BMC Biology

Full-length transcriptomic analysis in murine and human heart reveals diversity of PGC-1α promoters and isoforms regulated distinctly in myocardial ischemia and obesity

Daniel Oehler[1,2*] , André Spychala[3], Axel Gödecke[2,3], Alexander Lang[1,2], Norbert Gerdes[1,2], Jorge Ruas[4], Malte Kelm[1,2], Julia Szendroedi[5,6] and Ralf Westenfeld[1,2*]

Abstract

Background: Peroxisome proliferator-activated receptor gamma coactivator-1 alpha (PGC-1α) acts as a transcriptional coactivator and regulates mitochondrial function. Various isoforms are generated by alternative splicing and differentially regulated promoters. In the heart, total PGC-1α deficiency knockout leads to dilatative cardiomyopathy, but knowledge on the complexity of cardiac isoform expression of PGC-1α remains sparse. Thus, this study aims to generate a reliable dataset on cardiac isoform expression pattern by long-read mRNA sequencing, followed by investigation of differential regulation of PGC-1α isoforms under metabolic and ischemic stress, using high-fat-high-sucrose-diet-induced obesity and a murine model of myocardial infarction.

Results: Murine (C57Bl/6J) or human heart tissue (obtained during LVAD-surgery) was used for long-read mRNA sequencing, resulting in full-length transcriptomes including 58,000 mRNA isoforms with 99% sequence accuracy. Automatic bioinformatic analysis as well as manual similarity search against exonic sequences leads to identification of putative coding PGC-1α isoforms, validated by PCR and Sanger sequencing. Thereby, 12 novel transcripts generated by hitherto unknown splicing events were detected. In addition, we postulate a novel promoter with homologous and strongly conserved sequence in human heart. High-fat diet as well as ischemia/reperfusion (I/R) injury transiently reduced cardiac expression of PGC-1α isoforms, with the most pronounced effect in the infarcted area. Recovery of PGC-1α-isoform expression was even more decelerated when I/R was performed in diet-induced obese mice.

Conclusions: We deciphered for the first time a complete full-length transcriptome of the murine and human heart, identifying novel putative PGC-1α coding transcripts including a novel promoter. These transcripts are differentially regulated in I/R and obesity suggesting transcriptional regulation and alternative splicing that may modulate PGC-1α function in the injured and metabolically challenged heart.

Keywords: Ischemia/reperfusion, Long-read sequencing, PGC-1α, Diet-induced obesity

Background

The ubiquitously expressed Peroxisome proliferator-activated receptor gamma coactivator-1 alpha (PGC-1α) regulates mitochondrial function, hypoxia-induced angiogenesis and antioxidative capacity. It functions as a transcriptional coactivator, and its transcripts are assembled

*Correspondence: Daniel.Oehler@med.uni-duesseldorf.de; Ralf. Westenfeld@med.uni-duesseldorf.de

[1] Division of Cardiology, Pulmonology, and Vascular Medicine, Medical Faculty, Heinrich-Heine University, Moorenstr. 5, 40225 Düsseldorf, Germany
Full list of author information is available at the end of the article

Oehler *et al. BMC Biology*   (2022) 20:169

Page 2 of 20

by alternative splicing events and are regulated by different promoters [1–3]. Its biological roles range mitochondrial biogenesis and mitochondrial dynamics [4] through induction of hypertrophy and atrophy in muscle cells as well as controlling inflammation in apoptosis environments [5]. PGC-1α is studied in detail in liver, skeletal muscle, neuronal cells (neurodegenerative disorders) [6, 7] and cancer [8].

In skeletal muscle in mice, general activation of PGC-1α leads to protection from sarcopenia [9]. Additionally, PGC-1α activation in muscle promotes angiogenesis, oxygen consumption and energy supply [10–12], shown for isoforms PGC-1α1, PGC-1α-b, PGC-1α4 and NT-PGC-1α. Moreover, it leads to an anti-inflammatory environment [13], while loss of whole PGC-1α in muscle potentiates a systemic inflammatory response [14]. The most intensively studied isoform, PGC-1α1 (also known as PGC-1α-a), leads in cultured muscle cells and in skeletal muscle in vivo to an increased expression of several angiogenic factors (including VEGF), and consecutively to accelerated recovery after ischemia, which is severely impaired in PGC-1α-deficient mice [15, 16]. Interestingly, the short isoform PGC-1α4 induces muscle hypertrophy, most likely via a negative feedback mechanism to myostatin expression [1]. Thus it can be concluded that in skeletal muscle, PGC-1α-1 and PGC-1α4, generated through alternative splicing events, are differentially regulated and are interacting with different downstream targets (PGC-1α4 via IGF1/myostatin axis and PGC-1α-1 via changes of mitochondrial gene expression). In skeletal muscle, this means that depending on the expressed isoform, PGC-1α expression can have opposite effects on muscle mass and energy consumption, which could also be shown to be clinically relevant [1]. Besides alternative splicing, differential promoter usage represents another level of regulation, as for example systemic cold stress leads to increased expression of PGC-1α isoforms controlled by the alternative promoter [17].

The heart is crucially depending on energy supply, mitochondrial oxidative capacity, and mitochondrial biogenesis. The PGC-1 family of transcription factors is involved in cardiac metabolism as a main driver of mitochondrial biogenesis and induces scavengers of ROS-species [18, 19]. Additionally, it is involved in indirect transcriptional regulation of the mitochondrial genome, import and utilization of fatty acids and angiogenesis [20]. Moreover, previous work showed a relation between deficiency of total PGC-1α and impaired mitochondrial function [21–23]. Transgenic mice with myocardial inactivation/deletion/deficiency of total PGC-1α develop a dilatative cardiomyopathy phenotype with increased end systolic volume and reduced contractile function as well as metabolic alterations associated with heart failure [24], while overexpression of PGC-1α-a appears to enhance contractility without negative feedback on cardiac metabolism [25].

The crosstalk between metabolic dysfunction and heart failure in terms of 'diabetic cardiomyopathy' is still not fully understood [26]. PGC-1α hereby mediates adaptation to caloric restriction [27] and suppresses inflammatory processes mediated by NFκB [28] in mice fed with high-fat diet. Additionally, PGC-1α has a protective effect on mitochondrial function in insulin resistance [29]. Considering that in skeletal muscle alternative splicing and varying promoters generate different isoforms with distinct functions, we aimed to investigate the detailed regulation of PGC-1α expression (and function) also in heart.

As neither in-depth information about the cardiac transcriptome in general nor data on possible differential expression for PGC-1α in particular were available, this study is the first to fill this gap by latest-generation sequencing using long-read full-length transcriptomics. Second, as the functional impact of differential-expressed isoforms PGC-1α in the heart remains elusive, we investigate modified expression patterns under metabolic and cardiovascular challenges in cardiac tissue. Using ischemia/reperfusion (I/R) injury alone or in combination with diet-induced obesity and pre-diabetes, we aim model pathological conditions such as acute myocardial infarction in diabetes-prone patients.

Results

First full-length transcriptome in murine heart reveals over 58,000 unique isoforms

Classical short-read sequencing techniques are not suited to generate full-length transcriptomes with reliable sequence information, which is necessary for detection of diverse splice variants. Thus, high accuracy on sequence level is crucial for downstream experiments. We generated a comprehensive full-length transcriptome map using mRNA long-read sequencing (Fig. 1A). In murine wildtype heart, we identified 58440 unique isoforms with 99% predicted accuracy, originating from 12,789 genes (Fig. 2A–C). Exploring the underlying mechanism of transcript formation, we identified 117,182 known canonical (92.04%), 57 known non-canonical (0.04%), 6306 novel canonical (4.95%) and 3768 novel non-canonical (2.96%) splice junctions.

The majority of all found isoforms ($n = 22,186$) used junctions and corresponding exons matching the annotation ('Full-Splice-Match', FSM). Respectively, 16,579 isoforms used known splice junctions in consecutive order with some parts missing (e.g. last part of a transcript; 'Incomplete Splice Match', ISM). The minority of all isoforms were either novel in catalogue (NIC, $n = 8466$), using known splice junctions, but resulting in different transcripts, or novel not in catalogue (NNC, $n = 9009$),
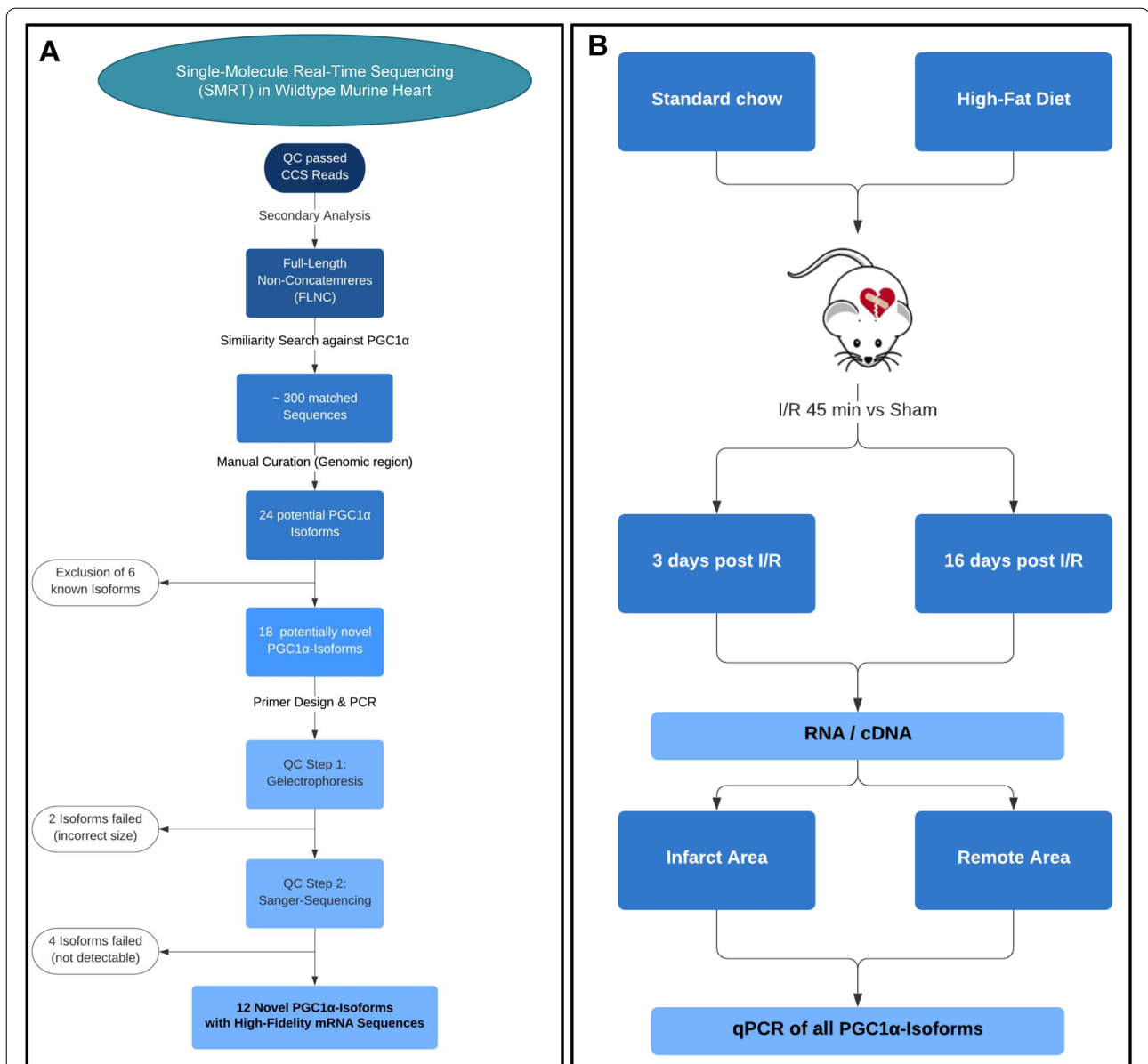
Fig. 1 Workflow overview. **A** Strategy for detection and validation of PGC-1α isoforms by SMRT sequencing in mice. Starting from raw data from SMRT sequencing, primary and secondary analyses were performed, resulting in full-length non-concatemers (FLNC). Then, a similarity search against PGC-1α was performed, yielding in 18 potential novel PGC-1α transcripts, resulting in 12 high-fidelity novel PGC-1α isoforms after quality control. **B** Strategy for investigating into differential expression of PGC-1α isoforms by diet-induced obesity and ischemia/reperfusion (I/R) injury in mice followed by qPCR. Mice fed either a standard chow or high-fat diet underwent I/R or sham surgery. Then, tissue from the infarct area and the remote area (distant from the infarcted area) as baseline as well as 3 and 16 days post I/R was collected and used for expression analysis using qPCR

using new splice junctions and resulting in previously not annotated transcripts. The overall quality of the dataset was excellent (see Additional file 1: Figure S1).

