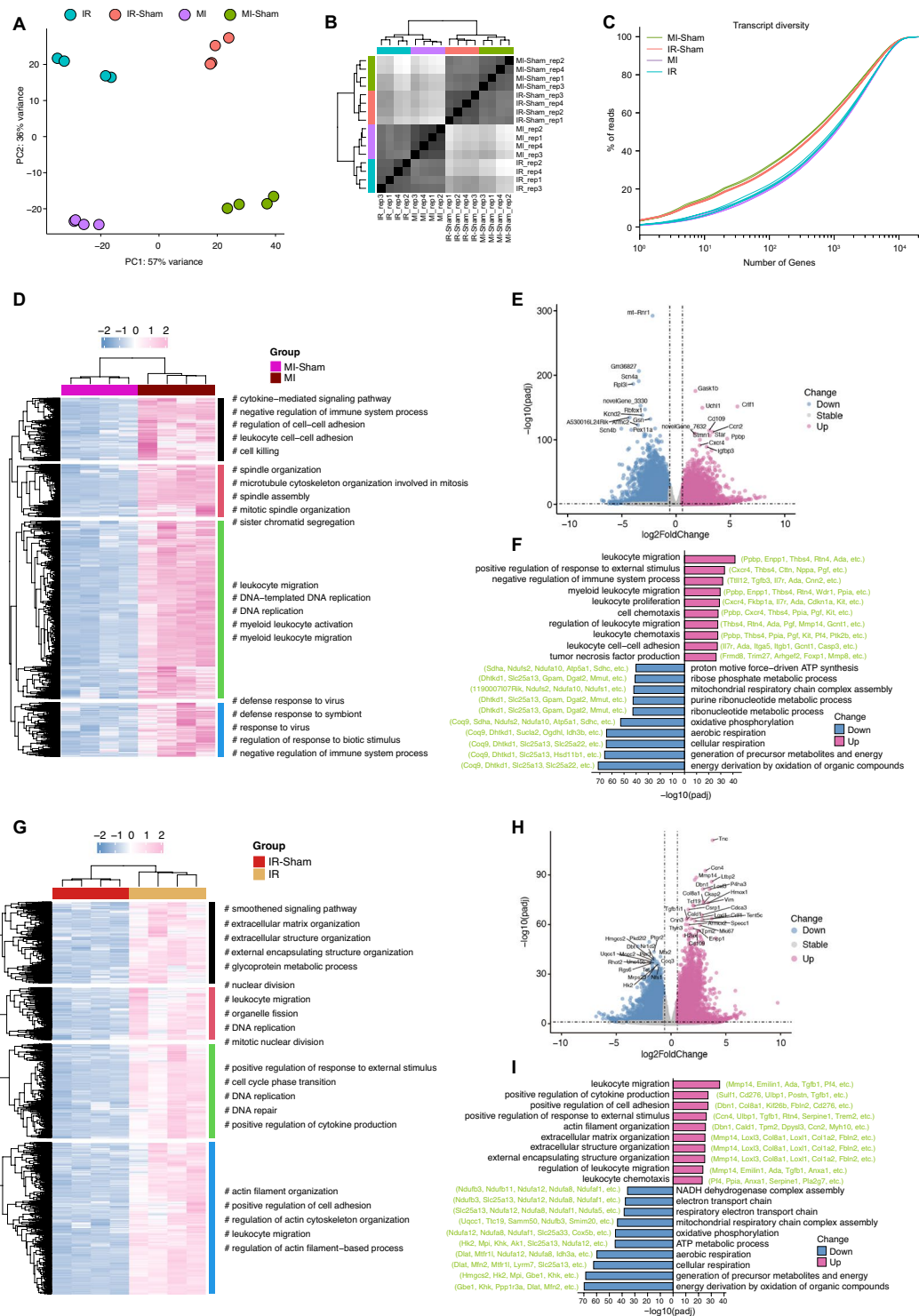


Fig. 3 Gene level expression differences and functional enrichment. **(A)** Principal Component Analysis (PCA) of samples based on top 500 genes. **(B)** Distance matrix of samples. **(C)** The number of genes identified versus NGS reads data, it can be seen that IR/MI can identify the same number of genes with less number of reads relative to two Sham groups. **(D)** Genes up regulated by MI relative to MI-Sham and their GO enrichment to terms. **(E)** Volcano plot of differential genes in the MI relative to IR-Sham. **(F)** Results of GO enrichment analysis of differential genes in MI relative to MI-Sham. **(G)** Genes up regulated by IR relative to IR-Sham and their GO enrichment to terms. **(H)** Volcano plot of differential genes in the IR relative to IR-Sham. **(I)** Results of GO enrichment analysis of differential genes in IR relative to IR-Sham.

