WWP2 regulates pathological cardiac fibrosis by modulating SMAD2 signaling

Huimei Chen1,22, Aida Moreno-Moral1,22, Francesco Pesce2, Nithya Devapragash1, Massimiliano Mancini3, Ee Ling Heng4, Maxime Rotival5, Prashant K. Srivastava6, Nathan Harmston1, Kirill Shkura6, Owen J.L. Rackham1, Wei-Ping Yu7,8, Xi-Ming Sun9, Nicole Gui Zhen Tee10, Elisabeth Li Sa Tan1, Paul J.R. Barton4,11, Leanne E. Felkin4,11, Enrique Lara-Pezzi12, Gianni Angelini4,13, Cristina Beltrami4, Michal Pravenec14, Sebastian Schafer1,10, Leonardo Bottolo15,16,17, Norbert Hubner18,19,20,21, Costanza Emanueli4,11, Stuart A. Cook1,9,10 & Enrico Petretto1,9

Cardiac fibrosis is a final common pathology in inherited and acquired heart diseases that causes cardiac electrical and pump failure. Here, we use systems genetics to identify a pro-fibrotic gene network in the diseased heart and show that this network is regulated by the E3 ubiquitin ligase WWP2, specifically by the WWP2-N terminal isoform. Importantly, the WWP2-regulated pro-fibrotic gene network is conserved across different cardiac diseases characterized by fibrosis: human and murine dilated cardiomyopathy and repaired tetralogy of Fallot. Transgenic mice lacking the N-terminal region of the WWP2 protein show improved cardiac function and reduced myocardial fibrosis in response to pressure overload or myocardial infarction. In primary cardiac fibroblasts, WWP2 positively regulates the expression of pro-fibrotic markers and extracellular matrix genes. TGFβ1 stimulation promotes nuclear translocation of the WWP2 isoforms containing the N-terminal region and their interaction with SMAD2. WWP2 mediates the TGFβ1-induced nucleocytoplasmic shuttling and transcriptional activity of SMAD2.

1 Programme in Cardiovascular and Metabolic Disorders, Duke-NUS Medical School, Singapore 169857, Republic of Singapore. 2 Department of Emergency and Organ Transplantation (DETO), University of Bari, 70124 Bari, Italy. 3 SOC di Anatomia Patologica, Ospedale San Giovanni di Dio, 50123 Florence, Italy. 4 National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK. 5 Unit of Human Evolutionary Genetics, Institute Pasteur, 75015 Paris, France. 6 Division of Brain Sciences, Imperial College Faculty of Medicine, London W12 ONN, UK. 7 Animal Gene Editing Laboratory, BRC, A*STAR Biopolis Way, Singapore 138668, Republic of Singapore. 8 Institute of Molecular and Cell Biology, A*STAR, 61 Biopolis Drive, Singapore 138673, Republic of Singapore. 9 MRC London Institute of Medical Sciences (LICM), Imperial College, London W12 ONN, UK. 10 National Heart Centre Singapore, Singapore 169609, Republic of Singapore. 11 Cardiovascular Research Centre, Royal Brompton and Harefield NHS Trust, London SW3 6NP, UK. 12 Centro Nacional de Investigaciones Cardiovasculares – CNIC, 28029 Madrid, Spain. 13 Bristol Heart Institute, Bristol Medical School, University of Bristol, Bristol BS2 8HW, UK. 14 Institute of Physiology, Czech Academy of Sciences, 142 08 Praha 4, Czech Republic. 15 Department of Medical Genetics, University of Cambridge, Cambridge CB2 0QQ, UK. 16 The Alan Turing Institute, London NW1 2DB, UK. 17 MRC Biostatistics Unit, University of Cambridge, Cambridge CB2 0SR, UK. 18 Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany. 19 DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, 13347 Berlin, Germany. 20 Charité-Universitätsmedizin, 10117 Berlin, Germany. 21 Berlin Institute of Health (BIH), 10178 Berlin, Germany. 22 These authors contributed equally: Huimei Chen, Aida Moreno-Moral. Correspondence and requests for materials should be addressed to E.P. (email: enrico.petretto@duke-nus.edu.sg)
Pathological cardiac fibrosis is a process characterized by an excessive deposition of extracellular matrix (ECM) which in turn can lead to the development of cardiac dysfunction, arrhythmias, and heart failure (HF). Prolonged and extensive cardiac fibrosis is associated with poor clinical outcome and is an independent predictor of sudden cardiac death and overall mortality. Cardiac fibrosis can have several and sometimes concurring triggers, including systemic hypertension, diabetes, and native cardiac ischemic insults causing ischemic cardiomyopathies and other forms of dilated cardiomyopathy (DCM). Despite diverse etiologies (e.g., ischemic vs non-ischemic origin), different heart conditions such as DCM and hypertrophic cardiomyopathy develop myocardial fibrosis and may share common pathogenic pathways that lead to the development of HF.