Deciphering novel full-length transcriptome in human heart by long-read sequencing

As knowledge on human data is even more limited, we aimed to unravel the human heart transcriptome by

means of the same approach. Thus, we created a transcriptomic map using mRNA derived from left ventricle from one patient undergoing LVAD-Implantation. In human heart, we identified 66,096 unique isoforms originating from 15,062 genes (Fig. 2D–F). Here, 120,036 known canonical (92.45%), 191 known non-canonical (0.15%), 6989 novel canonical (5.38) and 2616 novel non-canonical (2.01%) splice junctions were detected. The majority of all found isoforms ($n = 24649$) was
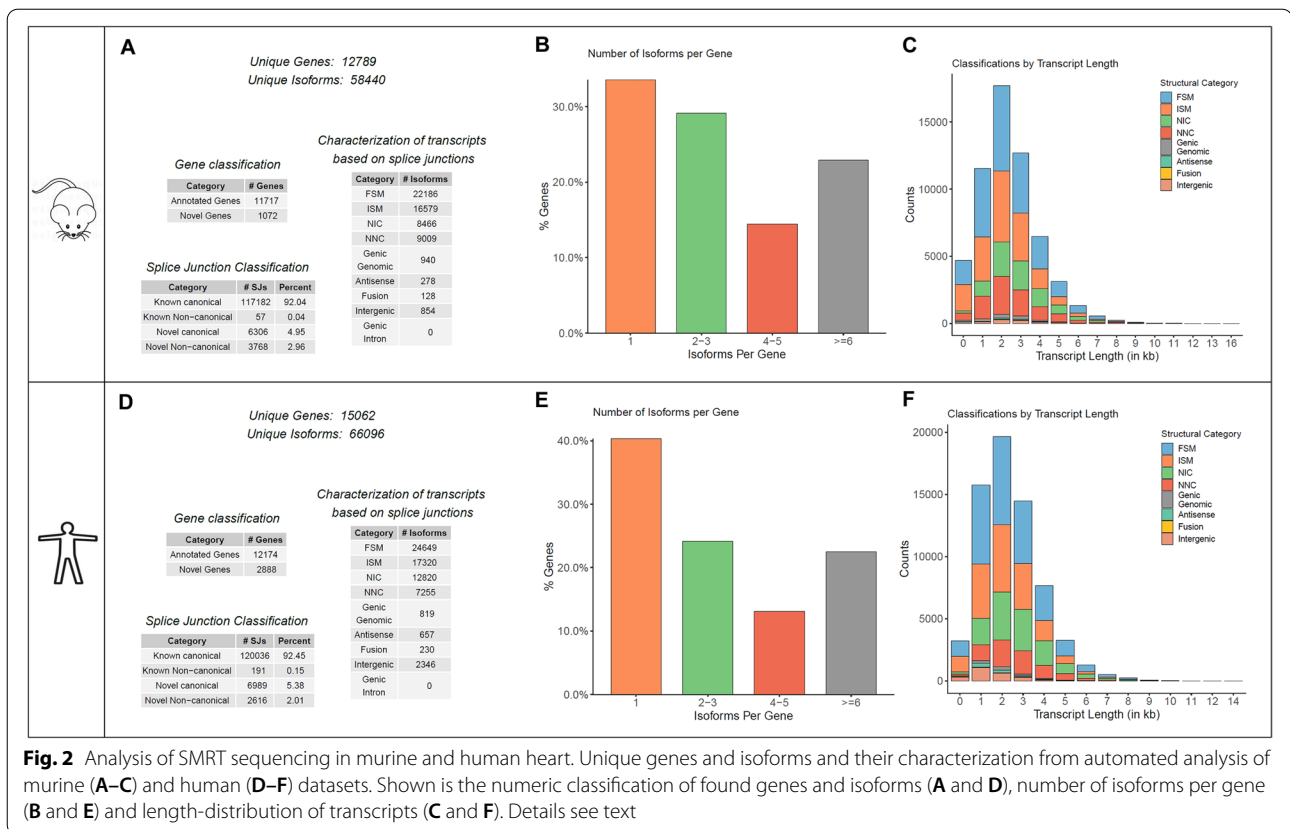
Oehler *et al. BMC Biology* (2022) 20:169

Page 4 of 20

Fig. 2 Analysis of SMRT sequencing in murine and human heart. Unique genes and isoforms and their characterization from automated analysis of murine (**A–C**) and human (**D–F**) datasets. Shown is the numeric classification of found genes and isoforms (**A** and **D**), number of isoforms per gene (**B** and **E**) and length-distribution of transcripts (**C** and **F**). Details see text

categorized as FSM, 17,320 isoforms as ISM. A substantial part of all isoforms was either NIC ($n = 12820$) or NNC ($n = 7255$). The overall quality of the transcriptomic data was also very good and comparable to the murine dataset (Additional file 1: Figure S1).

Detection of 12 novel PGC-1α isoforms within the full-length-transcriptome of the murine heart

Within our novel and complete cardiac transcriptome, we further evaluated expression of PGC-1α as an example for a complex genomic region/structure yielding highly diverse isoforms. These isoforms derive from multiple splice events as well as different promoter sites and regulate diverse biological functions in vivo [1–3]. Two approaches were performed (see also workflow in Fig. 1A): First, we used our high-quality, full-length, clustered transcripts after mapping to reference genome and filtered for those with potential open reading frame (ORF). Due to a potential loss of isoforms during filtering steps of the automated pipeline, we further performed a similarity search against PGC-1α within reads without previous clustering and mapping (full-length, non-concatemers; FLNC). As these reads are non-polished and therefore not error-corrected,

manual curation was necessary. This led not only to the identification of six known but also remarkably 18 potentially novel PGC-1α transcript variants with valid ORF prediction (Fig. 1A). We confirmed cardiac expression of 12 of the predicted novel transcripts using qPCR (Fig. 3 and Additional file 1: Figure S2).

Annotation and prediction of functional domains in known and novel transcripts

After identification of 12 novel and 6 known PGC-1α isoforms, we classified and annotated those also with known functional protein domains. We assigned known functional protein domains to the known isoforms and extrapolated those also on the potential novel transcript variants using motif annotation and prediction tools (detailed search strategy and references see methods) revealing existence of six main protein domains: Two transcription activation domains (AD1, residues 30–40 and AD2, residues 82–95), a protein-recognition motif involved in transcriptional regulatory processes (*LLXXLL*-motif, residues 141–147), the PDB domain 3D24|D involved in binding of the oestrogen-related receptor-alpha (ERRalpha, residues 198–218), a binding domain for interaction with the interaction partner

Oehler *et al. BMC Biology*   (2022) 20:169

Page 5 of 20

Fig. 3 Overview over PGC-1α isoforms in murine heart (passed QC). PGC-1α isoforms (mRNA) after SMRT sequencing which passed QC filtering (Fig. 1): Differing starting exons 1a, b, b' or c (blue boxes), canonical main exons (orange boxes), novel/altered exons (red boxes) and functional domains (green boxes, details see text). Length of boxes indicates relative length of nucleotides (true to scale, with exception of exon 8: shortened bp marked), asterisks and red background layer indicating novel isoforms. **A** Isoforms consisting of either exon 1a, b, b' or the novel exon 1c followed by only canonical main exons. **B** Isoforms starting with exon 1a or b followed by canonical exon 2 to exon 12 and then followed by a novel exon 13b and novel exon 14 (new C-terminal end). **C** Isoforms with starting exon 1a or b and novel exon 6b, ending in the canonical C-terminal end (two isoforms) or with a novel exon 13c (one isoform). **D** N-terminal isoforms (known), with either starting with exon 1a, b or b' and ending preliminary due to an alternative exon 7b. **E** Novel N-terminal isoform, shorter than the known (see D; therefore prefixed with 's'), ending in a novel exon 3b / exon 4b. This isoform is the shortest isoform in the overall pattern. **F** Novel isoform group consisting of different splicing events upstream exon 3, resulting in a shift of the start codon inside exon 3 with valid open reading frame. Three variants exist, differing in the 5'-end (either canonical C-, novel C- or N-terminal end)

PPARγ (residues 292–338) and finally a RNA-Binding-motif, possibly involved in splicing processes of downstream mRNA targets (residues 677–746).

Integrating the structural differences on mRNA sequence level as well as annotation of functional domains led to separation of 6 groups of isoforms:

The first group characterized by a differing starting exon in each isoform is built of 13 exons, encoding all known protein domains, and leading to protein lengths between 793 and 802 amino acids (aa) (Fig. 3A). This group includes the 'reference' long canonical isoform PGC-1α-a as well as two more known (PGC-1α-b and PGC-1α-c) and a novel isoform (named PGC1α-d). The latter contains a novel starting exon which we denominate exon 1c, according to the known starting exons 1a, 1b and 1b').

The second group of two novel isoforms (CT-PGC-1α-a and CT-PGC-1α-b) includes transcripts encoding all known PGC-1α protein domains (810 and 814 aa) and can be classified by a novel C-terminal end, built by alternative splicing events in exon 13 and the former 3'-UTR (Fig. 3B). They consist of either exon 1a or 1b followed by canonical exons 2–12 and end with two novel exons (exon 13b and exon 14, Additional file 1: Figure S3, A).

The third group consisting of three novel isoforms, which contain an alternative (shorter) exon 6b with preserved open reading frame (Fig. 3C and Additional file 1: Figure S3, C). In this group, no known protein

Oehler *et al. BMC Biology* (2022) 20:169

Page 6 of 20

domains are affected. Two isoforms end at the canonical C-terminus (PGC-1α-trE6-a and PGC-1α-trE6-b, 785 and 781 aa) and include a novel exon 13c (see also Additional file 1: Figure S3, B), resulting from a splice event between exon 13 and the 3′-UTR (PGC-1α-trE6-E13c, 783 aa).

The fourth group consists of the known N-terminal isoforms with premature stop within a truncated exon 7 (NT-PGC-1α-a, NT-PGC-1α-b and NT-PGC-1α-c; 257-270aa, see Additional file 1: Figure S3, D), missing the PPARγ-binding domain as well as the RNA recognition motif (Fig. 3D).

A fifth group with a short single isoform (sNT-PGC-1α, 117 aa), built up by exon 1a, exon 2 and two novel exons (Figs. 3B and 4B, see Additional file 1: Figure S3, E), only contains the two activation domains AD1 and AD2 but is missing all other functional domains (Fig. 3E).

