

RESEARCH

Open Access



# CVD-associated SNPs with regulatory potential reveal novel non-coding disease genes

Chaonan Zhu<sup>1,3†</sup>, Nina Baumgarten<sup>1,2,3†</sup>, Meiqian Wu<sup>1</sup>, Yue Wang<sup>1</sup>, Arka Provo Das<sup>1,3</sup>, Jaskiran Kaur<sup>1</sup>, Fatemeh Behjati Ardakani<sup>1,2,3</sup>, Thanh Thuy Duong<sup>5</sup>, Minh Duc Pham<sup>1,3,4,5</sup>, Maria Duda<sup>5</sup>, Stefanie Dimmeler<sup>1,2,3</sup>, Ting Yuan<sup>1,3,4\*</sup>, Marcel H. Schulz<sup>1,2,3\*</sup> and Jaya Krishnan<sup>1,2,3,4\*</sup>

## Abstract

**Background** Cardiovascular diseases (CVDs) are the leading cause of death worldwide. Genome-wide association studies (GWAS) have identified many single nucleotide polymorphisms (SNPs) appearing in non-coding genomic regions in CVDs. The SNPs may alter gene expression by modifying transcription factor (TF) binding sites and lead to functional consequences in cardiovascular traits or diseases. To understand the underlying molecular mechanisms, it is crucial to identify which variations are involved and how they affect TF binding.

**Methods** The SNEEP (SNP exploration and analysis using epigenomics data) pipeline was used to identify regulatory SNPs, which alter the binding behavior of TFs and link GWAS SNPs to their potential target genes for six CVDs. The human-induced pluripotent stem cells derived cardiomyocytes (hiPSC-CMs), monoculture cardiac organoids (MCOs) and self-organized cardiac organoids (SCOs) were used in the study. Gene expression, cardiomyocyte size and cardiac contractility were assessed.

**Results** By using our integrative computational pipeline, we identified 1905 regulatory SNPs in CVD GWAS data. These were associated with hundreds of genes, half of them non-coding RNAs (ncRNAs), suggesting novel CVD genes. We experimentally tested 40 CVD-associated non-coding RNAs, among them RP11-98F14.11, RPL23AP92, IGBP1P1, and CTD-2383120.1, which were upregulated in hiPSC-CMs, MCOs and SCOs under hypoxic conditions. Further experiments showed that IGBP1P1 depletion rescued expression of hypertrophic marker genes, reduced hypoxia-induced cardiomyocyte size and improved hypoxia-reduced cardiac contractility in hiPSC-CMs and MCOs.

**Conclusions** IGBP1P1 is a novel ncRNA with key regulatory functions in modulating cardiomyocyte size and cardiac function in our disease models. Our data suggest ncRNA IGBP1P1 as a potential therapeutic target to improve cardiac function in CVDs.

**Keywords** ncRNA, Cardiovascular disease, Regulatory SNPs, IGBP1P1

<sup>†</sup>Chaonan Zhu and Nina Baumgarten these authors contributed equally.

\*Correspondence:

Ting Yuan

yuan@med.uni-frankfurt.de

Marcel H. Schulz

marcel.schulz@em.uni-frankfurt.de

Jaya Krishnan

Krishnan@med.uni-frankfurt.de

Full list of author information is available at the end of the article



## Introduction

Cardiovascular diseases (CVDs) are among the most common causes of death in the world. Finding novel molecular biomarkers is an important research goal, to enable development of novel early detection, treatment and intervention strategies.

Recently, non-coding RNAs (ncRNAs) have been found to play important roles in cellular processes related to many CVDs [1–3]. The ncRNA HERN1 (hypoxia-inducible enhancer RNA 1), which is produced by direct hypoxia-inducible factor 1 $\alpha$  binding to a hypoxia response element, modulates the cardiac growth, metabolic, and contractile gene program in pressure-overload heart disease [4]. Similarly, the ncRNA CARMEN is derived from a human super-enhancer (SE) and regulates cardiomyocyte differentiation and homeostasis in human cardiac precursor cells [5]. Moreover, inhibition of the ncRNA MEG3 (maternally expressed gene 3) decreased cardiac fibrosis and improved diastolic performance by targeting cardiac matrix metalloproteinase-2 (MMP-2) [6].

Different approaches can be used to associate a ncRNA with the pathology of a CVD. Genome-wide methods have been especially successful using different types of assays measuring RNA [7, 8], genome [9], epigenome [10] variation or image-measured physiological differences [11] in disease models or from patient data directly.

Genome data in the form of mutations, such as single nucleotide polymorphisms (SNPs), that are associated with CVDs through genome-wide association studies (GWAS), provide an interesting source of information for detection of relevant ncRNAs. In particular, because many mutations found associated with CVDs reside outside of protein-coding genes and their functional role is often unknown [9], However, it is difficult to connect SNPs that reside in the non-coding regions of the genome with potential biological functionality.

One promising direction for deciphering the functional role of such non-coding SNPs is variant annotation methods that use transcription factor (TF) binding or epigenetic information such as DNase1-seq, ATAC-seq or histone ChIP-seq data [12, 13]. To predict TF binding, different approaches exist using position weight matrices (PWMs) [14, 15], that are available for the majority of human TFs, or more complex methods such as deep learning-based models [16, 17], which are currently more limited due to lack of TF-specific data. Specific statistical methods have been developed to assess whether a SNP has a regulatory effect on TF binding [18–20].

Alternatively, SNPs can be categorized as *functionally important* by a computational model that assesses whether changes in the DNA sequence will affect gene or epigenome activity more generally [21–23]. In other

words, all these methods assess whether a SNP is likely to have a regulatory effect and may allow to predict the tissue and cell-type relevance of such an effect [24–26].

While prioritization of regulatory SNPs (rSNPs) with approaches mentioned above is important and an area of active research, another problem is to associate rSNPs with their potential target genes. Several approaches for linking regions to target genes exist using diverse data types [27, 28], such as the Activity-by-Contact model [29] or STITCHIT [30]. For example, the EpiRegio database [31] contains 2.4 million regulatory elements (REMs) that were linked to human target genes using STITCHIT utilizing paired DNase1-seq and RNA-seq data of several cell types.

Here, we present a characterization of ncRNAs that can be linked to genetic mutations associated with the CVDs, including Aortic stenosis, Coronary artery disease, Cardiomyopathy, Cardiac arrhythmia, Myocardial infarction or Myocardial ischemia. By using an algorithm to detect rSNPs as part of the SNEEP pipeline, hundreds of cardiovascular associated ncRNAs have been identified that harbor rSNPs in their gene-regulatory elements. To study the functions of some interesting ncRNAs, we used two models: the 2D human-induced pluripotent stem cells (hiPSCs) derived cardiomyocytes (hiPSC-CMs) model and 3D human cardiac organoids, which display a similar microenvironment and contractile function to the human heart. Through assessing the cardiomyocyte size and contractile function response to pathophysiologic stress, our data demonstrated that ncRNA immunoglobulin (CD79A) binding protein 1 pseudogene 1 (IGBP1P1) drives cardiac hypertrophy and contractile dysfunction.

## Methods

### Collection of GWAS SNPs of cardiovascular diseases

We have collected the significant SNPs ( $P$ -value  $< 10^{-5}$ ) from the NHGRI-EBI GWAS catalog [32] for the following search terms including all the available child traits: Coronary artery disease (EFO\_0000378), Aortic stenosis (EFO\_0000266), Cardiac arrhythmia (EFO\_0004269), Cardiomyopathy (EFO\_0000407), Myocardial infarction (EFO0000612) and Myocardial ischemia (EFO0005672). All GWAS were downloaded on 10/26/2020 (see also Additional file 6: Table S5).

For each set of GWAS SNPs, we have obtained correlated SNPs that are in linkage disequilibrium (LD) with any of the original SNPs. We used the LDProxy Tool [33] and extracted the proxy SNPs via their API functionality. Proxy SNPs from the European cohort with an  $R^2 > 0.75$  and within a window of  $\pm 500,000$  bp centered around the original SNP were added to the GWAS SNPs. The combined set of proxy and lead SNPs was used as input set to SNEEP. There was no filtering of SNPs, which

have known roles, such as stop codon variants, as these SNPs are not expected to affect TF binding and to reside in regulatory elements.

### Detection of regulatory SNPs

To detect regulatory SNPs, we applied the SNEEP pipeline (<https://github.com/SchulzLab/SNEEP>) separately for each of the 6 cardiac GWAS. SNEEP (SNP exploration and analysis using epigenomics data) is a computational pipeline, which identifies rSNPs along with the affected TFs and further links the rSNPs to putative target genes. It can be used with any set of SNPs, TF motif and enhancer-gene resource.

To compute whether or not a SNP alters the binding behavior of a TF, a differential binding score is determined, which is the log-odds ratio between the binding affinity of the wildtype sequence (containing the wild type allele) and the mutated sequence (containing the alternative allele) [20]. For the log-odds ratio a differential binding *P*-value is computed. The approximation of the *P*-value depends on the characteristics like length, CG content etc. of the used TF PWM-motifs. Therefore, one needs to estimate a scale value per motif using the script *estimateScalePerMotif.sh* from the SNEEP pipeline. To estimate the scale parameter values, we used 200 000 sampled SNPs from the dbSNP database [34], and removed flanking bases of the PWMs with an entropy higher than 1.9, resulting in the following command: `bash estimateScalePerMotif.sh 200,000 <pathToMotifs> <outputDir> <motifNames> 1.9`

To run SNEEP, 632 human TF PWM-motifs in transfac format were gathered from the JASPAR database (version 2020) [35] and over 2.4 million regulatory elements linked to their putative target genes were downloaded from the EpiRegio database [31] (<https://doi.org/10.5281/zenodo.3758494>, file: REMAnnotationModelScore\_1.csv.gz).

Next, we applied the main SNEEP pipeline per GWAS with a differential binding *P*-value cutoff of 0.001:

```
./differentialBindingAffinity_multipleSNPs -o <SneepOutputDirectory>
-n 10 -p 0.5 -c 0.001 -r <EpiRegio_REMs> -g
<mappingEnsemblIdToGeneNameForREMs> -j 100 -l 123 -i
<pathToSneepDirectory> -s <estimatedScalesFromPreviousStep>
<pathToJasparMotifs> <InputSnpsPerCardiacGWAS> <pathTohg38.fa>
```

The SNEEP result is provided in Additional file 7: Table S6.

### Identification of disease associated genes using rSNPs

As part of the analysis of SNEEP, all rSNPs that overlap regulatory elements from EpiRegio constitute a candidate

disease gene. From the SNEEP output file protein-coding and non-coding genes were extracted that have overlapping rSNPs in their regulatory elements. Non-coding genes were understood to be all genes not labeled with the biotype ‘protein coding’ or ‘TEC’ (primary assembly annotation, version 39, downloaded from GENCODE) [36].

To label which protein-coding genes are already associated with the studied diseases (Fig. 2B, black dots), we used the *disease2gene* functionality of the R package of DisGeNET:

```
disease2gene(disease =< studiedDisease >,
database = ALL, score = c(0, 1))
```

The *studiedDisease* parameter needs to be provided as UMLS CUI identifiers. We used C1956346 (Coronary artery disease, CAD), C0003811 (Cardiac arrhythmia), C1449563 (Cardiomyopathy, Familial Idiopathic), C0151744 (Myocardial ischemia), C0027051 (Myocardial infarction) and C0340375 (Subaortic stenosis) as *studiedDisease* (see also Additional file 3: Table S2).