The fibrotic remodeling of the myocardium is also a feature of congenital heart diseases and their surgical repair, such as repaired tetralogy of Fallot (rTOF) or in other conditions causing progressive right ventricular dilation and failure. Therefore, there is a growing body of evidences supporting cardiac fibrosis as a pathophysiological pathway common to different heart diseases.

The cardiac fibroblast is a key effector cell in cardiac fibrosis, responsible for homeostasis of the ECM in the heart. Upon triggering of the fibrogenic response, activated cardiac fibroblasts ultimately differentiate into myofibroblasts, increase the synthesis of ACTA2 and other ECM components, and reduce their proliferation rate. At the molecular level, the development of tissue fibrosis comprises a complex signaling cascade, for which many regulators have been proposed such as angiotensin II (AngII), connective tissue growth factor (CTGF), bone morphogenetic protein (BMP), WNT, and cytokines such as interleukin 11 (IL-11), and the transforming growth factor beta (TGFβ) superfamily. In particular, TGFβ1 binds directly to receptors for signal transduction via downstream effector proteins known as SMADs, which leads to the transcription of ECM proteins.

Therapeutic targeting of cardiac fibrosis has been proposed for alleviating the progression of cardiovascular diseases and improving cardiac function. Systemic and local delivery of drugs modulating TGFβ, endothelin-1, AngII, CCN2, and PDGF are examples of anti-fibrotic therapies. Moreover, targeting multiple pro-fibrotic pathways may provide an additional therapeutic approach to control cardiac fibrosis. Since the physiological fibrosis process is required for normal wound healing and tissue repair, the identification of targets that regulate specifically cardiac fibrosis under pathological conditions is important to develop new therapies to improve clinical outcomes for patients with heart disease.

Here, using systems genetics in the fibrotic heart we identify WWP2, a E3 ubiquitin protein ligase, as a positive genetic regulator of a transcriptional ECM gene network, which is associated with pathological cardiac fibrosis. We demonstrate a previously unappreciated role for WWP2 in the pathophysiological cardiovascular system, and elucidate how this gene regulates SMAD function and the downstream fibrogenic response in the diseased heart. These findings suggest WWP2 as a potential target for treating heart diseases.

Results

Coordinated regulation of ECM genes in the diseased heart.

We set out to identify transcriptional programs conserved across species and associated to cardiac fibrosis. We first used a panel of 30 rat recombinant inbred (RI) strains, which allows integrative analyses of cardiac gene expression with quantitative pathophysiological traits (e.g., cardiac fibrosis), and genome-wide genetic data. This RI strains panel is an established model of cardiovascular traits, including cardiac hypertrophy, blood pressure (BP), and heart remodeling. We performed gene co-expression network analysis in the rat RI strains left ventricle (LV) transcriptome using RNA-sequencing (RNA-seq) data. This identified 41 distinct gene co-expression networks (Fig. 1a, Supplementary Data 1a, b). We then tested the association of these gene co-expression networks with quantitative histopathologic measurements of interstitial and perivascular fibrosis in the rat heart (Supplementary Fig. 1). We identified five gene co-expression networks associated with both interstitial and perivascular cardiac fibrosis in the rat (adjusted P < 0.05, permutation test in Gene Set Enrichment Analysis [GSEA] (Fig. 1b). To uncover transcriptional programs relevant to the fibrogenic processes in human heart disease, we performed a separate gene network analysis using LV RNA-seq data generated from a cohort of patients with DCM (n = 126) and a cohort of control heart samples (n = 92 organ donors whose hearts were explanted for transplantation). We inferred gene co-expression networks in the human DCM LV transcriptome, which resulted in 48 distinct networks (Supplementary Data 2a, b). Then, we assessed which of these human DCM gene networks was conserved in the rat LV. This identified 14 human DCM networks as having some degree of conservation with the rat networks (adjusted P < 0.05, Fisher’s exact test [FET]) (Fig. 1c). To obtain additional evidences supporting a role for these networks in human heart disease, we formally tested which gene co-expression pattern was present only in the LV from human DCM patients and not in LV from controls, i.e., we tested for differential co-expression between DCM and control hearts (see Methods). The differential co-expression paradigm assumes that the disease state is linked to perturbations of the structure of the regulatory network itself, and might reflect the dysregulation of the underlying transcription factors (TFs) in disease. Here, we found that eight human gene networks were both conserved across species and differentially co-expressed between human DCM and controls (adjusted P < 0.05, permutation test). Despite the fact that several networks were significantly conserved between rat and human DCM heart (see Extended Analyses in Supplementary Note 1), only one human network (Hs M47), containing 683 genes, was (1) significantly conserved in the rat (sharing 72 genes with the rat network M1, adjusted P = 8.4 × 10^{-45} in FET, Fig. 1d), (2) associated with both interstitial and perivascular fibrosis in the rat heart (Fig. 1b), and (3) differentially co-expressed in human DCM heart (Fig. 1c, right panel).