Finally, the last group of (novel) isoforms is characterized by splicing events upstream of exon 3, resulting in a shift of the start codon inside exon 3 with a valid open reading frame (Fig. 3F and Additional file 1: Figure S3, F). Here, three variants exist (PGC-1α-E3c, CT-PGC-1α-E3c and NT-PGC-1α-E3c), differing in the 5′-end (either canonical C-, novel C- or N-terminal end; see also Additional file 1: Figure S3, A and D, resp.). While all of them lost the activation domains AD1 and AD2, only the NT-PGC-1α-E3c misses also the PPARγ-binding domain as well as the RNA recognition motif.

Discovery of a novel promoter region of PGC-1α with high conservation in murine and human heart

The novel transcript PGC1a-d in the murine dataset contains a new starting exon (exon 1c; Fig. 3A), yet otherwise exhibits the same sequence as the known PGC-1α-a (Fig. 3A), which is the major and most abundant isoform in our and in all published datasets. Exon 1c originates downstream of the known promoter region for exon 1a (Fig. 4A). This finding suggests, particularly considering the known mechanisms of transcription initiation within this gene locus, identification of a new promoter site. Using promoter prediction tools, we identified a transcription initiator as well as a promoter element preceding the novel exon 1c, both necessary elements for mammalian transcription start sites (Fig. 4B). Remarkably, we found a homologous transcript with the new exon 1c in multiple full-length reads within the human SMRT dataset as well including hints for conservation across species by sharing most of the predicted amino acids (Additional file 1: Figure S4). For analysis of tissue-specific expression of the Exon1c-transcript, we designed primers covering the exon 1c–exon2 junction and

confirmed sequence identity of the amplified PCR fragment by direct sequencing (Additional file 1: Figure S2, F and Additional file 1: Figure S5). This novel Exon1c transcript is expressed in different murine organs, with highest expression in brown adipose tissue and skeletal muscle (Fig. 4C).

Diet-induced obesity (DIO) leads to lowered alternative promoter-driven expression

Gene regulation is accomplished by a variety of mechanisms including differential promoter usage. Thus, we analysed to what extent the different promoters were used for expression of PGC-1α using qPCR primer sets covering specifically each starting exon (Additional file 1: Table S1). Additionally, we were interested in metabolic effects on expression of PGC-1α. Thus, we investigated expression PGC-1α in metabolically dysregulated/challenged mice that were fed a high-fat, high-sucrose diet, leading to diet-induced obesity (DIO) and a pre-diabetic state (Additional file 1: Figure S6).

Notably, for exon 1a (under the control of the canonical promoter) as well as exon 1c (under the control of the novel promoter), comparison of direct expression levels between DIO and lean mice was not showing differences (Fig. 5A). On the other hand, exon 1b and exon 1b′, originating both from the alternative promoter, were significantly lower expressed in DIO compared to lean mice (fold-changes: exon 1b 0.71, exon 1b′ 0.36 with corresponding *p*-values of 0.03 and 0.04, resp.).

PGC-1α expression is reduced in DIO except for novel C-terminal isoforms, CT-PGC-1α-E3c and sNT-PGC-1α

Next, we investigated the influence of high-fat diet on expression levels of the different isoforms. Those isoforms which can be detected solely by covering common junctions (canonical C-terminus, novel C-terminus, shorter N-terminus, truncated exon 6b′, see also Additional file 1: Figure S7) showed significantly lowered expression under DIO with fold-changes in between 0.47 and 0.6 (Fig. 5B), *p*-values between 0.02 and 0.05, resp.), with exception of the novel C-terminal isoforms (*p*-value 0.97). Second, in those isoforms with ability to be detected uniquely by PCR (Fig. 5C), a more heterogenous pattern can be observed: PGC-1α-trE6-Ex13c as well as PGC-1α-E3c are also significantly lower expressed in DIO (fold-changes 0.36 and 0.53, *p*-values 0.006 and 0.02, resp.). CT-PGC-1α-E3c on the other hand is equally expressed under both conditions, and sNT-PGC-1α shows slightly but significantly higher expression in DIO compared to lean mice (fold-change 1.9, *p*-value 0.006).
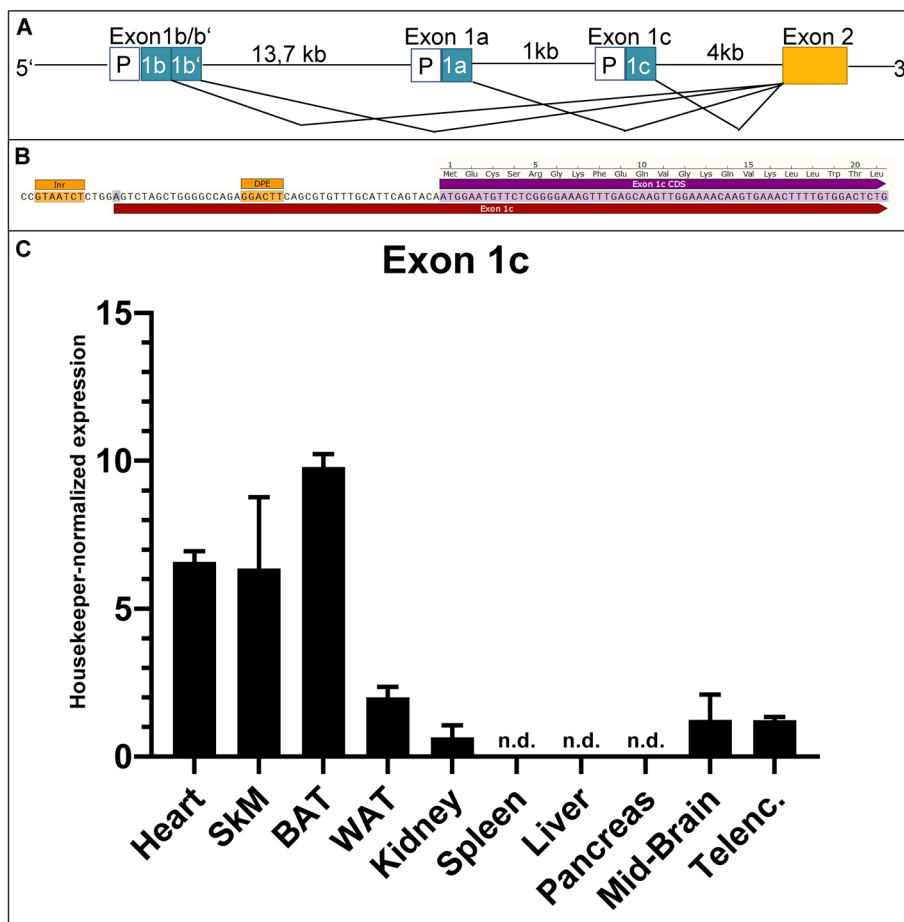
Fig. 4 Organ-specific expression profile of E1c-transcript (exon 1c). **A** Graphical illustration of PGC-1α-promoters and corresponding starting exons. The canonical promoter, responsible for exon 1a, lies between the alternative (known) promoter (regulating exons 1b and b') and the new putative promoter site controlling novel exon 1c. **B** Novel predicted promoter site associated to exon 1c. Prediction (using 'ElemeNT'[57]) of mammalian *in*itiato*r* element (Inr) and corresponding downstream promoter element (DPE) around the transcription start site (marked in grey) of exon 1c. **C** Expression data of DNA samples derived from qPCR with primers targeting new Exon1c-Exon2-Junction in heart, skeletal muscle (SkM), brown adipose tissue (BAT), white adipose tissue (WAT), kidney, spleen, liver, pancreas, mid-brain and telencephalon (Telenc.), normalized to housekeeper (NUDC) and factorized by 1000 for better visualization. Highest expression can be observed in BAT and muscle tissue (heart, SkM); lower, but detectable, expression levels in kidney, WAT and brain. No detection (n.d.) of the new junction could be seen in liver, pancreas and spleen

(See figure on next page.)

Fig. 5 PGC-1α isoform expression in heart in lean or diet-induced obesity (DIO) mice. Housekeeper-normalized expression levels at baseline for either lean mice (black bars) or mice with diet-induced obesity (DIO, grey bars). Shown is isoform expression for PGC-1α starting exons (**A**), for the four group-wise testable isoforms (**B**) and for the four single-detectable isoforms (**C**). **A** Isoform expression by promoter usage. While expression levels for exon 1a and exon 1c are similar between both conditions, transcripts containing exon 1b and exon 1b' are significantly lower expressed in DIO. **B** Expression of group-wise detectable isoforms. While isoforms with novel C-terminal end are unchanged between both diets, the other group-wise detectable isoforms show significantly lower expression in DIO. **C** Expression for isoforms which are single-detectable through PCR. While CT-PGC-1α-E3c is equally expressed under both conditions, the other single-detectable isoforms show significantly reduced expression in DIO. Data acquired by qPCR using cDNA using primers covering specific starting exon 1 (a, b, b' or c) and exon 2 resp. using primers covering specific exon-exon junctions. Expression values are normalized to housekeeper NUDC. $n = 4$ each, bars depict mean values, error bars represent SD. Significance calculated by unpaired Student's *t* test (ns $p > 0.05$, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$)
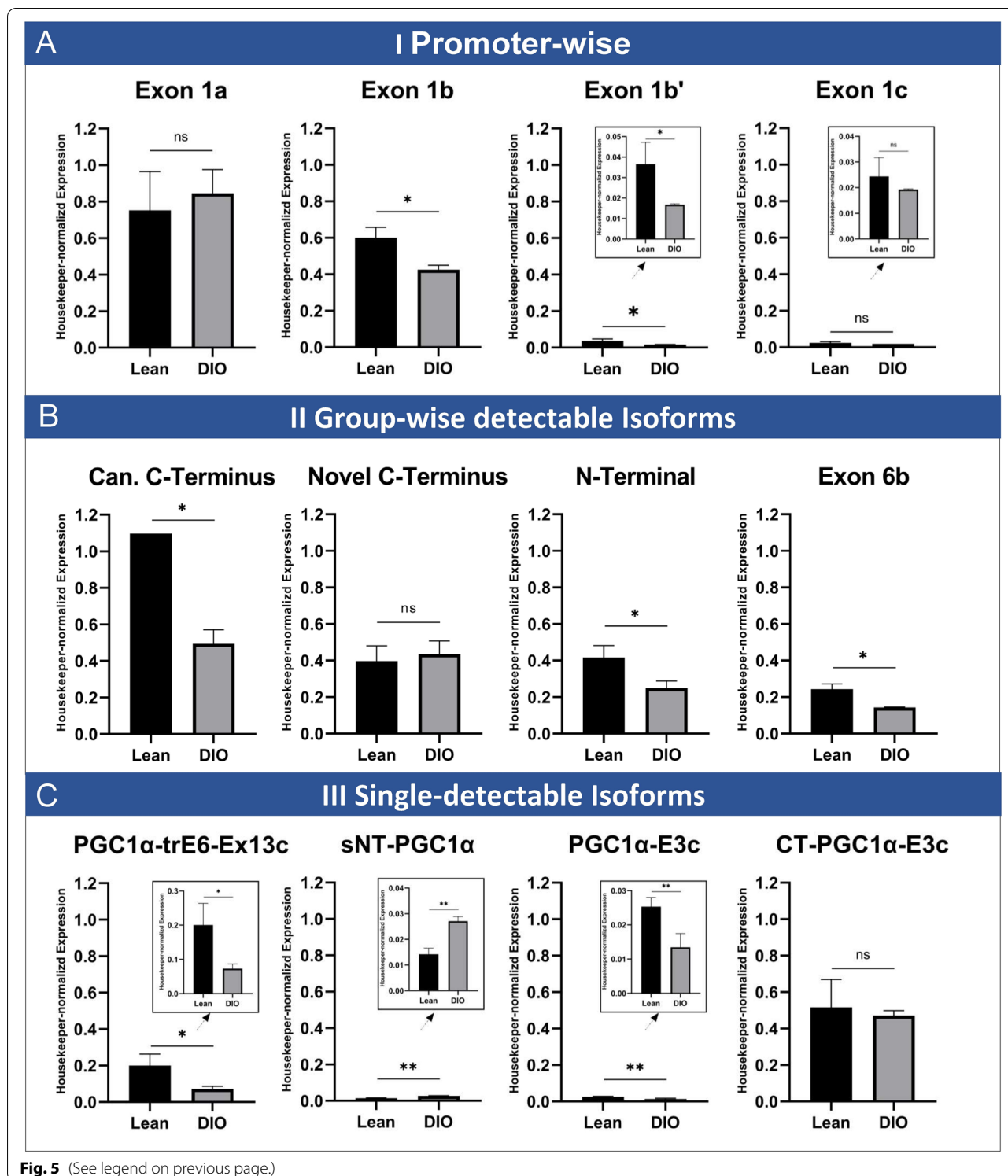
Fig. 5 (See legend on previous page.)

