Tenascin-C drives cardiovascular dysfunction in a mouse model of diabetic cardiomyopathy

Zsuzsanna Arnold1, Christopher Dostal1, Petra Lujza Szabo1, Ibrahim Aykac1, Ana Isabel Antunes Goncalves1, Silva Laura Sousa1, Simge Baydar1, Heidi Budde2, Barnabás Váradi3,4, György L Nadasy5, Mária Szekeres6, Camilla Hage7, Lars H. Lund7, Sarah Constantino8,9, Dietmar Abraham10, Karin Zins10, Seth Hallström1,11, Martin Bilban12, Bence Ágg3,4,13, Nazha Hamdani2, Péter Ferdinandy3,4,13, Francesco Paneni8,9, Attila Kiss1, Bruno K. Podesser1

1Ludwig Boltzmann Institute for Cardiovascular Research at the Center for Biomedical Research and Translational Surgery, Medical University of Vienna, 1090 Vienna, Austria

2Institute of Physiology, Cellular and Translational Physiology Department, Ruhr University Bochum, Bochum, Germany. Molecular and Experimental Cardiology, Ruhr University Bochum, Bochum, Germany. Department of Cardiology, St. Josef-Hospital, Ruhr University Bochum, Bochum, Germany, Department of Physiology, Cardiovascular Research Institute, University Maastricht, The Netherlands. HCEMM-SU Cardiovascular Comorbidities Research Group, Center for Pharmacology and Drug Research & Development, Department of Pharmacology and Pharmacotherapy, Semmelweis University, 1089 Budapest, Hungary

3Cardiometabolic and HUN-REN-SU System Pharmacology Research Group, Department of Pharmacology and Pharmacotherapy, Semmelweis University, Budapest, Hungary

4Center for Pharmacology and Drug Research & Development, Semmelweis University, Budapest, Hungary

5Department of Physiology, Faculty of Medicine, Semmelweis University, Budapest, Hungary

6Department of Morphology and Physiology, Faculty of Health Sciences, Semmelweis University, Budapest, Hungary

7Division of Cardiology, Department of Medicine, Karolinska Institutet, and Department of Cardiology, Karolinska University Hospital, Stockholm, Sweden

8Center for Translational and Experimental Cardiology (CTEC), Zurich University Hospital, University of Zurich, Zurich, Switzerland

9University Heart Center, University Hospital Zurich, University of Zurich, Zurich, Switzerland

10Center for Anatomy and Cell Biology, Medical University of Vienna, Vienna, Austria

11Division of Medicinal Chemistry, Otto Loewi Research Center, Medical University of Graz, Graz, Austria

12Department of Laboratory Medicine & Core Facilities (M.B.), Medical University of Vienna, Vienna, Austria

13Pharmahungary Group, Szeged, Hungary

Corresponding author: Dr. Attila Kiss, Ludwig Boltzmann Institute for Cardiovascular Research at the Center for Biomedical Research and Translational Surgery, Medical University of Vienna, 1090 Wahringer Gürtel 18-20, Vienna, Austria. Email: attila.kiss@meduniwien.ac.at

Supplemental Methods

2.14.2 Read Preprocessing and Alignment

For preprocessing, alignment and differential gene expression analysis of the RNA-sequencing data, a previously developed bioinformatics workflow was adapted as described below [1,2]. Raw reads underwent quality control using FastQC (v0.11.9) and MultiQC (v1.13) [3] and preprocessing using Cutadapt (v4.1) [4], removing poly(A) tails and reads containing ambiguous nucleotides (N's), trimming bases with a Phred quality score below 30 from the 3' ends, and filtering out reads shorter than 19nt. High-quality reads were aligned to the mouse reference genome (GRCm39) with Ensembl annotations using HISAT2 (v2.2.1) with default alignment parameters [5]. Aligned reads were sorted, indexed and converted to binary format using SAMtools (v1.10) [5, 6]. Gene-level quantification was performed with featureCounts (v2.0.1) [7], specifying the appropriate gene annotation files.

2.14.3 Differential Gene Expression Analysis

Differential expression analysis was conducted using the DESeq2 package (v1.34.0) in the R environment (v4.1.0) [8]. Raw count data were imported, and sample metadata were incorporated to define experimental groups: A/J non-diabetic, A/J diabetic, TNC-KO non-diabetic, and TNC-KO diabetic. Genes with low counts across all samples were filtered out using default DESeq2 settings to improve statistical power. The Benjamini-Hochberg method was applied to adjust p-values for multiple testing, controlling the false discovery rate (FDR) [9]. Genes with an adjusted p-value (q-value) less than 0.05 were considered significantly differentially expressed. Volcano plots were generated using decoupler (v1.8.0) to visualize the distribution of differentially expressed genes.

2.14.4 Transcription Factor Activity Inference

Transcription factor (TF) activities were inferred using the decoupler package (v1.8.0) in Python (v3.12). DESeq2 statistical results, specifically the Wald-test for gene expression changes, were used as input for the Univariate Linear Model (ULM) implemented in decoupler (1.8.0). The CollecTRI database was utilized via decoupler to obtain TF-target interaction networks specific to *Mus musculus*. TF activity scores and associated p-values were calculated for each contrast. TFs with significant activity changes were identified based on p-values adjusted for multiple testing. Bar plots were generated using the matplotlib library (v3.4.3) in Python (3.12) to visualize the top TFs with significant activity changes, focusing on those with the highest absolute activity scores.