### Identification of co-expressed genes for non-coding genes and disease enrichment

We conducted a co-expression analysis using the gene expression profiles of 9,662 GTEx RNA-seq samples [37], which is also visualized in Fig. 3A. We compared the expression of protein-coding genes with the non-coding genes using the Spearman correlation coefficient as the similarity metric (Additional file 8: Table S7). This allowed us to obtain a ranked list of protein-coding genes most similar to the expression of a selected non-coding gene in the GTEx data.

For each non-coding gene the top 10 co-expressed protein coding genes were extracted, varying this number did not change the further results. The joint set of all protein-coding genes that are co-expressed to any of the non-coding genes found for the same disease via the

GWAS analysis, were used to perform a disease enrichment analysis. The analysis was done separately for the resulting co-expressed protein-coding gene sets derived from Cardiac arrhythmia, CAD and Cardiomyopathy using the function *disease\_enrichment* from the R package of DisGeNET [38].

```
disease_enrichment(entities =
  < coExpressedProteinCodingGenesPerGWAS >,
  vocabulary = HGNC, database = ALL)
```

From the resulting list of enriched diseases (Additional file 4: Table S3), as part of the enrichment computation, cardiac phenotypes were selected and visualized as a dot plot using ggplot2 (Fig. 3C). For the GWAS Aortic stenosis, Myocardial infarction and Myocardial ischemia we did not apply the disease enrichment analysis because of the low number of associated non-coding genes (less than 30).

#### Preparation and maintenance of human iPSC-CMs

Human-induced pluripotent stem cells (hiPSCs) were purchased from Cellular Dynamics International (CMC-100-010-001) and cultured according to the manufacturer's protocol. The human iPSC-derived cardiomyocytes (hiPSC-CMs) were reprogrammed using the STEMdiff™ Cardiomyocyte Differentiation Kit (STEMCELL Technologies) as recommended by the manufacturer. Briefly, human iPSCs were plated at cell density of  $3.5 \times 10^5$  cells/well on Matrigel coated 12-well-plates using mTeSR™ medium supplemented with 5  $\mu$ M ROCK inhibitor (Y-27632, STEMCELL Technologies) for 24 h. After 1 day (-1), the medium was replaced with fresh TeSR™ medium. To induce cardiac differentiation, the TeSR™ medium was replaced with Medium A (STEMdiff™ Cardiomyocyte Differentiation Basal Medium containing Supplement A) at day 0, Medium B (STEMdiff™ Cardiomyocyte Differentiation Basal Medium containing Supplement B) at day 2, Medium C (STEMdiff™ Cardiomyocyte Differentiation Basal Medium containing Supplement C) at day 4 and day 6. On day 8, medium was switched to STEMdiff™ Cardiomyocyte Maintenance Medium with full medium changes every 2 days, to promote further differentiation into mature cardiomyocyte cells. All experiments were performed in the hiPSC-CMs at day 40. Hypoxic condition was achieved by using the Hypoxia chamber and the hiPSC-CMs were cultured at either 3% or 1% O<sub>2</sub> for 2 days.

#### Monoculture cardiac organoid formation technique

Monoculture cardiac organoids (MCOs) were created by hiPSC-CMs. Aggrewell™ 800 microwell culture plates were used to create the MCOs in STEMdiff™ Cardiomyocyte Support Medium (STEMCELL Technologies). At day 18, hiPSC-CMs were distributed into Aggrewell™ 800 microwell culture plates at a density of 900,000 hiPSC-CMs/well. After 2 days of culture, medium was switched

to STEMdiff™ Cardiomyocyte Maintenance Medium for long term culture. Hypoxic condition was achieved by using the Hypoxia chamber and the MCOs were cultured at either 3% or 1% O<sub>2</sub> for 3 days.

#### Self-organized cardiac organoids formation technique

Human iPSCs were plated at cell density of  $1.5 \times 10^5$  cells/well on Aggrewell™ 800 microwell culture plates to form embryoid bodies (EBs). At day 18 and day 20, 50 nM VEGF and 25 nM FGF were added into the Maintenance Medium. On day 22, medium was switched to Maintenance Medium with EGM-2 with full medium changes every 2 days. All experiments were performed in the self-organized cardiac organoids (SCOs) at day 40. Hypoxic condition was achieved by using the Hypoxia chamber and the SCOs were cultured at either 3% or 1% O<sub>2</sub> for 3 days.

#### Human heart biopsies

Human heart biopsies were provided by Prof. Dr. Silke Kauferstein (Goethe University Hospital, Frankfurt am Main, Germany). The human heart biopsies were conducted in compliance with the local ethics committee. Samples from healthy hearts as control and from hearts macroscopically visible signs of acute cardiac infarction as MI hearts, and RNA was extracted.

#### RNA isolation, reverse transcription and qRT-PCR

Samples were harvested in QIAzol Lysis Reagent (QIAGEN), and total mRNAs were isolated with RNeasy Kit (QIAGEN) according to the manufacturer's protocol. 200 ng total RNAs were reverse transcript into cDNA using QuantiTect Reverse Transcription Kit (QIAGEN) according to the manufacturer's protocol. The Applied Biosystems StepOnePlus Real-Time PCR system (Applied Biosystems, CA, USA) with Fast SYBR Green Master Mix (Thermo Fisher Scientific) were used for analysis. Gene expression levels were normalized against the housekeeping gene *HPRT1*. The qRT-PCR primers are listed in Table 1.

#### Antisense LNA gapmeRs

Antisense LNA GapmeRs were purchased from Qiagen. Four different antisense LNA GapmeRs were designed for each target (Table 2). The GapmeR negative control was used as the control.

#### Contractility measurement

Every single MCO was transferred into 96-well-plate and treated with either GapmeR control, GapmeR

**Table 1** List of qRT-PCR primers used in the study

Target gene name	Forward primer	Reverse primer
RP11-146N23.4	TGTTTTGTCTCCGGTGTCCA	ATAAGGCCAGGTGCGGTG
RP11-513M16.8	TATCTGCGCCTTAACCAGACC	AATGCAAGCTCTTTGTGGCA
RP11-77K12.9	CCATGAAAGTCGGCCCAAGA	GCGTAGGGTCACACTCTTCC
KRT18P62	GCCTTCATCGTTCTGCACAC	CATGGATGTCGCTCCTCACA
AP000487.5	GAGCCTCTTCATCTCCATCC	CACGCATGTGGCCCTTTCT
RP11-252K23.1	GACCAACGTTTCTGGCTTGA	GCTGGGACTCAGAAGTTGCTT
RP11-77K12.5 / TMEM231P1	AGGCAGAGCCATGAAACTCC	CACCCAGGCGAACTTTACCA
WBP1LP2	TACAGTGACTCCAGCTACGC	CTTGGGGTCTTGTGATGCT
ADORA2A-AS1	GCACCACATGCTTGTCTACGA	ATAGAGTCAGGGTTCCAGGCA
SNORD56	ATGTCAATAGTTTTTCATCAACAGCA	CCACTCAGACCCAAAGTATCGAC
RPL10AP1	CCCACGAATCCTCGGCATAG	TGCACAAGCTCATCGTCTGT
SUMO2P17	TGAGGTAGATCAGATTCCCATTCT	TATCTTCATCCTCCATTTCCAAC
HIF1AP1	TGGAACATTATTAACAGCAGCCAG	TTGCATCTTTTACACGTTTCTAGG
POM121L9P	CCAGCATCTTATTAGAGGACGGAA	TCCCTGAGGACTCTAGCACG
RP11-98F14.12	CCTGGATGCAGGCATGCTAA	GACCTGACCTGGCACAGTTG
OR7E7P	ATGGTGTAGTGGCGTCAGTG	CTCCGAGGGCACTTTGTAT
RP11-253E3.3	CACCTAGTGGCTCTTTGGGG	GAGTGCCAGACACACGGTAA
RP11-479F13.1	TGGACATGGATTGGTTGAGT	GTTTCTCATGCTGAGAGTGGC
RNA5SP320	TGAACACAAATGCGCAGAGT	AGTTCTCAGTTCATCTCCCATCC
RP11-98F14.11	GGAGGTCCTGTAGATCCGGT	GCTGAGAAGGCGCTGATTTT
RPL23AP92	GCGACCAACAAGTTCTCCCA	GCCCTGATCACGGTGTGAT
RP11-433J22.2-transcript201	AACTGTAAGGAGCTGCAGGG	GCCCTGGGGGAAAATTCTTGG
RP11-433J22.2-transcript202	GGCCATCTCACCCTACTCC	CACAGACAACCTGATCACCCT
KB-1440D3.13	ATCTCTGTGCCACCTTGAG	ACCTCCTTAGTTCCCAGCGT
IGBP1P1	GATCAGGGAATAGCCAAGGCA	CTCTGTTGGCTGCCATAGTC
RP11-326N17.2	AGTGCTGGTTACCAACTTTCTCT	GGAGTGCCAAGATCGCATGA
HLA-DQB1-AS1	CAGCTTGATGCAGATGTGTGG	CATGATGGTGGCTACTGCCT
ZMYM4-AS1	GACATGCTGTCAAGGGTAGGA	CCAGACTGACCTTATCATTGTGGT
HLA-DRB6	TTGGAGCAGGCTAAGTGTGAG	TCCGTAAGTGCCTGGAATC
Y RNA	GCTGGTAGTGAGTTATCTTG	ACAGACTAGCCAAGTGCAGTA
RP11-624L4.1-transcript 1–4	ACCAGAAGCACTCCAAGAACA	GACTTTGAAGTGACAGGCTGG
RP11-624L4.1-transcript 5	ATGCAGCCATCAGCCTCAAT	GCTTTTGGGTTGGTTGCCTT
RP11-624L4.1-transcript 6,9	CCTGCTGTGGGAGTAACCAT	AAGCCCTAGAGGGACAAGGT
RP11-624L4.1-transcript 7–8	TGGCCTGCATCCACTGTCT	AGACCCAAGATGGCCGAATAGG
RP11-649A18.4	TCAAGCTCAGCTCACAGCAT	TCACCAAGCAGGTAACCAATGT
Metazoa SRP	ATACTGATGGGGTGTCTGCA	GTCCCGAACTCCTGACCTC
CTD-2383I20.1	GACTCTGGCCTGAAGAAAGCA	TTGGCTCTCGGTGATCCTACT
RPL23AP36	GCTAAAAGGCATCCACCCCA	GGCTGCGTTTGGAAACCATAG
RP11-378J18.8	AATGACCCTCTGTCTTCTGT	AGGGTACAGTTGTTAGGGTAACG
RP11-20J15.3	CCTGTGACCCGGATCCAAC	CAAATGAACAGAAGCTGGGGG
MIR4269	CCTGCAGGCACAGACAGC	CCATCCCAGGCCCTGACAGA
CTC-215O4.4	AGGCCGCATTAAGAGCATGA	GTAGCAGGTCCTGTGTGAGG
KRT8P15	TTCGGCAACTGCTCCTATGC	CACTGGCCCCACCATAACTT
RP11-218C14.8	GGCCCTGTAAATGCCTCCC	AGATGAAAGGTGCAAGGGCG
COL3A1	TTGAAGGAGGATGTTCCCATCT	ACAGACACATATTTGGCATGGTT
MMP2	TACAGGATCATTGGCTACACACC	GGTCACATCGCTCCAGACT
TGFb1	CAATCTGCGGATACCTCAG	GCACAACCTCCGGTGACATCAA
NPPA	CAACGCAGACTGATGGATTT	AGCCCCCGCTTCTTCACTC
NPPB	TGGAACGTCGGGTTACAG	CTGATCCGGTCCATCTTCT