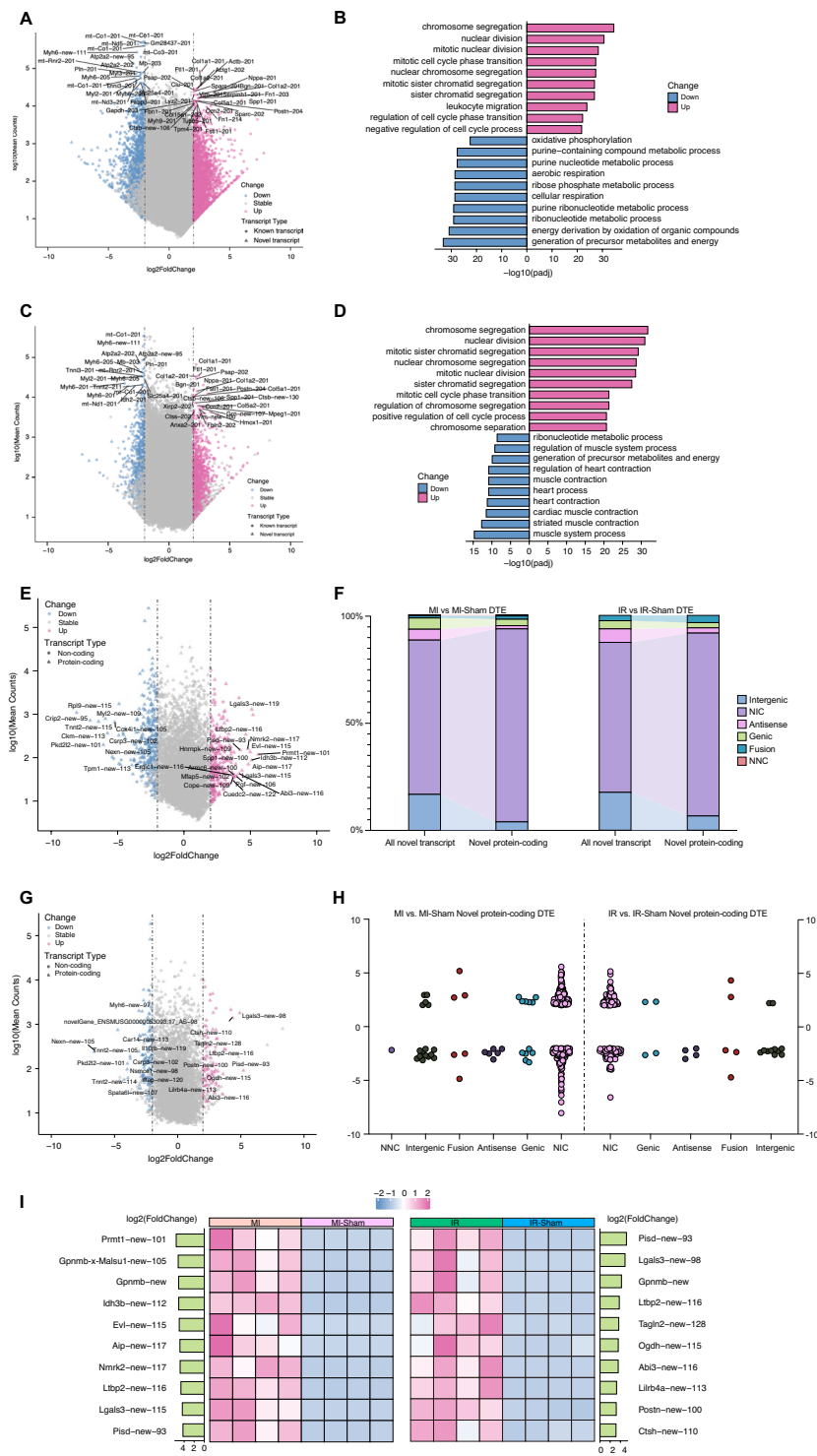


Fig. 4 Differential expression analysis at the transcript level with a focus on MI- and IR-specific transcripts. **(A)** MA plot of differentially expressed transcripts of MI relative to MI-Sham, triangles represent novel transcripts. **(B)** GO enrichment results for genes corresponding to differentially expressed transcripts in MI relative to MI-Sham. **(C)** MA plot of differentially expressed transcripts of IR relative to IR-Sham, triangles represent novel transcripts. **(D)** GO enrichment results for genes corresponding to differentially expressed transcripts in IR relative to IR-Sham. **(E)** MA plot of differentially expressed novel transcripts of MI relative to MI-Sham, triangles represent novel protein-coding transcripts. **(F)** Changes in the proportion of types of differentially expressed novel transcripts in protein-coding transcripts and novel transcripts. **(G)** MA plot of differentially expressed novel transcripts of IR relative to IR-Sham, triangles represent novel protein-coding transcripts. **(H)** Type distribution of differentially expressed protein-coding novel transcripts with fold change. **(I)** IR and MI group upregulation of differentially expressed novel protein-coding transcripts.

Technical Validation