2.14.5 Pathway Activity Analysis

Pathway activity inference was performed using the PROGENy framework (v1.16.0) to assess the activity of key signaling pathways based on gene expression data. The top 500 responsive genes for *Mus musculus* were obtained from the PROGENy database. The Multivariate Linear Model (MLM) method in decoupler was applied to calculate pathway activity scores and associated p-values for each contrast. Pathways with adjusted p-values less than 0.05 were considered to have significant activity changes. Bar plots were created using matplotlib in Python to display the activity scores of the significant pathways.

2.14.6 Pattern Analysis and Gene Selection

Pattern analysis was conducted to identify genes exhibiting specific expression patterns across different comparisons. Due to the stringent nature of q-value thresholds, which could exclude potentially relevant genes, specific p-value thresholds were applied instead. Genes with p-values less than 0.05 were considered for further analysis, even if they did not meet the FDR-adjusted q-value threshold. This approach ensured that genes showing significant expression changes were not overlooked due to multiple testing corrections.

Comparisons focused on included: A/J diabetic versus A/J non-diabetic; TNC-KO diabetic versus A/J diabetic; TNC-KO non-diabetic versus A/J non-diabetic and TNC-KO diabetic versus A/J non-diabetic. By applying p-value filtering, we captured a broader set of genes exhibiting significant changes, allowing for a more comprehensive downstream analysis.

2.14.7 Construction of Overlapping Protein-Protein Interaction Networks

To investigate the interactions among genes identified through pattern analysis, a second protein-protein interaction (PPI) network was constructed using Cytoscape software (v3.10.1) and the STRING app (v2.0.2). Genes from the TNC-KO diabetic versus A/J diabetic comparison that met the specific p-value threshold (p-value < 0.05) were imported into Cytoscape. The STRING database (v12.0) provided PPI data with a confidence score cutoff of 0.4, including medium-confidence interactions.

This second PPI network allowed us to explore interactions among genes that might have been excluded when using q-value filtering. By doing so, we identified additional genes of interest that, while not passing the FDR threshold, may play significant roles in the biological processes under study.

2.14.8 Overlap Analysis and Hub Gene Identification

An overlap analysis was performed between the hub genes identified in the initial PPI network (based on q-value filtering) and the genes in the second PPI network (based on p-value filtering). This analysis aimed to identify common genes that might be critical in the underlying biological mechanisms.

The CytoHubba plugin (v0.1) in Cytoscape was used again to identify hub genes within the second PPI network. The Maximal Clique Centrality (MCC) algorithm determined the top 10 hub genes, which included genes such as *Cpt1a*, *Slc27a1*, *Ccn2*, and *Corin*. These hub genes,

along with their immediate interacting partners, were visualized to highlight their potential significance.

2.14.9 Enrichment and Overrepresentation Analysis

Enrichment analysis was conducted on the genes from the second PPI network using Metascape 3.5. This analysis included overrepresentation tests for KEGG pathways, WikiPathways, Gene Ontology Biological Processes, Molecular Function (GO BP, MF), and Reactome pathways. By integrating multiple databases, we aimed to identify enriched biological functions and pathways associated with the genes identified through p-value filtering.

Enrichment terms with a p-value less than 0.05 and a minimum gene count of three were considered significant. This analysis provided insights into additional biological processes that may not have been detected using the more stringent q-value filtering.

2.14.10 Visualization of Gene Expression and Pathways

Normalized expression counts for selected genes (*e.g.*, *Cpt1a*, *Slc27a1*, *Ccn2*, *Corin*) were plotted using the ggplot2 package (v3.3.5) in R to visualize expression differences across experimental groups. Additionally, the fold changes from the TNC-KO diabetic versus A/J diabetic comparison were mapped onto the p53 signaling pathway diagram from WikiPathways using the pathview package (v1.32.0) in R. This allowed for the illustration of the impact of TNC-KO on the p53 pathway.

References

1. Sayour NV, Gergely TG, Váradi B, Tóth VÉ, Ágg B, Kovács T, Kucsera D, Kovácsházi C, Brenner GB, Giricz Z: **Comparison of mouse models of heart failure with reduced ejection fraction**. *ESC Heart Failure* 2024.
2. Sayour NV, Tóth VÉ, Nagy RN, Vörös I, Gergely TG, Onódi Z, Nagy N, Bödör C, Váradi B, Ruppert M: **Droplet digital PCR is a novel screening method identifying potential cardiac G-protein-coupled receptors as candidate pharmacological targets in a rat model of pressure-overload-induced cardiac dysfunction**. *International Journal of Molecular Sciences* 2023, **24**(18):13826.

3. Ewels P, Magnusson M, Lundin S, Käller M: **MultiQC: summarize analysis results for multiple tools and samples in a single report**. *Bioinformatics* 2016, **32**(19):3047-3048.
4. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads**. *EMBnet journal* 2011, **17**(1):10-12.
5. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nature methods* 2015, **12**(4):357-360.
6. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM: **Twelve years of SAMtools and BCFtools**. *Gigascience* 2021, **10**(2):giab008.
7. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.
8. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome biology* 2014, **15**:1-21.
9. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *Journal of the Royal statistical society: series B (Methodological)* 1995, **57**(1):289-300.