**Table 1** (continued)

Target gene name	Forward primer	Reverse primer
MYH7	GACCAGATGAATGAGCACCG	GGTGAGGTCGTTGACAGAAC
MYH6	CTCCTCTACGCAACTGCCG	CGACACCGTCTGGAAGGATGA
RP11-98F14.11	TAAATTGAAGCACGCGGAGAG	GCTGACCTCTGAGAAGCGTG
RPL23AP92	GCGACCAACAAGTTCTCCCA	GCCCTGATCACGGTGTGAT
IGBP1P1	GATCAGGGAATAGCCAAGGCA	CTCTGTTGGCTGCCATAGTC
CTD-2383I20.1	TGGCCTCATCAGAACCAAGAC	CAAGCTGGCCATTTTGTGTTGT
HPRT	CCTGGCGTCGTGATTAGTGAT	AGACGTTTCAGTCTGTCCATAA

**Table 2** Characteristics of the LNA GapmeRs used in the study

ID	Sequence (5'–3')
GapmeR negative Control	AACACGTCTATACGC
GapmeR RP11-98F14.11 #1	CGCGTGCTTCAATTTA
GapmeR RP11-98F14.11 #2	CTCCGCGTGCTTCAAT
GapmeR RP11-98F14.11 #3	CTGAGAAGGCGCTGAT
GapmeR RP11-98F14.11 #4	TCGCTGAGAAGGCGCT
GapmeR RPL23AP92 #1	TTGGTCGCATAGTGGT
GapmeR RPL23AP92 #2	TTTTGACCTCGACATC
GapmeR RPL23AP92 #3	CGACATCCACAATGAA
GapmeR RPL23AP92 #4	GCTTTTTGACCTCGAC
GapmeR IGBP1P1 #1	TGGCGAGATGAATTAG
GapmeR IGBP1P1 #2	GGATAAGCCATGGAGA
GapmeR IGBP1P1 #3	GAAATTAGCAGTGTGA
GapmeR IGBP1P1 #4	GCAATGAGGCTAGGAT
GapmeR CTD-2383I20.1 #1	GAAGGTTGGTCGCTCT
GapmeR CTD-2383I20.1 #2	GTTGGTCGCTTGGTT
GapmeR CTD-2383I20.1 #3	CTTGCCAGCATCTGAT
GapmeR CTD-2383I20.1 #4	AACAAGCTGGCCATTT

RP11-98F14.11, GapmeR RPL23AP92 or GapmeR IGBP1P1. Hypoxic condition was achieved by using the Hypoxia chamber and cardiac organoids were cultured at 3% O<sub>2</sub> for 3 days, then the contractility will be analyzed by IonOptix system. Units/pixels were determined by calibrating the system with a micrometer.

#### Calcium transient measurements

The MCOs were cultured in the 96-well-plate. 2 μM Cal-520 AM (AAT bioquest) with 0.04% Pluronic® F-127 (AAT bioquest) working solution was added into the plate, and then the plate was incubated in the incubator at 37 °C for 90 min. After washing the MCOs with PBS for 3 times, and with medium for 2 times, the MCOs will be transferred to the 384 well U-bottom plate and incubated at 37 °C for 1 h. The fluorescence will be measured by the fluorescence plate reader.

#### Immunofluorescence staining

Immunofluorescence staining was performed as described previously. After fixation with 4% paraformaldehyde (PFA)/PBS, the hiPSC-CMs or MCOs were permeabilized and incubated overnight at 4°C with primary antibodies against sarcomeric α-actinin (Sigma-Aldrich), diluted in 2% (v/v) HS/PBS. After 3 washes with PBS for 5 min, cells were incubated with 4',6-diamidino-2-phenylindole (DAPI Thermo Fisher Scientific) and AlexaFluor 555 anti-mouse (Thermo Fisher Scientific) secondary antibody for 1 h at room temperature. Dishes were mounted onto glass slides (Fisher Scientific) with a drop of ProLong™ Gold Antifade (Thermo Fisher Scientific). Fluorescent images were acquired with the SP5 confocal microscopy (Leica) using a 40× magnification. Cell size was quantified blindly using the software Image J.

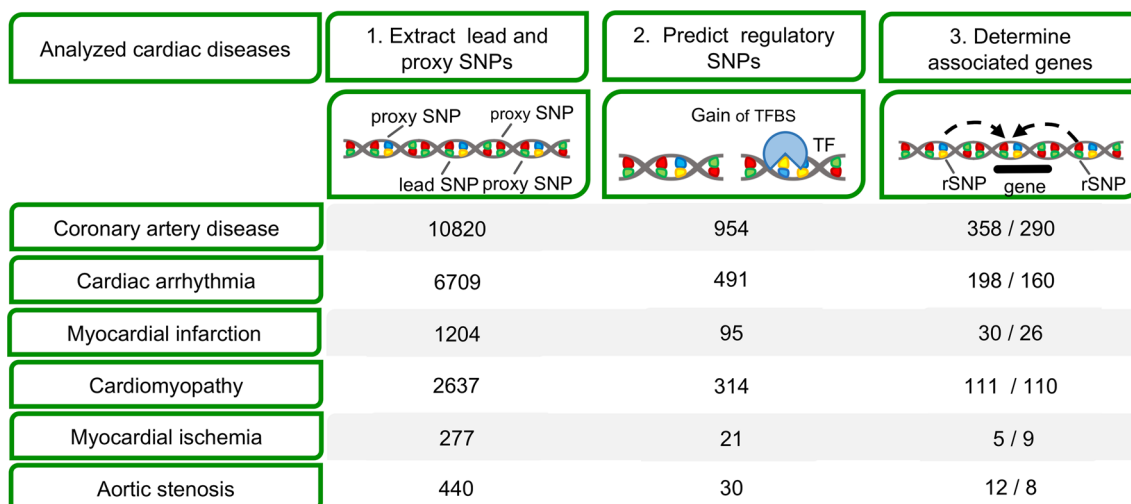
#### Statistical analysis

Data are represented as mean and error bars indicate the standard error of the mean (SEM). Two-tailed unpaired Student's *t*-tests or one-way ANOVA analyses followed by either a Dunnett's multiple comparison post-test (multiple comparisons to a single control) or Bonferroni correction (multiple comparisons between different groups) were used as indicated in the respective figure legends. *P* values were determined with Prism 9.0 (GraphPad) and *P* < 0.05 was considered statistically significant.

## Results

#### Identification of cardiovascular disease associated genes using regulatory SNPs and enhancer-gene linkage

Based on the previously postulated idea to identify rSNPs that have a regulatory effect, our pipeline SNEEP combines three sources of information for finding genes related to a CVD: (1) SNPs found significantly associated in GWAS with a particular CVD [32], (2) prediction which of these SNPs are potentially regulatory using human TF PWM-motifs [20, 35], and (3) enhancer-gene



**Fig. 1** Overview of the bioinformatic pipeline to identify non-coding genes associated with cardiovascular diseases. Step 1: SNPs for 6 different cardiac diseases were collected from the NHGRI-EBI GWAS catalog. Step 2: We used existing transcription factor binding models to filter regulatory SNPs (rSNPs), that are predicted to have an impact on transcription factor binding sites (TFBS). Step 3: rSNPs are linked to putative target genes using enhancer-gene links. For each step the number of SNPs or genes separated for coding and non-coding for each disease is given per row

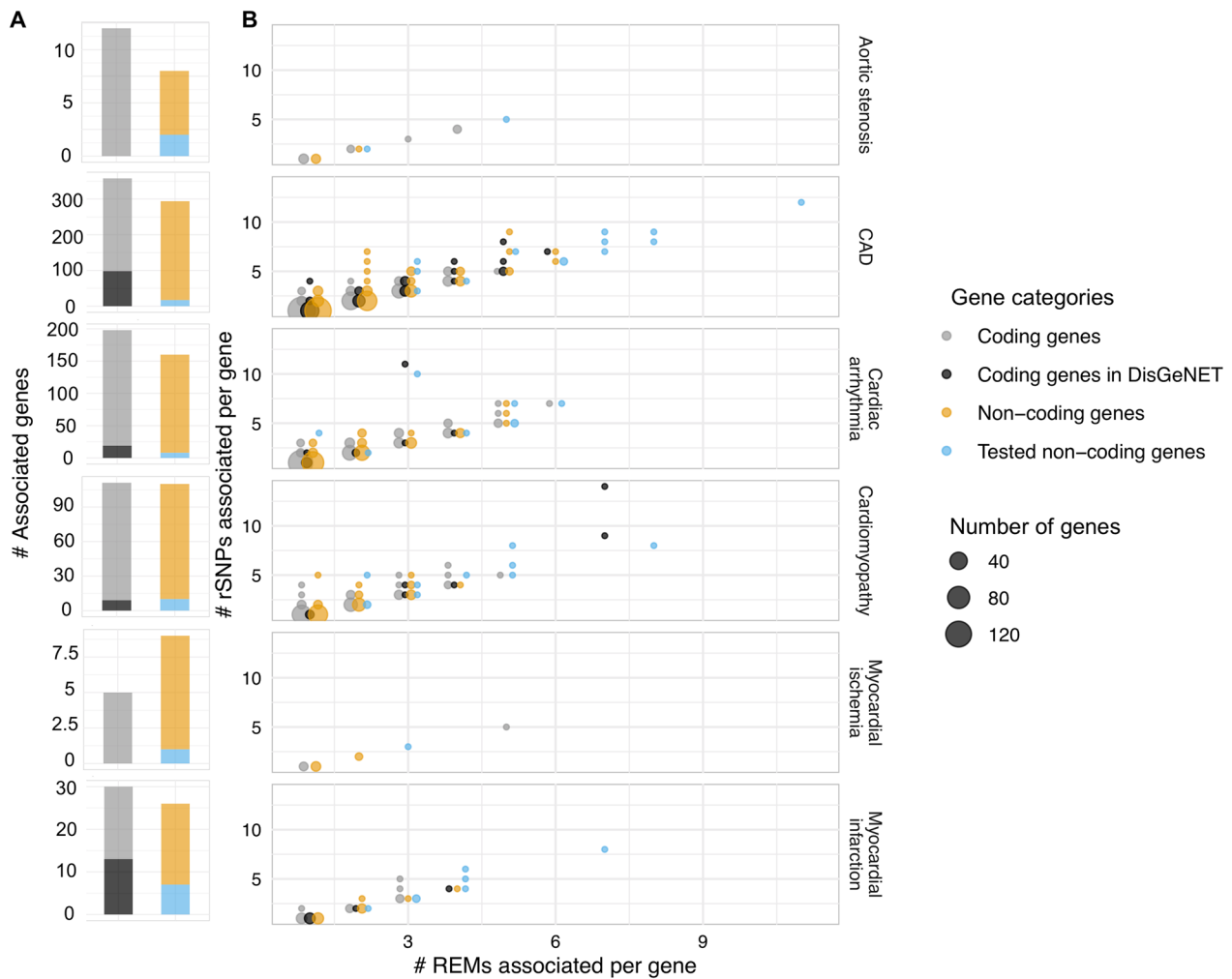
catalogue to map SNPs to putative target genes (Fig. 1) [31]. In short, SNEEP uses PWM descriptions of TF binding sites to assess whether a SNP would affect the binding of a known TF. Such an rSNP may lead to the loss of a TF binding site or the creation of a new binding site affecting the expression of one or several target genes related to the phenotype.