We analyzed samples from mouse models of myocardial infarction (MI) and ischemic reperfusion (IR), in which the experimental group had a permanent left main descending coronary artery (LCA) occlusion to create MI; the ligature was then removed to allow for reperfusion of the ischemic area and promote IR injury. As an experimental control, the sham group (Sham) had the same surgical procedure except that the LCA was not occluded (Fig. 1A). mRNA was extracted from heart tissues from the MI, IR, and Sham groups (n = 4), and full-length mRNA (cDNA) was generated with the SQK-PCB109 kit by Oxford Nanopore Technologies (ONT). This resulted in a total of 23.93 M (million) ONT reads (3.09–9.56 M per sample, on average 5.98 M) (Supplementary Table S1, Supplementary Figure S1A); and 825.19 M Illumina reads (18.87–42.20 M per sample, on average 25.79 M) (Supplementary Table S2).

Relying on front- and back-end primers (VNP, SSP) added during library construction, we filtered full-length transcripts from the ONT reads (Fig. 1B). In terms of the orientation of the reads, the “+” and “−” strands were roughly 1:1 in ratio, as expected (47.9%:52.1%) (Supplementary Figure S1B). The overall identification rate of full-length reads was 84.9% (Fig. 1D), and the numbers of full-length reads were 9.56 M, 6.68 M, 3.09 M, and 4.61 M for MI, MI-Sham, IR, and IR-Sham group, respectively, with a mean read length of 231.1 bp (N50 = 499) (Supplementary Table S1).

Code availability

The code supporting this study is openly available at GitHub repository (<https://github.com/devxia/FLTIQ>).