Expression of PGC-1α is reduced 3 days but partially recovers 16 days after I/R in the infarcted area

After providing evidence for DIO-associated differential PGC-1α isoform expression, we aimed to analyse

potential (dys-)functional regulation of the PGC-1α transcriptome in cardiac ischemia/reperfusion (I/R) injury serving as experimental model for myocardial infarction. We used a well-established protocol and

Oehler *et al. BMC Biology*   (2022) 20:169

Page 9 of 20

compared samples at baseline as well as 3 and 16 days post I/R from the infarcted area and remote area to sham-operated controls (workflow see Fig. 1B).

Under basal conditions without intervention, no differences in expression dependent on localization within the heart (corresponding areas to the infarcted and remote myocardium) could be observed (Fig. 6A–D, black bars in upper versus bottom row). I/R injury leads to downregulation of all starting exons 3 days post intervention in the infarcted area compared to sham control (0.32-to 0.38-fold, $*p < 0.05$ for exons 1a, 1b' and c and $***p < 0.001$ for exon 1b). In contrast, in the remote area the expression levels between sham and I/R are equal.

During recovery, 16 days post I/R, almost no difference is observed between I/R and sham (Fig. 6A–D). This is due to a significant increase of expression in the infarcted area from 3 to 16 days post I/R (upper rows; fold-change exon 1a 4.73, exon 1b 5.85, exon 1b' 12.88, exon 1c 5.56; corresponding p-values exon 1a 0.013, exon 1b 0.016, exon 1b' 0.04, exon 1c 0.02). In contrast to that, the expression in the remote area over time stays unchanged (bottom rows).

Additionally, we were also interested if changes in expression also occur in the group-wise or single-detectable isoforms. Here, a similar pattern as for the promoter variants was observed also for most of the splice variants: When compared to sham, I/R induces a transient decline of transcripts in the infarcted area on day 3 post I/R, which was followed by recovered expression on day 16. Expression in the remote myocardium was not altered except for the C-terminal isoforms and PGC-1α-trE6-Ex13c (Fig. 6E–L; fold-changes for the infarcted area ranging from 3.08 to 8.13, p-values from 0.0001 to 0.03).

I/R in DIO leads to heterogenous pattern of PGC-1α expression

As both DIO and I/R individually changed PGC-1α expression in cardiac tissue, we were curious if combining both hits would have an additional effect. Therefore, I/R was repeated in mice fed 10 weeks with high-fat diet, using the same protocol which was used for each modality separately before (see scheme in Fig. 1B). At baseline, the majority of all PGC-1α transcripts exhibit lower expression levels in DIO than lean mice (colour-coded Fig. 7A, Fig. 5 and Additional file 1: Figure S8). Exceptions are CT-PGC-1α-E3c (Additional file 1: Figure S8, horizontally half-filled diamond-shape) and isoforms with the novel C-Terminus (Additional file 1: Figure S8, empty square), which are equally expressed, and transcripts with exon 1b', which are higher expressed in DIO. Three days post I/R (colour-coded Fig. 7B, and Additional file 1: Figure S8), in the infarcted area this changes as, beside low absolute expression values in both groups, all transcripts are either equally or higher expressed in DIO. This effect appears transient, since 16 days post I/R (Additional file 1: Figure S8, C) a similar pattern to baseline can be observed for most transcripts. However, transcripts starting with exon 1a (filled circle) and exon 1b' (empty circle) show higher expression in DIO than lean mice at this time point, while isoforms with the N-terminal end (vertically half-filled square) are even lower expressed than before (see also Fig. 7). Thus, expression distribution at 3 days post I/R seems altered by combination of metabolic and ischemic hit, with incomplete recovery over time in the infarcted area.

The remote area shows also lower expression of all isoforms in DIO than lean mice at baseline (Figs. 7D, 5 and Additional file 1: Figure S8). Three days as well as 16 days post I/R in the remote area, expression in transcripts with exon 1a (Additional file 1: Figure S8, filled circle) and exon 1b' (Additional file 1: Figure S8, horizontally half-filled circle) are significantly higher in DIO than lean mice (Fig. 7E, F; Additional file 1: Figure S8), while all other transcripts resemble baseline expression values.

Discussion

PGC-1α, member of the PGC-1 gene family of transcription factors, is regulated on transcriptional and post transcriptional level. In skeletal muscle,

---

(See figure on next page.)

Fig. 6 PGC-1α isoform expression in heart at baseline and 3- and 16 days post I/R**.** Expression levels at baseline (black bars) and 3 or 16 days post sham surgery (sham, white bars) or ischemia / reperfusion injury (I/R, striped bars) for different PGC-1α starting exons (**A–D**), for the four group-wise testable isoforms (**E–H**) and for the four single-detectable isoforms (**I–L**). Housekeeper-normalized expression values in infarct area (upper row) and remote area (bottom row). **A** Exon 1a, originating from the canonical promoter. **B, C** Exons 1b and 1b', under control of the alternative (known) promoter. **D** Novel exon 1c, originating from a new promoter site. **E** Isoforms with canonical C-terminal ending. **F** Isoforms with novel C-terminal ending. **G** N-terminal isoforms. **H** Isoforms with the novel exon 6b. **I** PGC-1α-trE6-Ex13c. **J** sNT-PGC-1α. **K** PGC-1α-E3c. **L** CT-PGC-1α-E3c. I/R leads to downregulation of canonical and novel PGC-1α starting exons as well as most of the single- and group-wise-detectable PGC-1α isoforms 3 days post I/R, suggesting a general downregulation of PGC-1α in infarct area 3 days post I/R. At 16 days post I/R, a recovery can be observed. Moreover, PGC-1α-trE6-Ex13c is showing 'overcompensative' behaviour in infarcted and remote area. In all other isoforms, the expression in the remote area is not affected, neither at day 3 nor 16 days post infarction. Detailed description, see text. Data acquired by qPCR using cDNA using primers covering specific starting exon 1 (a, b, b' or c) and exon 2 resp. using primers covering specific exon-exon junctions. Expression values are normalized to housekeeper NUDC. $n = 4$ each, bars depict mean values, error bars represent SD. Significance calculated by unpaired Student's t test (ns $p > 0.05$, $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$)
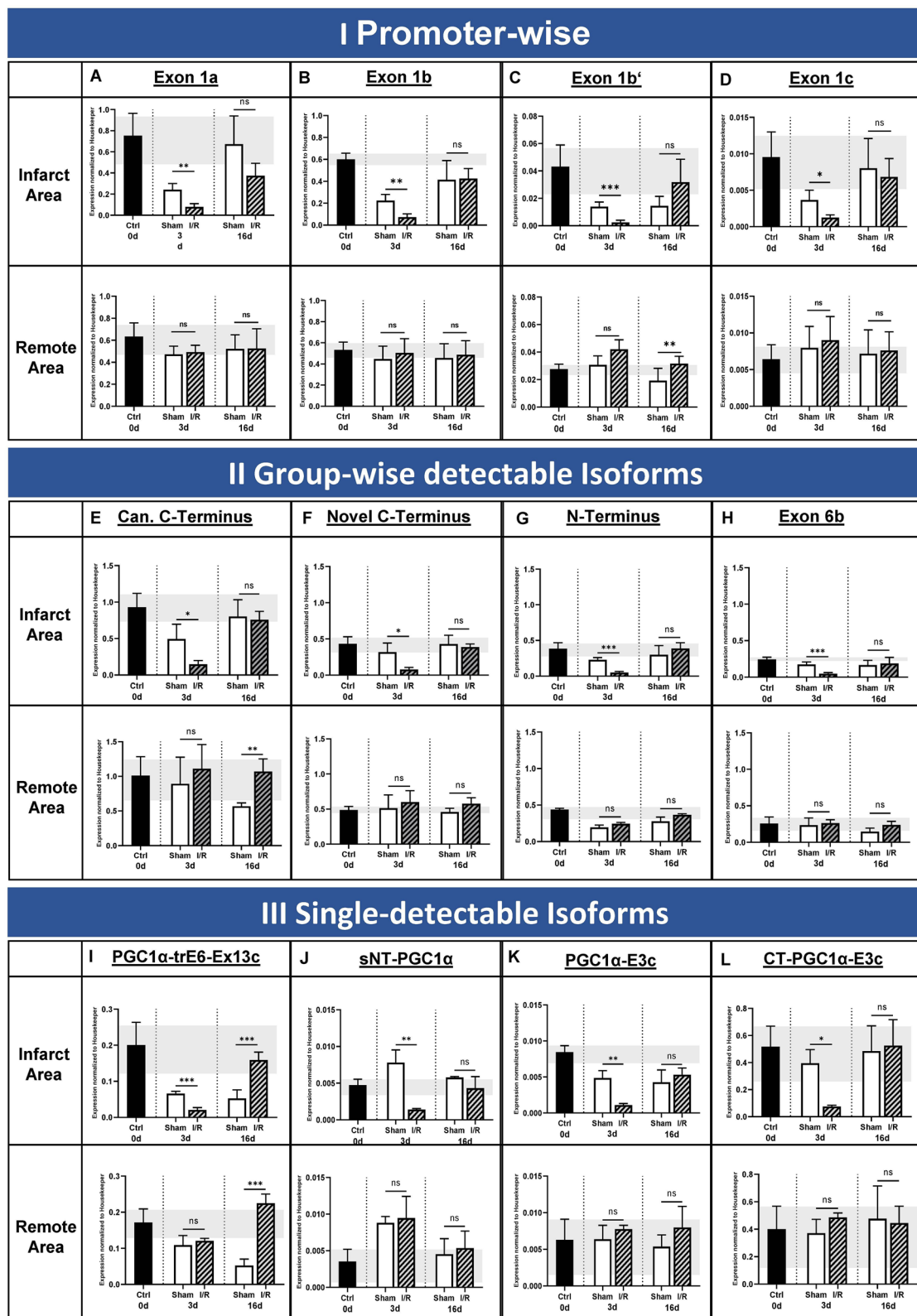
Fig. 6 (See legend on previous page.)

Oehler *et al. BMC Biology*   (2022) 20:169

Page 11 of 20

alternative splicing and different promoters generate various PGC-1α isoforms with distinct functions. Mice with cardiomyocyte-specific deficiency of total PGC-1α develop dilatative cardiomyopathy [24]; vice versa overexpression of PGC-1α1 enhances contractility [25]. Though playing a major role in the heart orchestrating mitochondrial biogenesis [4] and function, little in-depth information about the cardiac transcriptome or possible differential expression PGC-1α is available.