In the first step, we have retrieved significant lead SNPs and correlated SNPs in LD ( $R^2 \geq 0.75$ ) from GWAS for known cardiovascular diseases: 10,820 for Coronary artery disease (CAD), 440 for Aortic stenosis, 2637 for Cardiomyopathy, 6709 for Cardiac arrhythmia, 1204 for Myocardial infarction and 277 for Myocardial ischemia, which were then subjected to SNEEP analysis (Fig. 1). First, we filtered for those SNPs that may have a regulatory function (rSNPs) and affect TF binding. Then, we utilized the EpiRegio database, containing 2.4 million human regulatory elements, to link rSNPs to possible target genes for the respective indications.

These analyses led to the identification of protein-coding and non-coding genes that could be directly linked to a specific indication (Fig. 2A). Notably, there were often similarly many non-coding and protein-coding genes associated. We used the DisGeNET database [38], a large collection of known disease associated genes, for a positive control experiment. We conducted a disease enrichment analysis that assessed whether the newly identified protein-coding genes using our approach are enriched among previously associated disease genes. We found that for all tested indications we were able to get the corresponding disease phenotype as significantly enriched

(Fisher’s exact test,  $FDR \leq 0.05$ , Additional file 1: Fig. S1, Additional file 2: Table S1). For Myocardial ischemia and Aortic stenosis, we had less than 30 genes available and therefore enrichment analysis was omitted as it is statistically underpowered. We also computed enrichment of Gene Ontology and Human Phenotype Ontology terms and pathways using the g:Profiler webserver [39]. These analyses have revealed genes and biological processes that were previously associated with the diseases (Additional file 5: Table S4). For example, the terms *cardiac muscle contraction* or *regulation of ventricular cardiac muscle cell depolarization* were enriched for Cardiac Arrhythmia, including genes such as Gap Junction Protein Alpha 1 (GJA) and T-Box Transcription Factor 5 (TBX5) with known roles in the disease [40, 41]. Other examples are *PD-1 and IL6 signaling* (Cardiomyopathy) and *TGF-beta signaling* (Myocardial Infarction). For CAD a process enriched was *response to lipids* with many genes detected for that term, such as the Calcitonin receptor like receptor (CALCRL) gene involved in calcitonin regulation, connected with increased severity of CAD [42]. Together, the enrichment analysis results support that our approach, although limited to using rSNPs, is able to find many of the previously associated disease genes and recapitulates knowledge about involved biological functions and pathways.

We systematically compared the properties of the associated genes, looking at the number of rSNPs that could be linked to each gene and the number of regulatory elements of each gene with at least one rSNP (Fig. 2B, Additional file 3: Table S2). Protein- and non-coding

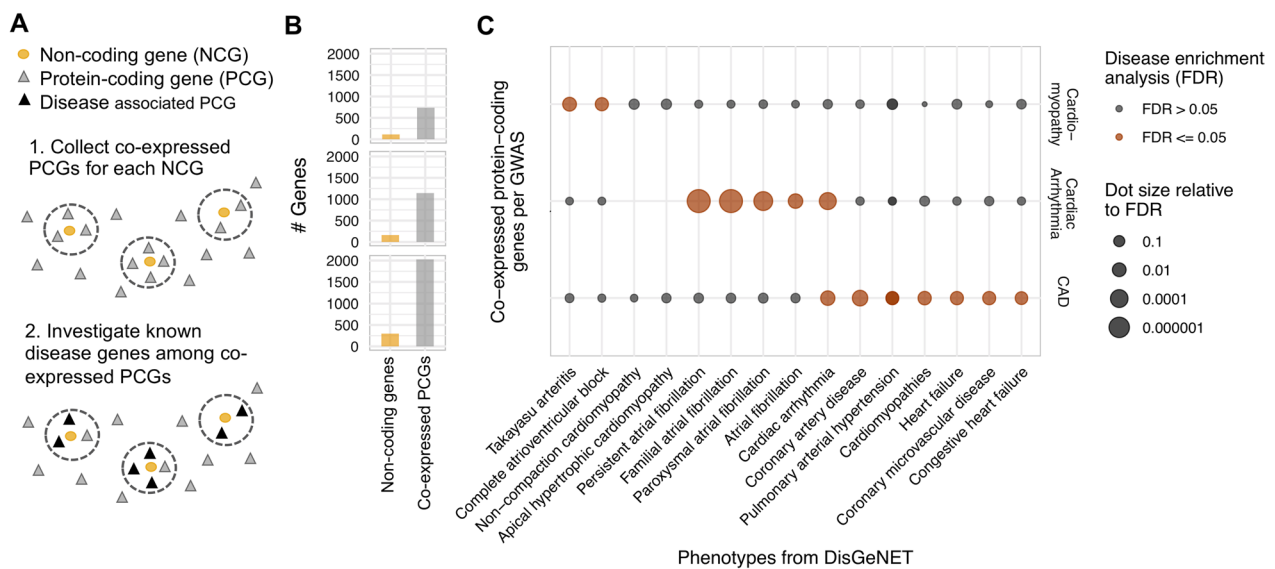


**Fig. 2** Analysis of protein and non-coding genes associated with rSNPs. **A** For each GWAS (row) a stacked bar plot shows the number of associated genes per category. **B** Dot plot visualizing per GWAS how many regulatory elements (REMs) and rSNPs are associated with a gene. Genes are separated into protein-coding, protein-coding associated with the disease according to DisGeNET, non-coding, and experimentally studied non-coding genes in this work (circle colour). The x-axis shows the number of REMs for a gene overlapping with at least one rSNP and the y-axis the number of rSNPs for all REMs associated with a gene. The size of the dot correlates with the number of genes having the same x- and y- coordinate values

genes had cases with many rSNPs and/or REMs associated per gene. Notably, identified genes had on average more rSNPs than disease genes listed in DisGeNET. This underlines a unique feature of our [40] approach, highlighting genes that have many non-coding rSNPs associated with the indication. Of particular note was the large number of non-coding genes we identified. We surveyed the literature and existing databases that list known RNA biomarkers to identify additional evidence for genes that we found in our analyses (Additional file 3: Table S2). For example, we checked the Heart Failure database for known RNA biomarkers (HFBD) [43], but none of the 49 ncRNAs listed there overlapped ours.

In speculating that it would be possible to find further evidence using existing OMICs data—we used a guilt-by association strategy and collected for each non-coding gene the top 10 most highly correlated protein-coding genes among all genes, according to a large RNA expression dataset from the GTEx resource [37] (Fig. 3A). Based on 9662 GTEx expression samples, protein-coding genes with the highest correlation for each of the non-coding genes were identified. Our rational being, that co-expressed protein-coding genes may be involved in similar pathways or biological functions than the non-coding gene. We then gathered all co-expressed protein-coding genes with respect to the non-coding genes





**Fig. 3** Disease enrichment analysis based on the protein coding genes co-expressed with GWAS-associated non-coding genes. **A** We sought to identify co-expressed protein-coding genes (PCGs, triangles) for each non-coding gene (NCG, circle) using a large dataset of RNA-seq samples from GTEx top correlated protein-coding genes are tested for an enrichment of previous indications with a cardiovascular disease. **B** Bar plot visualizing the number of non-coding and co-expressed protein-coding genes per GWAS (row). **C** Dot plot showing enriched cardiovascular phenotypes from DisGeNET (x-axis) for the top 10 co-expressed protein-coding genes of the associated non-coding genes separated per GWAS (y-axis). The dot coloring represents whether a phenotype is significantly enriched ( $FDR \leq 0.05$ ) and the dot size is relative to the FDR

we had found for each disease (Fig. 3B). This collection thus signifies all protein-coding genes that are highly correlated to the non-coding genes we found. The guilty-association principle relies on the idea that associated genes are likely to share the same function and may further be involved in the same disease. To test this idea, we conducted the DisGeNET enrichment analysis for all co-expressed protein-coding genes (Fig. 3C, Additional file 4: Table S3). We observed a strong enrichment for the expected diseases in each tested study in which we had more than 30 non-coding genes. Thus, we were positive that a majority of the non-coding genes could play an important role in the underlying disease.

#### Identification and characterization of ncRNAs in hiPSC-CMs

Our previous analyses suggested that many of the ncRNAs that we associated with CVDs could have important cardiovascular functions. Thus, we selected in total 40 ncRNA genes and prioritized ncRNA genes not listed in DisGeNET and with many rSNPs (Fig. 2B, Table 3 and Additional file 5: Table S4). We selected genes for each indication focusing on genes with many rSNPs and REMs per gene, but also taking cases with fewer rSNPs or REMs to explore the complete range of observed values for these parameters (Fig. 2B). We selected 12 ncRNA genes for CAD, 2 for Aortic stenosis, 10 for Cardiomyopathy, 8 for Cardiac arrhythmia, 7 for Myocardial infarction and one for Myocardial

ischemia. The gene KRT8P15 was additionally selected as it appeared in Myocardial infarction and CAD. For each of 38 out of 40 ncRNAs available in the GTEx expression dataset, we investigated the top 10 correlated protein-coding genes (as mentioned above). For all the 38 ncRNAs we were able to find at least one protein-coding gene among the top 10, which is connected with a CVD in the DisGeNET database (Additional file 5: Table S4). For example, the gene HLA-DQB1-AS1 has many correlated protein-coding genes from other HLA genes, for which the frequency of certain haplotypes was recently connected with heart failure [44]. Another example is the ncRNA RP11-433J22.2 which was found for Cardiac Arrhythmia and is co-expressed with two gap junction genes (GJA4 and GJA5) [40, 41].