Received: 4 June 2024; Accepted: 11 October 2024;

Published online: 18 October 2024

References

- Roth, G. A. *et al.* Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. *J Am Coll Cardiol* **76**, 2982–3021 (2020).
- Mechanic, O. J., Gavin, M. & Grossman, S. A. Acute Myocardial Infarction. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024).
- Stähli, B. E. *et al.* Timing of Complete Revascularization with Multivessel PCI for Myocardial Infarction. *N Engl J Med* **389**, 1368–1379 (2023).
- Barrère-Lemaire, S. *et al.* Mesenchymal stromal cells for improvement of cardiac function following acute myocardial infarction: a matter of timing. *Physiol Rev* **104**, 659–725 (2024).
- Martí-Pàmies, Í. *et al.* Brown Adipose Tissue and BMP3b Decrease Injury in Cardiac Ischemia-Reperfusion. *Circ Res* **133**, 353–365 (2023).
- Xia, N. *et al.* A Unique Population of Regulatory T Cells in Heart Potentiates Cardiac Protection From Myocardial Infarction. *Circulation* **142**, 1956–1973 (2020).
- Wang, N. *et al.* Histone Lactylation Boosts Reparative Gene Activation Post-Myocardial Infarction. *Circ Res* **131**, 893–908 (2022).
- Gladka, M. M. *et al.* Single-Cell Sequencing of the Healthy and Diseased Heart Reveals Cytoskeleton-Associated Protein 4 as a New Modulator of Fibroblasts Activation. *Circulation* **138**, 166–180 (2018).
- Molenaar, B. *et al.* Single-cell transcriptomics following ischemic injury identifies a role for B2M in cardiac repair. *Commun Biol* **4**, 1–15 (2021).
- Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* 1–12, <https://doi.org/10.1038/s41586-022-05060-x> (2022).
- Tardaguila, M. *et al.* SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res* **28**, 396–411 (2018).
- Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res* **49**, D916–D923 (2021).
- Mazin, P. V., Khaitovich, P., Cardoso-Moreira, M. & Kaessmann, H. Alternative splicing during mammalian organ development. *Nat Genet* **53**, 925–934 (2021).
- Byrne, A. *et al.* Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun* **8**, 16027 (2017).
- De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
- Tang, A. D. *et al.* Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun* **11**, 1438 (2020).
- Pardo-Palacios, F. J. *et al.* SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. *bioRxiv* 2023.05.17.541248 <https://doi.org/10.1101/2023.05.17.541248> (2023).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M. & Gautheret, D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* **10**, 1001–1010 (2000).
- Dainat, J., Hereñú, D., LucileSol & pascal-git. NBISweden/AGAT: AGAT-v0.8.1. *Zenodo* <https://doi.org/10.5281/zenodo.5834795> (2022).
- Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
- Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
- Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
- Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq 2. *Genome Biology* **15**, 550 (2014).
- Zhu, A., Srivastava, A., Ibrahim, J. G., Patro, R. & Love, M. I. Nonparametric expression analysis using inferential replicate counts. *Nucleic Acids Res* **47**, e105 (2019).
- NCBI GEO <https://identifiers.org/geo/GSE275835> (2024).

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2020YFA0113400, 2022YFC2303200), the National Natural Science Foundation of China (92368112, 81921006), Beijing Natural Science Foundation (JQ22017), CAS Project for Young Scientists in Basic Research (YSBR-076), the State Key Laboratory of Membrane Biology and Key Laboratory of Organ Regeneration and Reconstruction of the Chinese Academy of Sciences.

Author contributions

The study was conceptualized and managed by Jun Wang and Moshi Song. Binbin Xia developed the bioinformatics pipeline and analyzed the data. Hao Zhang performed the animal model establishment and surgery. Jianghua Shen and Siqi Chen carried out the experiments. Xuan Zhang prepared the library and sequencing. Binbin Xia prepared the initial draft of the manuscript, which was subsequently edited by Jun Wang and Moshi Song.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03998-3>.

Correspondence and requests for materials should be addressed to M.S. or J.W.

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

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



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

© The Author(s) 2024