This study aimed to fill this gap by combining latest-generation sequencing with metabolic and ischemic challenges to investigate PGC-1α isoform expression in cardiac tissue.

Here, we report three major findings:

1. For the first time, we generated a reliable, full-length dataset on cardiac transcript expression.
2. We analysed expression pattern of PGC-1α isoforms and identify novel and known PGC-1α transcripts being differentially expressed under high-fat diet, a model for diabetic cardiomyopathy.
3. In murine I/R injury, we demonstrate that promoter-driven PGC-1α transcripts are downregulated and partially recover after 16 days post I/R in the infarcted area.

Deciphering full-length, high-fidelity transcriptome with discovery of 58,000 unique isoforms in murine and human heart

In order to address the complexity of splice variants with high-quality long-read sequencing, we performed full-length, high-fidelity transcriptomics of both murine and human heart for the first time. More than 58,000 unique, full-length sequences with very good quality were achieved. To our knowledge, full-length transcriptomic data are not yet available for most organisms, especially

not for mammals. One earlier study reported even lower number of genes and transcripts in human and mouse brain, but overall comparable quality [30]. Previously, mainly short-read RNA-sequencing was used for transcriptomic analysis. However, as this method sequences fragments of RNA, it depends on reference annotation and is therefore good in detecting alterations on a gene or exon level but lacks depth in discovery of novel isoforms, especially when alternative splicing occurs. Using the SMRT technique, we add important information to the analysis of the cardiac transcriptome as this method avoids assembly of transcripts from classical short-read sequencing that can lead to wrong assumptions on sequence identity. Recently, long-read transcriptomic sequencing has been used in induced pluripotent stem cell-derived cardiomyocytes in vitro by Zhu et al., showing the potential of this method by investigating the impact of gene mutations on isoform expression patterns in a human DCM phenotype [31]. Our dataset, however, gives the possibility to investigate an unbiased whole heart long-read cardiac transcriptome for the first time, which is necessary for a reliable background of functional categorization of novel and known isoforms. Additionally, apart from PGC-1α, our dataset can be used as base for investigation in other possible target genes and biological pathways not only in I/R and obesity, as its generation was unbiased.

Naturally, there are always technical limitations regarding sequencing methods in general and the full-length transcriptome in special: First, good pre-analytics is crucial as the quality of the original RNA is limiting the data availability on a transcriptomic level. Therefore, only samples with high RNA integrity score (RIN > 9) were used for further analysis. Second, to achieve 99% sequence identity to the true full-length mRNA sequence, automated data analysis removes sequences for which quality control criteria cannot be reached. Therefore, some potentially true sequences get lost throughout

(See figure on next page.)
Fig. 7 Impact of high-fat diet on promoter-specific expression. Expression levels for diet-induced obesity at baseline (Ctrl) and 3 or 16 days post ischemia / reperfusion injury (I/R) for different PGC-1α starting exons (**A–D**), for the four group-wise testable isoforms (**E–H**) and for the four single-detectable isoforms (**I–L**). Housekeeper-normalized expression values in area-at-risk (upper row) and remote area (bottom row). Colours of bars represent fold-change difference between high-fat (HFD) compared to standard chow (SD) diet, colour scheme likewise to heat map illustrations (i.e. dark green means higher, dark red means lesser expression in HFD compared to SD). **A** Exon 1a, originating from the canonical promoter. **B, C** Exons 1b and 1b′, under control by the alternative (known) promoter. **D** Novel exon 1c, originating from a new promoter site. **E** Isoforms with canonical C-terminal ending. **F** Isoforms with novel C-terminal ending. **G** Isoforms with N-terminal isoforms. **H** Isoforms with the novel exon 6b. **I** PGC-1α-trE6-Ex13c. **J** sNT-PGC-1α. **K** PGC-1α-E3c. **L** CT-PGC-1α-E3c. The recovery of expression in the infarct zone after 16 days (ratio between 3 and 16 days) under high-fat diet is impaired, leading to a continuing downregulation of exon 1a and b. For the group-wise detectable isoforms, the downregulation is dominant in the infarct zone and less in the remote area in direct comparison of expression values. Most of the isoforms recover expression values compared over time with exception of the N-terminal isoforms in infarct area. Data acquired by qPCR using primers covering specific exon-exon junctions. Expression values are normalized to housekeeper NUDC. $n = 4$ each, bars depict mean values, error bars represent SD. Significance calculated by unpaired Student's *t* test (ns $p \leq 0.05$, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$)
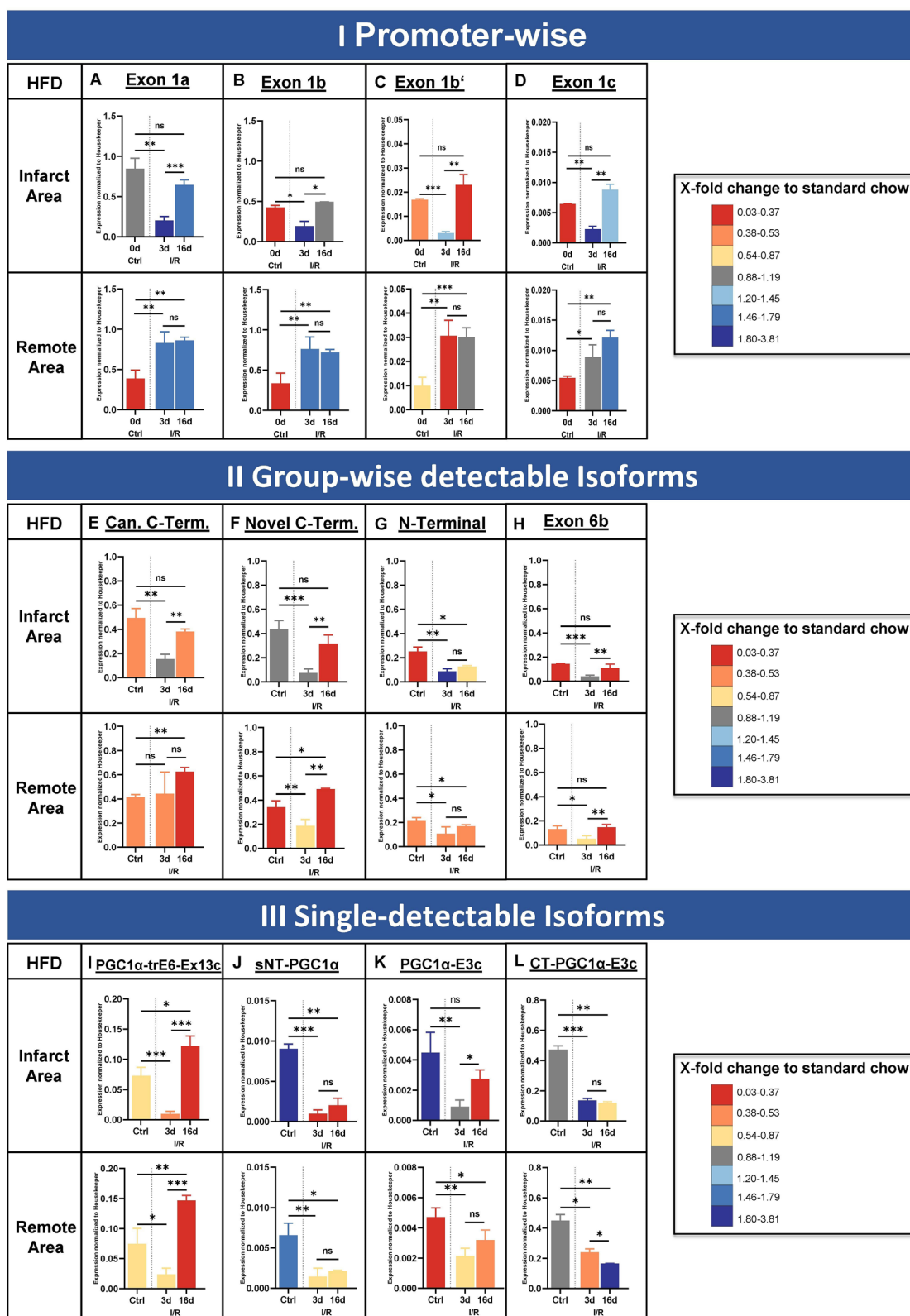
Fig. 7 (See legend on previous page.)

Oehler *et al. BMC Biology*   (2022) 20:169

Page 13 of 20

the process. To overcome this, we performed, a second, manual readthrough, using the non-filtered, full-length, non-concatemers, after the automatic pipeline assessment, thus minimizing loss of true mRNA sequences. Third, although we do not need to map short reads for assembly, mapping of the long-read sequences to the genome for correct assignment of the transcriptomic reads to genomic loci must be performed. The fidelity of the mapping relies on the underlying genomic data (which is in mice and human overall good) and additionally on the algorithm used. To overcome potential errors regarding software issues, we used different available mapping tools (minimap2 [32], STAR [33] and GMAP [34]) as well as different genomic database sources (GENCODE, ENSEMBL) resulting in comparable results. Fourth, if focussing on PGC-1α, limitations are also that low abundant transcripts are more difficult to detect. Here, long-read sequencing using SMRT can detect one unique transcript even if there is only one single molecule inside the sample. Even using enough so called Zero-Mode Waveguides (each containing the 'hole' in which one molecule is sequenced), sensitivity is always limited, and we cannot exclude that some very low abundant transcripts / isoforms of PGC-1α may not be (physically) detectable.

Investigation of PGC-1α expression pattern leads to discovery of 12 novel PGC-1α isoforms and a novel promoter

Focussing on PGC-1α, our bioinformatic pipeline identified most of the known PGC-1α transcripts, and strikingly, 12 novel isoforms. Furthermore, we discovered a putative novel promoter, controlling expression of a new starting exon, which we called 'exon 1c', following the terminology of known PGC-1α promoters. Notably, we identified this promoter also for human heart tissue enabled by existence of conserved transcripts. The novel predicted promoter site associated to exon 1c has a mammalian initiator element (Inr) and corresponding downstream promoter element (DPE) around the transcription start site of exon 1c, raising confidence about the existence of the promoter element.

The newly detected isoforms are in line with regulated gene expression which is commonly considered a combination of alternative splicing events and usage of different promoters: PGC-1α1, PGC-1α-b and PGC-1α-c, all known from literature, share the same 'sequence body' but start with either exon 1a, exon 1b or exon 1b'. In this present study, we demonstrated a new transcription start with the novel exon 1c and the very same sequence afterwards (hereafter called PGC-1α-d).

In the same manner, also the novel isoforms are derivatives of known sequences, but result from alternative internal splicing, giving rise to new exonic features within the mRNA sequence. Interestingly, most of those isoforms also follow the principle of shared core sequence with different starting exons. Although we did not perform SMRT sequencing in other organs, we confirmed expression of exon 1c in various organs including heart and skeletal muscle as well as brown adipose tissue by quantitative PCR. Therefore, we suggest that a novel promoter specifically active in those tissues drives expression of exon 1c. Obviously, although prediction tools improved over the years and have a currently high accuracy in forecasting promoter elements, proof can only be done experimentally, limiting the finding to only be 'putative'. However, our findings on the postulated promoter elements being common in different isoforms and organs in our dataset and across species strengthen this finding.

Of special interest is also one group of three novel isoforms lacking 'classical starting exons' due to alternative splicing events shifting the open reading frame. All of them share the same novel starting exon 3c, which originates from within the sequence of canonical exon 3 (see also Additional file 1: Figure S3, F). This mechanism of 'skipping' exon 1 and exon 2 makes this group different to all other known and novel isoforms discussed here and is of special interest as it resembles in that manner the structure of an isoform previously known in human liver (L-PGC-1α [35]). L-PGC-1α could not been detected in other vertebrates before, probably because the strategy for detection was based on sequence (which is different) rather than the common splicing mechanism. As they share a similar splicing mechanism (novel ORF in canonical exon 3 due to alternative splicing upstream with disruption of canonical ORF) and structural elements, they seem to be closely related and possibly functionally similar. However, future studies are needed to evaluate this aspect further.