To determine the roles of the 40 ncRNAs in human cardiomyocytes, the hiPSC derived CMs, monoculture cardiac organoids (MCOs) and self-organized cardiac organoids (SCOs) were used in the study (Fig. 4A). The hiPSC-CMs, MCOs and SCOs were incubated in normoxia and hypoxia (1%  $O_2$  or 3%  $O_2$ , respectively), and the expression levels of these 40 ncRNAs were quantified by qRT-PCR (Fig. 4B). As shown in Fig. 4B, we have selected the ncRNAs RP11-98F14.11, RPL23AP92, IGBP1P1, and CTD-2383I20.1 for further experiments, as they are consistently upregulated in hypoxia in hiPSC-CMs, MCOs and SCOs. Furthermore, the expression levels of the ncRNAs RP11-98F14.11,

**Table 3** List of ncRNAs

Gene ID	Gene name	Related human disease
ENSG00000232978	RP11-146N23.4	Aortic Stenosis
ENSG00000273226	RP11-513M16.8	Aortic Stenosis
ENSG00000274220	RP11-77K12.9	CAD
ENSG00000233471	KRT18P62	CAD
ENSG00000246889	AP000487.5	CAD
ENSG00000259999	RP11-252K23.1	CAD
ENSG00000262583	RP11-77K12.5 / TMEM231P1	CAD
ENSG00000250474	WBP1LP2	CAD
ENSG00000178803	ADORA2A-AS1	CAD
ENSG00000201151	SNORD56	CAD
ENSG00000244691	RPL10AP1	CAD
ENSG00000248278	SUMO2P17	CAD
ENSG00000258978	HIF1AP1	CAD
ENSG00000128262	POM121L9P	CAD
ENSG00000283828	RP11-98F14.12	Cardiac Arrhythmia
ENSG00000238228	OR7E7P	Cardiac Arrhythmia
ENSG00000250899	RP11-253E3.3	Cardiac Arrhythmia
ENSG00000271146	RP11-479F13.1	Cardiac Arrhythmia
ENSG00000252072	RNA5SP320	Cardiac Arrhythmia
ENSG00000269125	RP11-98F14.11	Cardiac Arrhythmia
ENSG00000270723	RPL23AP92	Cardiac Arrhythmia
ENSG00000274415	RP11-433J22.2	Cardiac Arrhythmia
ENSG00000272954	KB-1440D3.13	Cardiomyopathy
ENSG00000226677	IGBP1P1	Cardiomyopathy
ENSG00000274281	RP11-326N17.2	Cardiomyopathy
ENSG00000223534	HLA-DQB1-AS1	Cardiomyopathy
ENSG00000227409	ZMYM4-AS1	Cardiomyopathy
ENSG00000229391	HLA-DRB6	Cardiomyopathy
ENSG00000252042	Y RNA	Cardiomyopathy
ENSG00000259345	RP11-624L4.1	Cardiomyopathy
ENSG00000263786	RP11-649A18.4	Cardiomyopathy
ENSG00000275293	Metazoa SRP	Cardiomyopathy
ENSG00000249994	CTD-2383I20.1	Myocardial Infarction
ENSG00000225124	RPL23AP36	Myocardial Infarction
ENSG00000272750	RP11-378J18.8	Myocardial Infarction
ENSG00000229116	RP11-20J15.3	Myocardial Infarction
ENSG00000265215	MIR4269	Myocardial Infarction
ENSG00000266936	CTC-215O4.4	Myocardial Infarction
ENSG00000233579	KRT8P15	Myocardial Infarction & CAD
ENSG00000270001	RP11-218C14.8	Myocardial Ischemia

RPL23AP92, IGBP1P1, and CTD-2383I20.1 in human MI hearts and healthy hearts were also quantified by qRT-PCR, and our data exhibited an increased expression of RP11-98F14.11, RPL23AP92, IGBP1P1, and CTD-2383I20.1 in human MI heart compared to the healthy hearts (Ctrl) (Additional file 1: Fig. S2).

In order to investigate the function of RP11-98F14.11, RPL23AP92, IGBP1P1, and CTD-2383I20.1 in hiPSC-CMs, we performed loss-of-function through GapmeR-mediated knockdown of the respective target ncRNAs. GapmeRs are chimeric anti-sense oligonucleotides that contain a central block of deoxynucleotide monomers to induce RNaseH cleavage [45]. Four GapmeRs were designed and synthesized for each ncRNA target (Table 2), and the knockdown efficiency was quantified by qRT-PCR in hiPSC-CMs (Additional file 1: Fig. S3A-S3D). We selected GapmeRs RP11-98F14.11 #2 (GM RP11-98F14.11), GapmeR RPL23AP92 #1 (GM RPL23AP92), GapmeR IGBP1P1 #1 (GM IGBP1P1), and GapmeR CTD-2383I20.1 #1 (GM CTD-2383I20.1) for further experiments. As shown in Fig. 4C, the expression levels of pathologic hypertrophy markers, including atrial natriuretic peptide A (*NPPA*), *NPPB* and beta-myosin heavy chain 7 (*MYH7*), were reduced after GapmeR RP11-98F14.11 (GM RP11-98F14.11) treatment in hypoxia compared to the GapmeR Control (GM Ctrl) treatment. We also profiled expression of key fibrotic marker genes including Collagen type III alpha 1 chain (*COL3A1*), matrix metalloproteinase-2 (*MMP2*) and Transforming growth factor  $\beta$ 1 (*TGF $\beta$ 1*). However, GM RP11-98F14.11 treatment did not alter hypoxia-induced fibrotic marker gene expression in hiPSC-CMs compared to the GM Ctrl. Similarly, Knockdown of RPL23AP92 and IGBP1P1 reduced hypoxia-induced pathologic hypertrophy marker expression, but not of hypoxia-induced fibrotic marker genes (Fig. 4D and E). CTD-2383I20.1 did not alter either hypoxia-induced pathologic hypertrophy markers or hypoxia-induced fibrotic marker gene expression compared to GM Ctrl (Additional file 1: Fig. S4).

To better define the function of RP11-98F14.11, RPL23AP92, IGBP1P1, and CTD-2383I20.1 in human cardiomyocytes, we analyzed cardiomyocyte cell size in a loss-of-function setting. In order to directly visualize the cells, cardiomyocytes were stained for  $\alpha$ -actinin and DAPI and imaged by confocal microscopy (Fig. 4F). As shown in Fig. 4G, hypoxia led to increased cell size, which was rescued upon RP11-98F14.11, RPL23AP92 and IGBP1P1 inhibition, while CTD-2383I20.1 did not affect the cell size. Together, this suggests that RP11-98F14.11, RPL23AP92, IGBP1P1 play key roles in regulation of cell size to determine hypertrophic response in human cardiomyocytes, but not in pathology of fibrosis in 2D monolayer cultures.

#### IGBP1P1 drives pathologic hypertrophy and contractile dysfunction in human cardiac tissue mimetics

We determined the function of RP11-98F14.11, RPL23AP92, IGBP1P1 in human iPSC-derived cardiac

organoids, serving as a model for native heart tissue. On day 18 of differentiation, the monolayer was dissociated into single cells and seeded on Aggrewell™ 800 microwell culture plates in order to induce the formation of cardiac organoids (Fig. 4A). Cardiac mimetics were then cultured under control normoxic conditions, or in hypoxia (3% O<sub>2</sub>) to mimic myocardial hypoxia in vitro. As shown in Fig. 5A, hypoxia-induced pathologic hypertrophy markers and fibrotic marker genes were decreased by inhibition of RP11-98F14.11. While RPL23AP92 deletion could partially rescue the pathologic hypertrophy markers and fibrotic marker genes in hypoxia compared to GM Ctrl (Fig. 5B). Similarly, IGBP1P1 depletion also reduced the expression levels of pathologic hypertrophy markers and fibrotic marker genes in hypoxia. In addition, the cell size of cardiomyocytes in MCOs was also measured in both normoxia and hypoxia (Fig. 5C). As shown in Fig. 5D and E, inhibition of IGBP1P1 reduced the cell size in hypoxia compared to the GM Ctrl.

Furthermore, cardiac contractility was also measured by IonOptix and through calcium flux analysis. Both assays revealed that the Hypoxia-induced beating frequency was reduced by IGBP1P1 inhibition (Fig. 5F), while hypoxia-reduced contractile amplitude was rescued by IGBP1P1 inhibition compared to GM Ctrl (Fig. 5G and 5H, Additional file 1: Fig. S5A and S5B). However, there is no difference of beating frequency and contractile amplitude between GM RP11-98F14.11 and GM Ctrl (Additional file 1: Fig. S5C-S5E). Similarly, knockdown of RPL23AP92 did not rescue the hypoxia-induced beating frequency and hypoxia-reduced contractile amplitude in hypoxia compared to GM Ctrl (Additional file 1: Fig. S5F-S5H). Together, IGBP1P1 regulates both cell size and cardiomyocyte contractility in 3D human cardiac tissue mimetics.

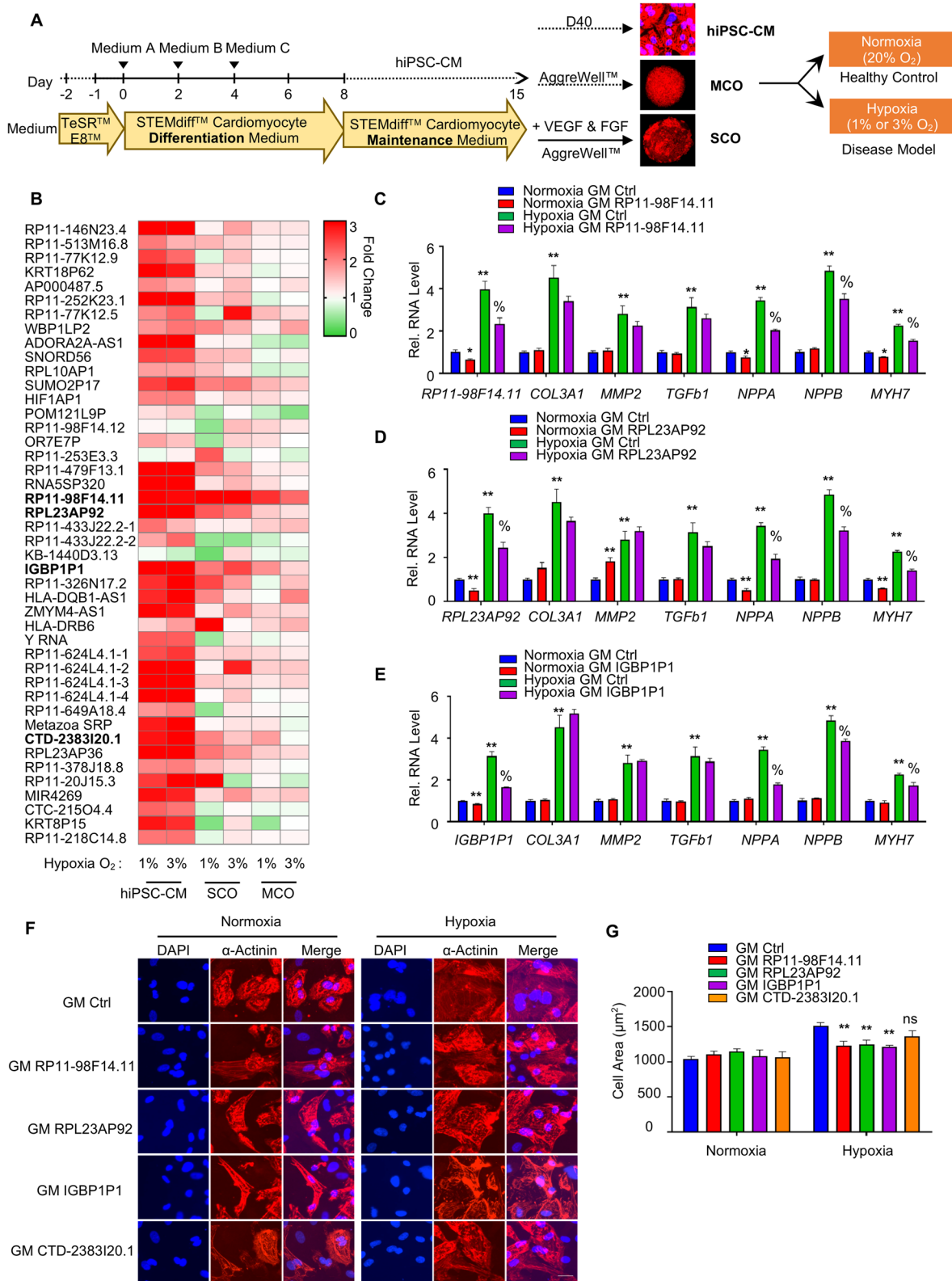
## Discussion

In this study we have used known position-weight matrix models of TF binding to predict whether SNPs have a regulatory effect. These models do not consider dependency between positions. Although alternative models exist, such as SLIM [46], the absolute number of available TFs is lower than what we used here from the JASPAR database. Exploring more complex models may increase the number of rSNPs that can be detected and thus may reveal additional genes of interest. Compared to using all SNPs that are significant according to a GWAS, we limit relevant gene associations to rSNPs, to enrich for variants that are involved in transcriptional regulation and therefore should reside in regulatory elements of their target genes. Instead of predicting rSNPs after conducting the GWAS, Arloth et al. have first determined rSNPs to filter [26]. They then recomputed the association significance for a smaller subset of genome-wide rSNPs, which may further boost the ability to detect disease genes.

The conducted analysis workflow is not limited to applications to CVD GWAS and deviates from other studies that use GWAS SNPs for the identification of disease genes. Common approaches for linking SNPs to target genes include using expression quantitative trait loci, or high-throughput chromosome capture (Hi-C) data that measures DNA-DNA contacts [47]. Both types of data need to exist for the cell types that are involved in the disease, which is challenging given the large number of cell types affected by CVDs. We rely on enhancer-gene associations learned from paired DNase1-seq and RNA-seq data [30]. A recent generalisation of the Activity-by-Contact model was shown to improve accuracy of predicted enhancer-gene interactions from epigenomics data [48] and could be used to increase considered enhancer-gene interactions specific for CVD-relevant cell types.

(See figure on next page.)

**Fig. 4** Identification and characterization of ncRNAs in hiPSC-CMs **A** Schematic protocol for differentiation of hiPS cells into cardiomyocytes, monoculture cardiac organoids (MCOs) and self-organized cardiac organoids (SCOs) in vitro. **B** Heat map shows the relative mRNA expression levels of the 40 ncRNAs in the 40-day-old hiPSC-CMs, or 40-day-old human cardiac organoids (including MCOs and SCOs) treated with either 1% O<sub>2</sub> or 3% O<sub>2</sub> for 2 or 3 days, respectively. Data are normalized to either normoxia hiPSC-CMs, or normoxia MCOs/SCOs. The red and green colors indicate high and low expression values, respectively. Means of  $n=3$  biological replicates per group. **C** Relative mRNA expression of *RP11-98F14.11*, *COL3A1*, *MMP2*, *TGFb1*, *NPPA*, *NPPB* and *MYH7* in hiPSC-CMs transduced with either GapmeR RP11-98F14.11 (GM RP11-98F14.11) or GapmeR negative control (GM Ctrl) in both normoxia and hypoxia. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \* $P<0.05$ , \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **D** Relative mRNA expression of *RPL23AP9*, *COL3A1*, *MMP2*, *TGFb1*, *NPPA*, *NPPB* and *MYH7* in hiPSC-CMs transduced with either GapmeR RPL23AP9 (GM RPL23AP9) or GM Ctrl in both normoxia and hypoxia. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \* $P<0.05$ , \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **E** Relative mRNA expression of *IGBP1P1*, *COL3A1*, *MMP2*, *TGFb1*, *NPPA*, *NPPB* and *MYH7* in hiPSC-CMs transduced with either GapmeR IGBP1P1 (GM IGBP1P1) or GM Ctrl in both normoxia and hypoxia. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \* $P<0.05$ , \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **F** Representative images of  $\alpha$ -actinin (red) and DAPI (blue) in 40-day-old hiPSC-CMs after GM RP11-98F14.11, GM RPL23AP9, GM IGBP1P1, GM CTD-2383120.1, or the GM Ctrl under normoxia or hypoxia for 2 days. Scale bar is 25  $\mu$ m. **G** Cell size of the hiPSC-CMs was assessed by the Image J. 50 cells were analyzed for each condition. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \*\* $P<0.01$  vs. Hypoxia GM Ctrl. Two-tailed unpaired t-test. hiPSC-CMs, human-induced pluripotent stem cell-derived cardiomyocytes; MCOs, monoculture cardiac organoids; SCOs: Self-organized cardiac organoids



**Fig. 4** (See legend on previous page.)

Pseudogenes are defined as regions of the genome that contain defective copies of genes and are often considered as nonfunctional ncRNA. However, recent studies have shown that pseudogenes may play important roles in CVDs [49]. Elevated low-density lipoprotein cholesterol (LDL-C) level is a main risk factor for CVDs, while knockdown of zinc finger protein 542 pseudogene (ZNF542P) increases the LDL-C level response to simvastatin in a human hepatoma cell line [50]. The mRNA levels of octamer-binding transcription factor 4 (Oct 4) pseudogene Oct-4-psG1 and Oct-4-psG5 are significantly down-regulated in pulmonary arterial smooth muscle cells (PASMC) in patients with idiopathic pulmonary arterial hypertension (IPAH), indicating that Oct-4-psG1 and Oct-4-psG5 are involved in IPAH [51]. Moreover, expression level of NMRA-like protein NMRAL1 pseudogene (NMRAL2P) is significantly decreased in the right ventricle in heart failure, suggesting that NMRAL2P is involved in heart failure [52]. Together, these studies indicate that pseudogenes are closely related to CVDs. IGBP1P1 is a pseudogene of IGBP1, a phosphoprotein associated with the B cell receptor complex and leads to multiple signal transduction pathways [53]. IGBP1 is a novel biomarker in lupus nephritis (LN) patients, its expression level is increased in the plasma and urine of patients with LN compared with systemic lupus erythematosus (SLE) patients without nephritis and healthy controls [54, 55]. Recently it also showed that IGBP1 is upregulated in esophageal squamous cell carcinoma (ESCC), and its expression is significantly associated with ESCC patient survival [56]. IGBP1 is also expressed in the heart, but its function has not been studied in the heart yet. In our study, we have identified that IGBP1P1 is upregulated in the human disease model in vitro, and its depletion could reduce hypoxia-induced cell size and improve cardiac contractility in human cardiac tissue mimetics. It may modulate cardiomyocyte size and cardiac contractility through regulating the expression level

of IGBP1 at both transcriptional and translational levels, but the mechanism is not known yet.

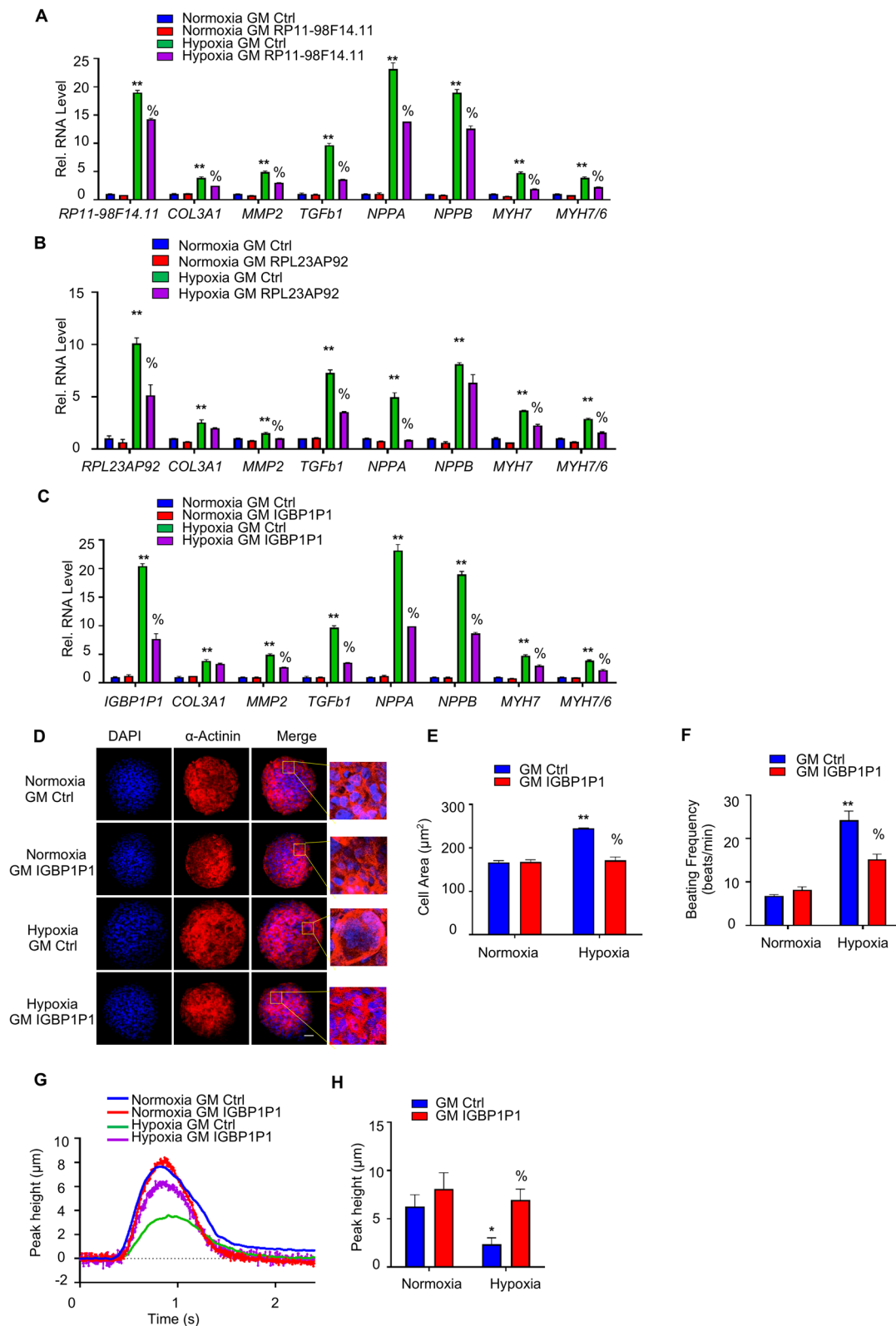
Many studies have demonstrated that ncRNAs play important roles in the development of CVDs [57]. However, the study of human-specific ncRNAs has been limiting and challenging. The majority of the human lncRNAs are poorly conserved in mouse, conventional mouse models are not a suitable tool to study their function in vivo regulation and function. Here, we utilized a human 3D cardiac organoid model, in order to best recapitulate the biological and molecular properties of native heart tissue thus enabling us to study ncRNA function in a physiologically relevant context-lending greater credence to the validity of our study and its findings. The 3D human cardiac organoids are composed of different cell types including cardiomyocytes, endothelial cells, and fibroblasts, which are able to self-organize into complex organ-like structures and have a similar microenvironment to the human heart. In addition, the human cardiac organoids can be cultured for longer term in vitro, and also display molecular, metabolic and contractile characteristics of adult native myocardium and respond to pathophysiologic stressors (Fig. 5G and H, Additional file 1: Fig. S5A and S5B). The organoid model is a useful biological tool to study the biological functions of ncRNAs in our study. In the future, human cardiac organoids will provide useful for research in disease modeling, developmental biology, and drug screening.