Assigning known functional protein domains to our dataset of novel and known isoforms using prediction tools revealed existence of six main protein domains, including transcriptional activity domains as well as protein and RNA recognition motifs. As stated above, we grouped the isoform pattern of novel and known isoforms by structural categories in regard to their mRNA sequence and predicted thereby also potential functional domains within. While the first and third group exhibit all known domains ('canonical'), others miss some domains (such the PPARγ-binding domain as well as the RNA recognition motif in group 4 or the activation domains in

Oehler *et al. BMC Biology*   (2022) 20:169

Page 14 of 20

the fifth group). Even from interest for functional studies in future, further prediction of target specificity including differential transcriptional regulation is speculative and needs future in vitro and in vivo experiments to elucidate this further. In addition, although prediction algorithms improved, it is probable that novel functional domains exist in these isoforms.

Pre-diabetic state in mice leads to reduced alternative promoter-driven expression as well as reduction of PGC-1α expression in general

Others and we found that obesogenic diet in mice leads to a pre-diabetic metabolic state with higher blood glucose and insulin levels, evolving in insulin resistance due to diet-induced obesity (DIO). PGC-1α is functionally implicated in obesity-associated syndromes as high-fat diet counteracts expression of the PGC1 gene family in drosophila [36] and leads to upregulated apoptosis pathways in murine hepatocytes through inhibition of PGC-1α-mediated suppression of NFκB [28]. We found that DIO is associated with a reduction for most PGC-1α transcripts.

Transient promoter-driven PGC-1α downregulation in I/R in infarcted area in mice

In the heart, total knockout of PGC-1α leads to a DCM-like heart failure phenotype [24]. Data on isoform-specific functions in the heart are only available for PGC-1α1, suggesting a cardioprotective function when overexpressed [25].

We found that I/R led to a general downregulation of the whole PGC-1α isoform pattern in the infarct zone when compared to sham control, while no effect on the remote area was observed 3 days post intervention. Usually, scar-associated gene programmes (like extracellular matrix proteins and inflammatory pathways) are rapidly activated after myocardial infarction [37] in the infarct zone. In contrast to this, PGC-1α expression in general was downregulated in the infarcted area in our experiments, supporting the theory of a stunned myocardium with reprogrammed metabolic pathways due to lowered energy supply ('hibernating myocardium') [38]. Thus, transcriptomic regulation of PGC-1α mainly occurs in the infarcted area 3 days post I/R, while expression equalizes to sham controls for all isoforms, indicating a recovery of expression in the subacute phase after I/R injury.

Combined metabolic and ischemic stress leads to altered PGC-1α expression response

It could be shown over the last decade that myocardial ischemia and diabetic conditions are closely related and associated to each other and that in human cardiac tissue, lipid accumulation and insulin resistance increase vulnerability to ischemia-induced cardiac dysfunction [39]. As PGC-1α's role in energy metabolism as a 'mitochondrial masterswitch' in skeletal muscle and liver was studied comprehensively, but limited knowledge on its influence in cardiac metabolism exist, we aimed here to investigate both ischemic and metabolic stressors in the context of possible differential regulation of PGC-1α isoforms. We thereby found a differential expression of PGC-1α isoforms in the remote area uniquely in DIO, which could be explained by a shifted metabolic state in the remote area, while in the infarct area the hibernating myocardium is not capable to do so. Strikingly, when comparing relative expression values directly to those under standard chow, in DIO both the infarcted area and the remote myocardium exhibit downregulated expression of PGC-1α, which is different to lean mice, where only expression in the infarct zone itself is altered. Additionally, one novel isoform, PGC-1α-trE6-Ex13c, depicted a 13-fold (overcompensatory) increase in the infarct zone but no recovery at all in the remote area.

Our working hypothesis is that metabolic changes in diet-induced obesity shift isoform pattern towards distinct roles, e.g. towards hypertrophy or changes in metabolic pathways. To what extent the newly identified isoforms are cardioprotective or worsen outcome after myocardial infarction, especially under high-fat-diet conditions and pre-diabetic state, remains to be elucidated. Future studies are needed to evaluate possible time-course dependency of recovery in I/R during recovery from injury and if these changes are related to metabolic shifts. Another open question is hereby the underlying cell type which causes the observed expression changes (e.g. cardiomyocytes, infiltrating immune cells, fibroblasts). Additionally, as direct protein proof of novel isoforms is challenging, we cannot exclude that some of the isoforms are not translated. However, as we selected only sequences with poly-A-tail bioinformatically, and we have confidence on the existence of a transcript because of a full-length sequence, it is very likely that the respective protein is synthesized in vivo. Also, to what extent lower abundancy is also associated with lesser biological impact (e.g. relative transcriptional activity) cannot be concluded from our dataset of isoforms and need to be investigated in protein-related studies in future.

From a translational perspective, the differences of PGC-1α expression between lean and DIO mice in I/R could be extrapolated on diabetic patients and their

Oehler *et al. BMC Biology*   (2022) 20:169

Page 15 of 20

unfavourable outcome after myocardial infarction. Future prospective studies, including transcriptomics and metabolomics, are needed to elucidate pathomechanistical insights into the complex regulatory mechanisms of PGC-1α expression and function in the heart of those patients.

Conclusions

We deciphered for the first time a complete full-length-transcriptome of the murine and human heart, identifying novel putative PGC-1α coding transcripts including a novel promoter. These transcripts are differentially regulated in I/R and obesity suggesting transcriptional regulation and alternative splicing that may modulate PGC-1α function in the injured and metabolically challenged heart.

Methods
Mice
Male 12-week-old C57BL/6J mice (Janvier Labs, Le Genest-Saint-Ile, France) were used for all mice experiments. For experiments involving metabolic changes by diet-induced obesity, mice were fed 10 weeks before I/R with a high-fat, high-sucrose diet (24% Sucrose, 60 kJ & Fat; ID S7200-E010; ssniff Spezialdiäten GmbH, Soest, Germany). All other mice were fed with normal chow.

Glucose and insulin tolerance tests
For measurements of serum insulin levels, blood was isolated by cardiac puncture and centrifuged at $2000 \times g$ for 20 min and respective serum analysed for insulin levels using the Ultra Sensitive Rat Insulin ELISA Kit (Crystal Chemicals) according to the manufacturer's protocol. For GTT, mice were fasted for 17 h, challenged with i.p. injected glucose (2 mg/g body weight) and blood glucose levels measured before and 15, 30, 45, 60 and 120 min after glucose injection using a blood glucose metre (Ascensia Contour, Bayer). For ITT, mice were fasted for 5 h before i.p. injection of insulin (0.75 U insulin/kg body weight) and blood glucose levels measured before and 15, 30, 45, 60 and 90 min after insulin injection.

Ischemia/reperfusion (I/R) and sham surgery
Mice used for functional testing of isoforms underwent myocardial ischemia/reperfusion injury (I/R) or sham surgery as described previously [40]. In brief, mice were anesthetized with 2% isoflurane, orotracheally intubated and ventilated with oxygen-enriched gas (40% oxygen) using a rodent ventilator (Minivent microventilator, Hugo Sachs, Germany). Mice were placed in a supine position on a warming plate (Uno, Zevenaar, the Netherlands) to keep body temperature at 37.5 °C

and received buprenorphine (0.1 mg/kg body weight, subcutaneously [s.c.]) for analgesia. Electrocardiography (ECG) was recorded continuously. After lateral thoracotomy, the pericardium was dissected, and a 7-0 surgical prolene suture was cautiously passed underneath the LAD coronary artery at a position 1 mm from the tip of the left auricle. The suture ends were passed through silicon tubing to form a snare occluder. Myocardial ischemia was produced by tightening the snare and confirmed by blanching of the myocardium and change in ECG (decrease in S wave amplitude). After 45 min, the snare occluder was opened to initiate reperfusion. The sham-operated controls underwent the same procedure but without ligation of the LAD. Afterward, the suture was removed, and the chest was closed. At the end of the experimental procedures, mice were extubated after they regained spontaneous breathing. Animals received buprenorphine (0.05–0.1 mg/kg body weight, s.c.) every 8 h for up to 5 days for postoperative analgesia. Mice were excluded from the experiment when certain criteria of suffering were observed. These included weight losses greater than 20% of body weight, cessation of food and water ingestion or lack of voluntary movement.

Permission for animal experiments was granted by the Landesamt für Natur, Umwelt und Verbraucherschutz (LANUV) Nordrhein-Westfalen, Aktenzeichen 81-02.04.2017.A401.

Human tissue
Fresh tissue from human left ventricle, gained as remnant of left-ventricular assist-device (LVAD) implantation in a 58-year-old female with ischemic cardiomyopathy, was aseptically obtained at the University of Duesseldorf cardiovascular department (Institutional Review Board approval 5263R/2015104434), following the Declaration of Helsinki Principles. Directly after surgery, the heart sample underwent deep freezing in liquid nitrogen and was then immediately used for RNA isolation as described below.

RNA isolation and cDNA Synthesis for RT-PCR
Total RNA was isolated using the Fibrous Tissue RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Shortly, the homogenate of tissue and RLT buffer was treated with proteinase K for 20 min before RNA was extracted on Qiagen RNeasy columns. A DNAse digestion took place on-column. Quality of eluted RNA was assessed by microvolume spectrophotometer analysis (Nanodrop™, Thermo Fisher Scientific) and 2100 Bioanalyzer (Agilent). Only samples with RNA Integrity Number (RIN) above 9 were used. cDNA was synthesized

from 1 µg RNA using the QuantiTect reverse transcription kit (QIAGEN).

Library preparation

Total RNA from murine wildtype total heart and human heart sample from left ventricle was isolated as described above and used for library preparation according to the manufacturer's protocols (Iso-Seq™ Template Preparation for Sequel® Systems) without size selection.

Long-read RNA-sequencing

Long-read sequencing was performed at the Genomics & Transcriptomics Laboratory of the Biological-Medical Research Center (BMFZ, Heinrich Heine University Düsseldorf) using the Sequel I System by Pacific Biosciences (SMRT), according to the manufacturer's protocols. For each murine and human sample ($n = 1$ per species), 4 single-molecule real-time (SMRT) cells were used ($= 4 \times 10^6$ zero-mode waveguide holes per sample).

Bioinformatic analysis of SMRT data

Bioinformatic analysis used a modified Iso-Seq3-pipeline followed by an own pipeline of open-source tools and scripts. For primary analysis, the Isoseq3-Pipeline from Pacific Biosciences has been used; followed by a secondary Analysis Pipeline with open-source tools and modified scripts. This included mapping to genome with Minimap2 [32], collapsing transcripts with cDNA Cupcake and Annotation with SQANTI2 [41]. Due to the nature of the algorithm behind, transcripts with very low abundance and therefore low count number will be excluded by the pipeline automatically within the polishing steps. Usually, if not focusing on a specific gene locus, it is impossible to detect whether how many in fact true transcripts get lost by that manner. Hence, interested in specifically PGC-1α isoforms, we created a second search strategy (Fig. 1) by a similarity search strategy against sequence structures of the (previously)

longest known transcript PGC-1α-a ($=$ PGC-1α-1). Finding a balance between on the one hand losing transcripts by clustering, merging and polishing transcripts and on the other hand analysing artificial, false-positive reads resulting from too low quality, we decided to use sequences from an intermediate step, called full-length non-concatemer (FLNC) reads. Furthermore, starting from the raw data, only necessary quality control steps were performed to generate first circular-consensus reads (CCS) with primer removal and correct orientation ('full-length' or 'FL' reads) followed by refinement (poly-A trimming, concatemer removal). Those FLNC were then used as database for the similarity search resulting in about 300 sequences which were then manually annotated to known sequences of PGC-1α-transcripts using SnapGene [42]. This step is similar to the automatic filtering step in the manufacturer's IsoSeq3-pipeline; however, the automatic analysis is prone to eliminate often real sequences by merging similar transcripts to the same isoform, even if they only share similarities but are not identical. We therefore needed to work with pre-filtered sequences, and by aligning those sequences to the genome, single-nucleotide changes due to assembly errors of the SMRT polymerase could and had to be manually detected and corrected. As cut-off, we used only those transcripts which only had one single anticipated error within each sequence and therefore reaching similar accuracy values as the automatic pipeline but higher output of real PGC-1α isoforms.