## Conclusion

Taken together, we have identified 40 CVD-associated non-coding RNAs by using a computational pipeline that integrates GWAS, TF motif and enhancer-gene information. Then, we have demonstrated that the ncRNA IGBP1P1, is a pathologic stress-induced modulator of cardiomyocyte hypertrophy and contractile function in 2D hiPSC-CMs and 3D human cardiac organoids.

(See figure on next page.)

**Fig. 5** IGBP1P1 is vital in cell size and contractility in human cardiac tissue mimetics. **A** Relative mRNA expression of *RP11-98F14.11*, *COL3A1*, *MMP2*, *TGFb1*, *NPPA*, *NPPB*, *MYH7* and ratio *MYH7/6* in MCOs transduced with either GM *RP11-98F14.11* or GM Ctrl in both normoxia and hypoxia. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \* $P<0.05$ , \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **B** Relative mRNA expression of *RPL23AP92*, *COL3A1*, *MMP2*, *TGFb1*, *NPPA*, *NPPB*, *MYH7* and ratio *MYH7/6* were detected. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \* $P<0.05$ , \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **C** Relative mRNA expression of *IGBP1P1*, *COL3A1*, *MMP2*, *TGFb1*, *NPPA*, *NPPB*, *MYH7* and ratio *MYH7/6* in MCOs transduced with either GM *IGBP1P1* or GM Ctrl in both normoxia and hypoxia. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \* $P<0.05$ , \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **D** Representative images of  $\alpha$ -actinin (red) and DAPI (blue) in 40-day-old human MCOs after GM *IGBP1P1* or the GM Ctrl in normoxia or hypoxia for 3 days. Scale bar is 50  $\mu$ m. **E** Cardiomyocyte cell size in the MCOs was assessed by the Image J. 50 cells were analyzed for each condition. Data are represented as Mean  $\pm$  SEM;  $n=3$ ; \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **F** Human MCOs were treated with GM *IGBP1P1* or GM Ctrl in both normoxia and hypoxia for 3 days, and then the frequency of CMO contraction was determined by counting beats per minute. Data are represented as Mean  $\pm$  SEM; 4–5 organoids per group; \*\* $P<0.01$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. **G** Representative traces of contractile MCOs. **H** The contractility assays were performed by determining the amplitude peak of contracting MCOs. Data are represented as Mean  $\pm$  SEM; 4–5 organoids per group; \* $P<0.05$  vs. Normoxia GM Ctrl, % $P<0.05$  vs. Hypoxia GM Ctrl. All by one-way ANOVA analyses followed by a Dunnett's multiple comparison post-test. MCOs, monoculture cardiac organoids



**Fig. 5** (See legend on previous page.)

IGBP1P1 depletion rescued cardiomyocyte size and improved cardiac contractility. Thus, blocking the ncRNA IGBP1P1 could be a promising strategy to improve cardiac function in cardiovascular disease.

#### Abbreviations

CVDs	Cardiovascular diseases
GWAS	Genome-wide association studies
SNPs	Single nucleotide polymorphisms
TF	Transcription factor
ncRNA	Non-coding RNA
hiPSC-CMs	Human pluripotent stem cells derived cardiomyocytes
MCO	Monoculture cardiac organoid
SCO	Self-organized cardiac organoid
MMP-2	Matrix metalloproteinase-2
PWMs	Position weight matrices
REMs	Regulatory elements
IGBP1P1	Immunoglobulin (CD79A) binding protein 1 pseudogene 1
LD	Linkage disequilibrium
SNEEP	SNP exploration and analysis using epigenomics data
ZNF542P	Zinc finger protein 542 pseudogene
LN	Lupus nephritis
SLE	Systemic lupus erythematosus
NMRAL2P	NMRA-like protein NMRAL1 pseudogene
CAD	Coronary artery disease
EBs	Embryoid bodies
HFBD	Heart Failure database for known RNA biomarkers
LDL-C	Low-density lipoprotein cholesterol
PASMC	Pulmonary arterial smooth muscle cell
ESCC	Esophageal squamous cell carcinoma

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-023-00513-4>.

**Additional file 1: Fig. S1.** Disease enrichment analysis for protein-coding genes. Bar plots representing per GWAS selected phenotypes enriched for protein coding genes identified with SNEEP. The x-axis shows the  $-\log_{10}$  FDR corrected  $P$ -value of the disease enrichment analysis performed with the DisGeNET software (usage of DisGeNET similar to method section 'Identification of disease associated genes using rSNPs', as input the protein-coding genes from the SNEEP result per GWAS are taken) (see also Additional file 2: Table S1). For the GWAS Myocardial ischemia and Aortic stenosis the disease enrichment analysis was not possible, because only 5 and 12 protein coding genes were associated. **Fig. S2.** Expression of ncRNAs in human MI hearts. Relative RNA expression level of ncRNAs RP11-98F14.11, RPL23AP92, IGBP1P1, and CTD-2383I20.1 in human MI hearts compared to healthy adult hearts (Ctrl). Data are represented as Mean  $\pm$  SEM; Ctrl:  $n = 4$ ; MI:  $n \geq 9$ ;  $*P < 0.05$  vs. Ctrl. Two-tailed unpaired t-test. MI: Myocardial infarction. **Fig. S3.** Identification of the best GapmeRs in hiPSC-CMs. **A** Relative RNA expression level of *RP11-98F14.11* in hiPSC-CMs treated with GM *RP11-98F14.11* #1, #2, #3, #4 or GM GapmeR Ctrl. Data are normalized to hiPSC-CMs expressing GM Ctrl. Data are represented as Mean  $\pm$  SEM;  $n = 3$ ;  $*P < 0.05$  vs. GM Ctrl. **B** Relative RNA expression level of *RPL23AP92* in hiPSC-CMs treated with GM *RPL23AP92* #1, #2, #3, #4 or GM Ctrl. Data are represented as Mean  $\pm$  SEM;  $n = 3$ ;  $*P < 0.05$ ,  $**P < 0.01$  vs. GM Ctrl. **C** Relative RNA expression level of *IGBP1P1* in hiPSC-CMs treated with GM *IGBP1P1* #1, #2, #3, #4 or GM Ctrl. Data are represented as Mean  $\pm$  SEM;  $n = 3$ ;  $*P < 0.05$ ,  $**P < 0.01$  vs. GM Ctrl. **D** Relative RNA expression level of *CTD-2383I20.1* in hiPSC-CMs treated with GM *CTD-2383I20.1* #1, #2, #3, #4 or GM Ctrl. Data are represented as Mean  $\pm$  SEM;  $n = 3$ ;  $*P < 0.05$  vs. GM Ctrl. Two-tailed unpaired t-test. **Fig. S4.** Characterization of *CTD-2383I20.1* in hiPSC-CMs. Relative RNA expression of *CTD-2383I20.1*, *COL3A1*, *MMP2*, *TGFb1*, *NPPA*, *NPPB* and *MYH7* in hiPSC-CMs transduced with either GM *CTD-2383I20.1* or GM Ctrl in both normoxia and hypoxia. Data are represented as Mean  $\pm$  SEM;  $n = 3$ ;

$*P < 0.05$ ,  $**P < 0.01$  vs. Normoxia GM Ctrl,  $\%P < 0.05$  vs. Hypoxia GM Ctrl. Two-tailed unpaired t-test. **Fig. S5.** IGBP1P1 inactivation improved contractility in human MCOs. Human MCOs were treated with GM IGBP1P1 or GM Ctrl in both normoxia and hypoxia for 3 days, and then the contractility assays were performed by calcium transient. **A** Representative traces of contractile MCOs. **B** Data are represented as Mean  $\pm$  SEM;  $n > 5$  organoids per group;  $**P < 0.01$  vs. Normoxia GM Ctrl,  $\%P < 0.05$  vs. Hypoxia GM Ctrl. Two-tailed unpaired t-test. **C** Human MCOs were treated with GM *RP11-98F14.11* or GM Ctrl in both normoxia and hypoxia for 3 days, and then the frequency of CMO contraction was determined by counting beats per minute. Data are expressed as means  $\pm$  SEM.  $n > 5$  organoids per group;  $**P < 0.01$  vs. Normoxia GM Ctrl, ns vs. Hypoxia GM Ctrl. Two-tailed unpaired t-test. **D** Representative traces of contractile MCOs. **E** The contractility assays were performed by calcium transient. Data are expressed as means  $\pm$  SEM.  $n > 5$  organoids per group;  $**P < 0.01$  vs. Normoxia GM Ctrl, ns vs. Hypoxia GM Ctrl. Two-tailed unpaired t-test. **F** The frequency of CMO contraction was determined by counting beats per minute. Data are represented as Mean  $\pm$  SEM; 4–5 organoids per group;  $**P < 0.01$  vs. Normoxia GM Ctrl,  $\%P < 0.05$  vs. Hypoxia GM Ctrl. Two-tailed unpaired t-test. **G** Representative traces of contractile MCOs. **H** The contractility assays were performed by determining the amplitude peak of contracting MCOs. Data are represented as Mean  $\pm$  SEM; 4–5 organoids per group;  $*P < 0.05$  vs. Normoxia GM Ctrl,  $\%P < 0.05$  vs. Hypoxia GM Ctrl. Two-tailed unpaired t-test.

**Additional file 2: Table S1.** Disease Enrichment Protein Coding.

**Additional file 3: Table S2.** Associated Genes.

**Additional file 4: Table S3.** Disease Enrichment Noncoding.

**Additional file 5: Table S4.** Tested Noncoding RNAs.

**Additional file 6: Table S5.** SNPs Per GWAS.

**Additional file 7: Table S6.** Result SNEEP Per GWAS.

**Additional file 8: Table S7.** Co-expressed PCG.

#### Acknowledgements

We thank Silke Kauferstein (Goethe University Hospital, Frankfurt am Main, Germany) for providing human heart biopsies.

#### Author contributions

C.Z., N.B., T.Y., M.S. and J.K. planned and designed the experiments and wrote the paper. N.B., F.B.A. and M.S. performed the bioinformatics analysis. C.Z. executed most of the experiments with help from M.W., Y.W., A.P.D., M.D. and S.D.. J.K., T.T.D. and M.D.P. provided cells and material for 2D and 3D experiments.

#### Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by SFB-TRR 267 (Non-coding RNA in the cardiovascular system) and the German Research Foundation (Exc2026) to J.K., M.S. and S.D. and the European Innovation Council (GA: 822455) to J.K., the LOEWE Center for Cell and Gene Therapy to S.D. and J.K., the European Research Council (Angiolnc) to S.D.. C.Z. and Y.W. were supported by the China Scholarship Council (CSC) Grant #202008080200 and #202108080020.

#### Declarations

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Institute for Cardiovascular Regeneration, Goethe University, 60590 Frankfurt Am Main, Germany. <sup>2</sup>German Center for Cardiovascular Research, Partner Site Rhein-Main, 60590 Frankfurt Am Main, Germany. <sup>3</sup>Cardio-Pulmonary Institute, Goethe University Hospital, 60590 Frankfurt Am Main, Germany. <sup>4</sup>Department of Medicine III, Cardiology/Angiology/ Nephrology, Goethe University Hospital, Frankfurt, Germany. <sup>5</sup>Genome Biologics, Theodor-Stern-Kai 7, 60590 Frankfurt Am Main, Germany.

Received: 26 February 2023 Accepted: 12 July 2023  
Published online: 25 July 2023

## References

- Das S, Shah R, Dimmeler S, Freedman JE, Holley C, Lee J-M, et al. Noncoding RNAs in cardiovascular disease: current knowledge, tools and technologies for investigation, and future directions: a scientific statement from the American Heart Association. *Circ Genom Precis Med*. 2020;13:e000062.
- Josefs T, Boon RA. The long non-coding road to atherosclerosis. *Curr Atheroscler Rep*. 2020;22:55.
- Juni RP, Hart KC, Houtkooper RH, Boon RA. Long noncoding RNAs in cardiometabolic disorders. *FEBS Lett*. 2022;596:1367–87.
- Mirtschink P, Bischof C, Pham M-D, Sharma R, Khadayate S, Rossi G, et al. Inhibition of the hypoxia-inducible factor 1 $\alpha$ -induced cardioprotective HERN1 enhance-templated RNA protects from heart disease. *Circulation*. 2019;139:2778–92.
- Ounzain S, Micheletti R, Arnan C, Plaisance I, Cecchi D, Schroen B, et al. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *J Mol Cell Cardiol*. 2015;89:98–112.
- Piccoli M-T, Gupta SK, Viereck J, Foinquinos A, Samolovac S, Kramer FL, et al. Inhibition of the cardiac fibroblast-enriched lncRNA Meg3 prevents cardiac fibrosis and diastolic dysfunction. *Circ Res*. 2017;121:575–83.
- Fasolo F, Jin H, Winski G, Chernogubova E, Pauli J, Winter H, et al. Long noncoding RNA *MIAT* controls advanced atherosclerotic lesion formation and plaque destabilization. *Circulation*. 2021;144:1567–83.
- Stanicek L, Lozano-Vidal N, Bink DJ, Hooglugt A, Yao W, Wittig I, et al. Long non-coding RNA *LASSIE* regulates shear stress sensing and endothelial barrier function. *Commun Biol*. 2020;3:265.
- Li L, Chen Z, von Scheidt M, Li S, Steiner A, Güldener U, et al. Transcriptome-wide association study of coronary artery disease identifies novel susceptibility genes. *Basic Res Cardiol*. 2022;117:6.
- Anene-Nzulu CG, Tan WLW, Lee CJM, Wenhao Z, Perrin A, Dashi A, et al. Assigning distal genomic enhancers to cardiac disease-causing genes. *Circulation*. 2020;142:910–2.
- Kirchler M, Konigorski S, Norden M, Meltendorf C, Kloft M, Schurmann C, et al. TransferGWAS: GWAS of images using deep transfer learning. *Bioinformatics*. 2022;38:3621–8.
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790–7.
- Amlie-Wolf A, Tang M, Mlynarski EE, Kuksa PP, Valladares O, Katanic Z, et al. INFERNO: inferring the molecular mechanisms of noncoding genetic variants. *Nucleic Acids Res*. 2018;46:8740–53.
- Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27:1017–8.
- Schmidt F, Kern F, Ebert P, Baumgarten N, Schulz MH. TEPI2—an extended framework for transcription factor binding prediction and integrative epigenomic analysis. *Bioinformatics*. 2019;35:1608–9.
- Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics*. 2021;37:2112–20.
- Avsec Ž, Weilert M, Shrikumar A, Krueger S, Alexandari A, Dalal K, et al. Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat Genet*. 2021;53:354–66.
- Macintyre G, Bailey J, Haviv I, Kowalczyk A. is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinforma Oxf Engl*. 2010;26:i524–530.
- Zuo C, Shin S, Keleş S. atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinforma Oxf Engl*. 2015;31:3353–5.
- Baumgarten N, Rumpf L, Kessler T, Schulz MH. A statistical approach to identify regulatory DNA variations [Internet]. (2023) <https://doi.org/10.1101/2023.01.31.526404>
- Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods*. 2015;12:931–4.
- Chen L, Wang Y, Zhao F. Exploiting deep transfer learning for the prediction of functional non-coding variants using genomic sequence. *Bioinformatics*. 2022;38:3164–72.
- Wang Y, Chen L. DeepPerVar: a multi-modal deep learning framework for functional interpretation of genetic variants in personal genome. *Bioinformatics*. 2022;5:btac696.
- Yao Q, Ferragina P, Reshef Y, Lettre G, Bauer DE, Pinello L. Motif-Raptor: a cell type-specific and transcription factor centric approach for post-GWAS prioritization of causal regulators. *Bioinformatics*. 2021;8:btab072.
- Yang H, Chen R, Wang Q, Wei Q, Ji Y, Zhong X, et al. TVAR: assessing tissue-specific functional effects of non-coding variants with deep learning. *Bioinformatics*. 2022;38:4697–704.
- Arloth J, Eraslan G, Andlauer TFM, Martins J, Iurato S, Kühnel B, et al. DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. *PLOS Comput Biol*. 2020;16:e1007616.
- Scherer M, Schmidt F, Lazareva O, Walter J, Baumbach J, Schulz MH, et al. Machine learning for deciphering cell heterogeneity and gene regulation. *Nat Comput Sci*. 2021;1:183–91.
- Dey KK, Gazal S, van de Geijn B, Kim SS, Nasser J, Engreitt JM, et al. SNP-to-gene linking strategies reveal contributions of enhancer-related and candidate master-regulator genes to autoimmune disease. *Cell Genomics*. 2022;2:100145.
- Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. 2019;51:1664–9.
- Schmidt F, Marx A, Baumgarten N, Hebel M, Wegner M, Kaulich M, et al. Integrative analysis of epigenetics data identifies gene-specific regulatory elements. *Nucleic Acids Res*. 2021;49:10397–418.
- Baumgarten N, Hecker D, Karunanithi S, Schmidt F, List M, Schulz MH. EpiRegio: analysis and retrieval of regulatory elements linked to genes. *Nucleic Acids Res*. 2020;48:W193–9.
- Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005–12.
- Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555–7.
- Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29:308–11.
- Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2019;5:gkz1001.
- Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. *Nucleic Acids Res*. 2021;49:D916–23.
- GTEX Consortium, Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet*. 2018;50:956–67.
- Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*. 2019;5:gkz1021.
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47:W191–8.
- Guo Y-H, Yang Y-Q. Atrial fibrillation: focus on myocardial connexins and gap junctions. *Biology*. 2022;11:489.
- Bosada FM, Van Duijvenboden K, Giouvou AE, Rivaud MR, Uhm J-S, Verkerk AO, et al. An atrial fibrillation-associated regulatory region modulates cardiac *Tbx5* levels and arrhythmia susceptibility. *eLife*. 2023;12:580317.
- Kanbay M, Wolf M, Selcoki Y, Solak Y, İkizel M, Uysal S, et al. Association of serum calcitonin with coronary artery disease in individuals with and without chronic kidney disease. *Int Urol Nephrol*. 2012;44:1169–75.
- He H, Shi M, Lin Y, Zhan C, Wu R, Bi C, et al. HFBD: a biomarker knowledge database for heart failure heterogeneity and personalized applications. *Bioinformatics*. 2021;37:4534–9.
- Roura S, Rudilla F, Gastelurrutia P, Enrich E, Campos E, Lupón J, et al. Determination of HLA-A, -B, -C, -DRB1 and -DQB1 allele and haplotype frequencies in heart failure patients. *ESC Heart Fail*. 2019;6:388–95.
- Wan WB, Migawa MT, Vasquez G, Murray HM, Nichols JG, Gaus H, et al. Synthesis, biophysical properties and biological activity of second



- generation antisense oligonucleotides containing chiral phosphorothioate linkages. *Nucleic Acids Res.* 2014;42:13456–68.
46. Keilwagen J, Grau J. Varying levels of complexity in transcription factor binding motifs. *Nucleic Acids Res.* 2015;43:e119–e119.
  47. Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nat Rev Methods Primer.* 2021;1:59.
  48. Hecker D, Behjati Ardakani F, Karollus A, Gagneur J, Schulz MH. The adapted Activity-By-Contact model for enhancer–gene assignment and its application to single-cell data. *Bioinformatics.* 2023;39:btad062.
  49. Qi Y, Wang X, Li W, Chen D, Meng H, An S. Pseudogenes in cardiovascular disease. *Front Mol Biosci.* 2020;7:622540.
  50. Kim K, Theusch E, Kuang Y-L, Dose A, Mitchel K, Cubitt C, et al. ZNF542P is a pseudogene associated with LDL response to simvastatin treatment. *Sci Rep.* 2018;8:12443.
  51. Firth AL, Yao W, Remillard CV, Ogawa A, Yuan JX-J. Upregulation of Oct-4 isoforms in pulmonary artery smooth muscle cells from patients with pulmonary arterial hypertension. *Am J Physiol Lung Cell Mol Physiol.* 2010;298:5548–57.
  52. Garciandia A, Suarez T. The NMRA/NMRAL1 homologue PadA modulates the expression of extracellular cAMP relay genes during aggregation in *Dictyostelium discoideum*. *Dev Biol.* 2013;381:411–22.
  53. Kuwahara K, Matsuo T, Nomura J, Igarashi H, Kimoto M, Inui S, et al. Identification of a 52-kDa molecule (p52) coprecipitated with the Ig receptor-related MB-1 protein that is inducibly phosphorylated by the stimulation with phorbol myristate acetate. *J Immunol Baltim Md.* 1950;1994(152):2742–52.
  54. Lee E-J, Kwon OC, Ghang B, Lim D-H, Kim DH, Hong S, et al. Immunoglobulin binding protein 1 as a potential urine biomarker in patients with lupus nephritis. *Int J Mol Sci.* 2019;20:2606.
  55. Kwon OC, Lee E-J, Oh JS, Hong S, Lee C-K, Yoo B, et al. Plasma immunoglobulin binding protein 1 as a predictor of development of lupus nephritis. *Lupus.* 2020;29:547–53.
  56. Jiang S, Li D, Liang Z, Wang Y, Pei X, Tang J. High expression of IGBP1 correlates with poor prognosis in esophageal squamous cell carcinoma. *Int J Biol Markers.* 2020;35:33–40.
  57. Sallam T, Sandhu J, Tontonoz P. Long noncoding RNA discovery in cardiovascular disease: decoding form to function. *Circ Res.* 2018;122:155–66.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