Scripts for SMRT analysis

The following scripts have been used for analysis of SMRT data on the High-Performance Cluster using BioConda-Environment. As coding editor, 'Sublime Text' for Windows (Version Build 4113) was used. The Scripts here containing the essential information on settings; please be aware that not every step (e.g. copying or moving of files) is mentioned.

---

1. Creation of Circular Consensus Reads:
Used CCS Version: 6.0.0 (commit v6.0.0-2-gf165cc26)
Using libraries:

    unanimity : 6.0.0 (commit v6.0.0-2-gf165cc26)
    pbbam : 1.6.1 (commit SEQII-release-10.0.0-35-g8fc7d89)
    pbcopper : 1.8.0 (commit SEQII-release-10.0.0-38-g56f07ff)
    boost : 1.73
    htslib : 1.10.2
    zlib : 1.2.11

```
ccs --log-level INFO -j 20 --log-file CCS_"$PROJECTNAME"_"$i"_logfile.txt --reportFile
CCS_"$PROJECTNAME"_"$i"_report.txt "$rawfiles" "$PROJECTNAME"_"$i".ccs.bam;
```

2. Primer removal using LIMA-Tool:
   Used LIMA Version: lima 2.0.0 (commit v2.0.0)

```
lima "$PROJECTNAME"_"$i".ccs.bam "$primerpath" "$PROJECTNAME"_"$i"_demux.bam --isoseq --
peek-guess --dump-removed --dump-clips --log-level INFO --log-file
Lima_logfile_"$PROJECTNAME"_"$i".txt;
```

3. Refine Step (Creation of FLNC reads)
   Used Isoseq3-Version: isoseq3 3.4.0 (commit v3.4.0)

```
isoseq3 refine "$PROJECTNAME"_"$i"_demux.5p--3p.bam "$primerpath"
"$PROJECTNAME"_"$i"_flnc.bam --require-polya --log-level INFO --log-file
Refine_logfile_"$PROJECTNAME"_"$i".txt;
```

4. Clustering Step
   Used Isoseq3-Version: isoseq3 3.4.0 (commit v3.4.0)

```
isoseq3 cluster  flnc.fofn "$PROJECTNAME"_flnc_clustered.bam --verbose --use-qvs --log-
level INFO --log-file Cluster_logfile_"$PROJECTNAME".txt
```

5. Mapping of Long-Reads to Genome using minimap2:
   Used minimap2-Version: 2.18-r1015
   Used Genome and annotation files:
   Mice:

   - GRCm39.primary_assembly.genome.fa.gz (ENCODE)
   - gencode.vM27.annotation.gtf.gz (ENCODE)

   Note: For comparison reasons, analysis was also done additionally using ENSEMBL-Genome (Mus_musculus.GRCm39.dna.primary_assembly.fa.gz and Mus_musculus.GRCm39.104.chr.gtf.gz)
   Human:

   - GRCh38.primary_assembly.genome.fa.gz (GENCODE)
   - gencode.v38.annotation.gtf.gz (GENCODE)

   Note: For comparison reasons, analysis was also done additionally using ENSEMBL-Genome (Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz and Homo_sapiens.GRCh38.104.chr.gtf.gz)

```
minimap2 -t 20 -ax splice:hq -uf --secondary=no "$genomepath"/"$ensemblgene" \
"$PROJECTNAME"_flnc_clustered.fq > aligned_ensembl_"$PROJECTNAME"_flnc_clustered.sam 2> \
aligned_ensembl_"$PROJECTNAME"_flnc_clustered.sam.log
```

followed by sorting:

```
sort -k 3,3 -k 4,4n $DATEINAME.sam > sorted"$DATEINAME".sam;
```

6. Collapsing isoforms using Cupcake-Tool:
   Used collapse_isoforms_by_sam.py from Cupcake-Tools and seqkit.

   A) Removing Duplicates:

```
seqkit rmdup -D duplicate_list_"$DATEINAME".txt "$DATEINAME".fq > "$DATEINAME"_uni.fq;
```

   B) Collapse Transcripts:

```
collapse_isoforms_by_sam.py --input "$DATEINAME".fq --fq -s "$p" --dun-merge-5-shorter -o
"$p";
```

7. SQANTI3 for Reporting of Isoforms and Gene Usage
   Used R scripting front-end version 3.6.1 (2019-07-05) as well as SQANTI Version 3.0

```
sqanti3_qc.py --gtf "$p" "$PROJECTNAME"/SQANTI3/genome_SQ_G/"$gencodeannotation_SQ"
"$PROJECTNAME"/SQANTI3/genome_SQ_G/"$gencodegene_SQ" --aligner_choice=minimap2 -t 20;
```

Oehler *et al. BMC Biology*   (2022) 20:169

Page 18 of 20

Running environment of all Scripts was CentOS 7.7.1908-based at the High-Performance Cluster (HILBERT, Centre for Information and Media Technology (CIM/ZIM); Heinrich Heine University, Düsseldorf). Annotation of functional domains in transcripts was performed using open-accessible online prediction tools.

For the Motif annotation and prediction, the following tools were used: NCBI's Conserved Domain Database (CDD) [43], SMART protein domain annotation resource [44] and InterPro [45] as well as literature search on specific domain annotations of PGC-1α and detailed information on nuclear receptor domains [46–50]. The promoter prediction was done using ElemeNT [51].

Analysis of the longest canonical isoform PGC-1α-a by literature search as well as motif annotation and prediction tools [43–50] exhibited existence of mainly involvement of six protein domains: a transcription activation domain (AD1 [50], residues 30–40) accompanied by a second (AD2 [50], residues 82–95), 3) a *LLXXLL*-motif [47, 48, 50] (residues 141–147) involved in transcriptional regulatory processes, a binding domain (residues 292–338) for interaction with the upstream target PPARγ [46], a RNA-Binding-motif [43–45] (residues 677–746), possibly involved in splicing processes of downstream mRNA targets and the PDB domain 3D24|D [47, 49] (residues 198–218), involved in binding of the oestrogen-related receptor-alpha (ERRalpha).

Primer design and Sanger sequencing for validation of PGC-1α isoform sequences

Where possible, primer pairs (Additional file 1: Table S1 and Additional file 1: Tables S2 and S3) were designed to target specifically novel, unique exon-exon junctions or UTR-exon-junctions within the identified PGC-1α transcripts from long-read sequencing data, followed by quantitative real-time PCR (qPCR) and classical Sanger sequencing for confirmation and validation. In addition, qPCR primers to detect the known isoforms and starting exons were designed (Fig. 1, Additional file 1: Table S1 and Additional file 1: Tables S2 and S3).

Quantitative RT-PCR

Total RNA was isolated from heart tissue followed by cDNA synthesis as described above. qPCR was performed on the Step-One Plus real-time PCR system (Applied Biosystems) with Maxima SYBR Green and ROX qPCR Master Mix (Thermo Scientific). Transcript quantities were normalized to Nuclear Distribution Protein C (NUDC) mRNA. Bioinformatic analysis used a modified X0-Method approach, which is based on the comparative threshold cycle (CT) method [52].

Quality control and validation of qPCR primers

The primer set used for qPCR was tested using the basic local alignment search tool (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi) against both the NCBI database and internally in our unfiltered long-read transcriptomic data to provide specificity ('virtual PCR' approach, following published principles [53]). Quality control of the primers after qPCR was also performed by obtaining and evaluating the melting plots for every primer pair as well as gel electrophoresis of amplified cDNA fragments to check for estimated and real fragment length, followed by Sanger sequencing of the band to proof sequence identity.

Statistical analysis and figure making

Quantitative variables were compared by Students' *t* test. The tests were performed bilaterally, and the threshold of significance was set at .05. Unless otherwise stated, bars in graphs depict mean values and error bars represent standard deviation. Statistical analysis was performed using GraphPad Prism version 9 (GraphPad Software, Inc.) and IBM SPSS Statistics software version 27 (SPSS). Figures were created using GraphPad Prism and Microsoft PowerPoint. Individual data points for where $N < 6$ are provided in a separated file (see Additional file 2: Excel-File with individual data points for Figs. 5, 6 and 7).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-022-01360-w.

> **Additional file 1: Figure S1**: Quality Control for Primary, Secondary and Tertiary Analysis of SMRT-Sequencing. **Figure S2**: Sanger-Sequencing Results. **Figure S3**: Non-canonical exons (novel and previously known) in genomic context. **Figure S4**: Predicted Open reading Frame Ex1c (murine and human). **Figure S5**: Sanger-Sequencing Results of flanking primers approach for murine Exon1c/Exon2-junction. **Table S1**: Primers used for detecting starting exons with q(PCR). **Figure S6**: Pre-diabetic phenotype. **Figure S7**: Strategy for Detection of PGC1α-Isoforms in qPCR. **Figure S8**: Distribution of PGC-1α isoform expression in lean vs DIO under I/R. **Table S2**: Primers used for detecting isoform-pattern with q(PCR). **Table S3**: Primers used for flanking approach for Exon1c/Exon2 junction. 7
>
> **Additional file 2.** Individual data points for Figures 5, 6 and 7.

Authors' contributions

DO, AG and RW designed the study. DO conducted the experiments involving RNA isolation, PCR and sequencing. AS conducted the animal experiments including induction of myocardial infarction. Bioinformatical analysis was conducted by DO and AL. DO wrote the initial draft of the manuscript and DO, AG, NG and RW were the major contributors in writing the manuscript. DO, AG, NG, RW, JS and JR substantively revised the manuscript. RW, AG and MK supervised the project. All authors read and approved the final manuscript.

Oehler *et al. BMC Biology* (2022) 20:169

Page 19 of 20

Availability of data and materials
Raw transcriptomic sequencing data underlying this article are available through GenBank (NCBI) / Sequence Read Archive (SRA), and can be accessed with accession ID SRR16352840 (mus musculus, [54] and SRR16352587 (homo sapiens, [55]). Individual data points for Figs. 5, 6 and 7 are in Additional File 2.

Declarations

Ethics approval and consent to participate
Permission for animal experiments was granted by the Landesamt für Natur, Umwelt und Verbraucherschutz (LANUV) Nordrhein-Westfalen (Aktenzeichen 81-02.04.2017.A401). The study conformed to the principles of the Declaration of Helsinki and Good Clinical Practices. The human subject participated voluntarily with signed informed consent, and the study was approved by our local ethics committee (Institutional Review Board approval 5263R/2015104434).

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Division of Cardiology, Pulmonology, and Vascular Medicine, Medical Faculty, Heinrich-Heine University, Moorenstr. 5, 40225 Düsseldorf, Germany. [2]Cardiovascular Research Institute Düsseldorf (CARID), Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany. [3]Department of Cardiovascular Physiology, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany. [4]Molecular and Cellular Exercise Physiology, Department of Physiology and Pharmacology, Karolinska Institutet, SE-17177 Stockholm, Sweden. [5]Joint Heidelberg-IDC Translational Diabetes Program, Internal Medicine, Heidelberg University Hospital, Heidelberg, Germany. [6]German Center for Diabetes Research, Neuherberg, Germany.

References
1. Ruas JL, White JP, Rao RR, Kleiner S, Brannan KT, Harrison BC, et al. A PGC-1α isoform induced by resistance training regulates skeletal muscle hypertrophy. Cell. 2012;151(6):1319–31.
2. Martínez-Redondo V, Jannig PR, Correia JC, Ferreira DM, Cervenka I, Lindvall JM, et al. Peroxisome proliferator-activated receptor γ coactivator-1 α isoforms selectively regulate multiple splicing events on target genes. J Biol Chem. 2016;291(29):15169–84.
3. Martínez-Redondo V, Pettersson AT, Ruas JL. The hitchhiker's guide to PGC-1α isoform structure and biological functions. Diabetologia. 2015;58(9):1969–77.
4. Safdar A, Little JP, Stokl AJ, Hettinga BP, Akhtar M, Tarnopolsky MA. Exercise increases mitochondrial PGC-1 α content and promotes nuclear-mitochondrial cross-talk to coordinate mitochondrial biogenesis. J Biol Chem. 2018;293(13):4953.
5. Léveillé M, Besse-Patin A, Jouvet N, Gunes A, Sczelecki S, Jeromson S, et al. PGC-1α isoforms coordinate to balance hepatic metabolism and apoptosis in inflammatory environments. Mol Metab. 2020;34:72–84.
6. Agarwal S, Yadav A, Chaturvedi RK. Peroxisome proliferator-activated receptors (PPARs) as therapeutic target in neurodegenerative disorders. Biochem Biophys Res Commun. 2017;483(4):1166–77.
7. Mills R, Taylor-Weiner H, Correia JC, Agudelo LZ, Allodi I, Kolonelou C, et al. Neurturin is a PGC-1α1-controlled myokine that promotes motor neuron recruitment and neuromuscular junction formation. Mol Metab. 2018;7:12–22.
8. Girnun GD. The diverse role of the PPARγ coactivator 1 family of transcriptional coactivators in cancer. Semin Cell Dev Biol. 2012;23(4):381–8.
9. Yang S, Loro E, Wada S, Kim B, Tseng WJ, Li K, et al. Functional effects of muscle PGC-1alpha in aged animals. Skelet Muscle. 2020;10(1):14.
10. Kadlec AO, Barnes C, Durand MJ, Gutterman DD. Microvascular adaptations to exercise: protective effect of PGC-1 alpha. Am J Hypertens. 2018;31(2):240–6.
11. Correia JC, Ferreira DM, Ruas JL. Intercellular: local and systemic actions of skeletal muscle PGC-1s. Trends Endocrinol Metab. 2015;26(6):305–14.
12. Ruas JL, Lendahl U, Poellinger L. Modulation of vascular gene expression by hypoxia. Curr Opin Lipidol. 2007;18(5):508–14.
13. Eisele PS, Furrer R, Beer M, Handschin C. The PGC-1 coactivators promote an anti-inflammatory environment in skeletal muscle in vivo. Biochem Biophys Res Commun. 2015;464(3):692–7.
14. Sczelecki S, Besse-Patin A, Abboud A, Kleiner S, Laznik-Bogoslavski D, Wrann CD, et al. Loss of Pgc-1α expression in aging mouse muscle potentiates glucose intolerance and systemic inflammation. Am J Physiol Endocrinol Metab. 2014;306(2):E157–67.
15. Arany Z, Foo SY, Ma Y, Ruas JL, Bommi-Reddy A, Girnun G, et al. HIF-independent regulation of VEGF and angiogenesis by the transcriptional coactivator PGC-1alpha. Nature. 2008;451(7181):1008–12.
16. Thom R, Rowe GC, Jang C, Safdar A, Arany Z. Hypoxic induction of vascular endothelial growth factor (VEGF) and angiogenesis in muscle by truncated peroxisome proliferator-activated receptor γ coactivator (PGC)-1α. J Biol Chem. 2014;289(13):8810–7.
17. Allan R, Morton JP, Close GL, Drust B, Gregson W, Sharples AP. PGC-1α alternative promoter (Exon 1b) controls augmentation of total PGC-1α gene expression in response to cold water immersion and low glycogen availability. Eur J Appl Physiol. 2020;120(11):2487–93.
18. Patten IS, Arany Z. PGC-1 coactivators in the cardiovascular system. Trends Endocrinol Metab. 2012;23(2):90–7.
19. Di W, Lv J, Jiang S, Lu C, Yang Z, Ma Z, et al. PGC-1: the energetic regulator in cardiac metabolism. Curr Issues Mol Biol. 2018;28:29–46.
20. Rowe GC, Jiang A, Arany Z. PGC-1 coactivators in cardiac development and disease. Circ Res. 2010;107(7):825–38.
21. Garnier A, Fortin D, Deloménie C, Momken I, Veksler V, Ventura-Clapier R. Depressed mitochondrial transcription factors and oxidative capacity in rat failing cardiac and skeletal muscles. J Physiol. 2003;551(Pt 2):491–501.
22. Lehman JJ, Kelly DP. Transcriptional activation of energy metabolic switches in the developing and hypertrophied heart. Clin Exp Pharmacol Physiol. 2002;29(4):339–45.
23. Riehle C, Abel ED. PGC-1 proteins and heart failure. Trends Cardiovasc Med. 2012;22(4):98–105.
24. Kärkkäinen O, Tuomainen T, Mutikainen M, Lehtonen M, Ruas JL, Hanhineva K, et al. Heart specific PGC-1α deletion identifies metabolome of cardiac restricted metabolic heart failure. Cardiovasc Res. 2019;115(1):107–18.
25. Mutikainen M, Tuomainen T, Naumenko N, Huusko J, Smirin B, Laidinen S, et al. Peroxisome proliferator-activated receptor-γ coactivator 1 α1 induces a cardiac excitation-contraction coupling phenotype without metabolic remodelling. J Physiol. 2016;594(23):7049–71.
26. Lee WS, Kim J. Diabetic cardiomyopathy: where we are and where we are going. Korean J Intern Med. 2017;32(3):404–21.
27. Waldman M, Cohen K, Yadin D, Nudelman V, Gorfil D, Laniado-Schwartzman M, et al. Regulation of diabetic cardiomyopathy by caloric restriction is mediated by intracellular signaling pathways involving 'SIRT1 and PGC-1α'. Cardiovasc Diabetol. 2018;17(1):111.
28. Barroso WA, Victorino VJ, Jeremias IC, Petroni RC, Ariga SKK, Salles TA, et al. High-fat diet inhibits PGC-1α suppressive effect on NFκB signaling in hepatocytes. Eur J Nutr. 2018;57(5):1891–900.
29. Mitra R, Nogee DP, Zechner JF, Yea K, Gierasch CM, Kovacs A, et al. The transcriptional coactivators, PGC-1α and β, cooperate to maintain cardiac mitochondrial function during the early stages of insulin resistance. J Mol Cell Cardiol. 2012;52(3):701–10.
30. Leung SK, Jeffries AR, Castanho I, et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. Cell Rep. 2021;37(7):110022.

Oehler *et al. BMC Biology*   (2022) 20:169

Page 20 of 20

31. Zhu C, Wu J, Sun H, Briganti F, Meder B, Wei W, et al. Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes. Nat Commun. 2021;12(1):4203.
32. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
33. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
34. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for Genomic Sequence Alignment: enhancements to speed, accuracy, and functionality. Methods Mol Biol. 2016;1418:283–334.
35. Felder TK, Soyal SM, Oberkofler H, Hahne P, Auer S, Weiss R, et al. Characterization of novel peroxisome proliferator-activated receptor γ coactivator-1α (PGC-1α) isoform in human liver. J Biol Chem. 2011;286(50):42923–36.
36. Diop SB, Bisharat-Kernizan J, Birse RT, Oldham S, Ocorr K, Bodmer R. PGC-1/Spargel counteracts high-fat-diet-induced obesity and cardiac lipotoxicity downstream of TOR and brummer ATGL lipase. Cell Rep. 2015;10(9):1572–84.
37. Roy S, Khanna S, Kuhn DE, Rink C, Williams WT, Zweier JL, et al. Transcriptome analysis of the ischemia-reperfused remodeling myocardium: temporal changes in inflammation and extracellular matrix. Physiol Genomics. 2006;25(3):364–74.
38. Kloner RA. Stunned and hibernating myocardium: where are we nearly 4 decades later? J Am Heart Assoc. 2020;9(3):e015502.
39. Jelenik T, Flögel U, Álvarez-Hernández E, Scheiber D, Zweck E, Ding Z, et al. Insulin resistance and vulnerability to cardiac ischemia. Diabetes. 2018;67(12):2695–702.
40. Heinen A, Nederlof R, Panjwani P, Spychala A, Tschaidse T, Reffelt H, et al. IGF1 treatment improves cardiac remodeling after infarction by targeting myeloid cells. Mol Ther. 2019;27(1):46–58.
41. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018;28(3):396–411.
42. SnapGene software (from Insightful Science; available at snapgene.com).
43. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45(D1):D200–3.
44. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493–D6.
45. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2019;47(D1):D351–D60.
46. Teyssier C, Ma H, Emter R, Kralli A, Stallcup MR. Activation of nuclear receptor coactivator PGC-1alpha by arginine methylation. Genes Dev. 2005;19(12):1466–73.
47. Rha GB, Wu G, Shoelson SE, Chi YI. Multiple binding modes between HNF4alpha and the LXXLL motifs of PGC-1alpha lead to full activation. J Biol Chem. 2009;284(50):35165–76.
48. Greschik H, Althage M, Flaig R, Sato Y, Chavant V, Peluso-Iltis C, et al. Communication between the ERRalpha homodimer interface and the PGC-1alpha binding surface via the helix 8-9 loop. J Biol Chem. 2008;283(29):20220–30.
49. Jin L, Li Y. Structural and functional insights into nuclear receptor signaling. Adv Drug Deliv Rev. 2010;62(13):1218–26.
50. Sadana P, Park EA. Characterization of the transactivation domain in the peroxisome-proliferator-activated receptor gamma co-activator (PGC-1). Biochem J. 2007;403(3):511–8.
51. Sloutskin A, Danino YM, Orenstein Y, Zehavi Y, Doniger T, Shamir R, et al. ElemeNT: a computational tool for detecting core promoter elements. Transcription. 2015;6(3):41–50.
52. Thomsen R, Sølvsten CA, Linnet TE, Blechingberg J, Nielsen AL. Analysis of qPCR data by converting exponentially related Ct values into linearly related X0 values. J Bioinforma Comput Biol. 2010;8(5):885–900.
53. Govindkumar B, Kavyashree B, Patel K, Sasidharan K, Siva Arumugam T, Thomas L, et al. Ex-Ex primer: an experimentally validated tool for designing oligonucleotides spanning spliced nucleic acid regions from multiple species. J Biotechnol. 2022;343:1–6.
54. Long-Read Sequencing of mus musculus: adult heart. NCBI SRA accession: SRR16352840. 2022. https://www.ncbi.nlm.nih.gov/sra/?term=SRR16352840.
55. Long-Read Sequencing of homo sapiens: Heart (LVAD). NCBI SRA accession: SRR16352587. 2022. https://www.ncbi.nlm.nih.gov/sra/?term=SRR16352587.

Publisher's Note