The Hs M47 was functionally relevant for ECM regulation, as it was significantly enriched for genes belonging to the specific biological pathways and processes: “ECM-receptor interaction”, “TGFβ signaling pathway”, and “focal adhesion” (Fig. 1e). Henceforth, we will refer to Hs M47 as human ECM-network (or hECM-network). This hECM-network contains 237 strongly co-expressed genes annotated to encode for extracellular ECM region proteins (Supplementary Data 3b), and among these we highlight: 21 genes encoding for collagenes; focal adhesion molecules such as ITGB5, COMP, MAPK10, and THBS4; several extracellular genes involved in TGFβ-signaling (e.g., Dcn, CHRD, TGFβ3); three members of the BMP family (BMP4, BMP6, and BMP8B), and other important matricellular proteins, such as CTGF and PDGF, which contribute to the fibrogenic response.

We further investigated whether the hECM network was specific to the ECM remodeling processes undergoing in DCM and/or in LV tissue. To this aim, we considered a separate heart condition, rTOF, which has a very different etiology from DCM but is characterized by the presence of cardiac dysfunction and diffuse and pathologic myocardial fibrosis of both the right ventricle (RV) and LV. We analyzed RNA-seq data from RV in
Fig. 1 Co-expression networks analysis in rats and humans identifies a disease-associated pro-fibrotic ECM transcriptional program. **a** Hierarchical clustering of the first principal component of the rat co-expression networks (M1 to M39) inferred in the LV of the RI rat strains. Each branch denotes a distinct rat network with a different number of genes (full list in Supplementary Data 1a, b). **b** Level plot displaying the GSEA results with the significance of the association of each network to interstitial and perivascular fibrosis. Only DCM networks that are significantly intersecting (adjusted \( P < 0.05 \), FET) to any rat network are shown (full list of networks in Supplementary Data 2a, b). The right panel highlights in red the human DCM modules that are significantly differentially co-expressed between DCM and control LV samples (red denotes adjusted \( P < 0.05 \), permutation test). **c** Each row represents a human co-expression network inferred in the LV of DCM patients. Only DCM networks that are significantly intersecting (adjusted \( P < 0.05 \), FET) to any rat network are shown (full list of networks in Supplementary Data 2a, b). The right panel highlights in red the human DCM modules that are significantly differentially co-expressed between DCM and control LV samples (red denotes adjusted \( P < 0.05 \), permutation test). **d** Shortlisted rat and human networks: rat network M1 (rat ECM-network) and human network Hs M47 (hecM-network). These two networks have a significant overlap and share 72 genes (adjusted \( P = 8.39 \times 10^{-45} \), FET). **e** Functional enrichment of heCM-network. **f** Pairwise gene correlation in heCM network (heart fibrotic disease vs healthy). **g** % of heCM-network genes activated by TGFβ in human cardiac fibroblasts. **h** Enrichment of SMAD targets in heCM-network. The network contains 683 genes; 23% of these genes are targets of any SMAD (i.e., considering SMAD1/2/3/4; full list of SMADs targets in Supplementary Data 3d)
a cohort of adult patients with rTOF \((n = 27)\) who underwent surgery for pulmonary valve replacement and age-matched control donor samples \((n = 11)\), full details in Supplementary Table 1. The hECM gene co-expression network found in DCM LV was significantly conserved in rTOF RV \((P < 1 \times 10^{-3}\), permutation test) but not in control RV \((P \sim 1\), permutation test). The gene co-expression pattern of the hECM-network in rTOF RV mirrored the pattern observed in DCM LV (i.e., strongest gene-gene co-expression in disease), and this was significantly different between rTOF patients and controls (Fig. 1f). The consistent differential gene co-expression in DCM and rTOF heart thus suggests that the hECM network is capturing common ECM remodeling processes taking place across a wider range of human cardiac fibrotic diseases, irrespective of the specific disease etiology and the heart tissue, i.e., LV and RV. In addition, using longitudinal cardiac transcriptome data from a mouse genetic model of HF\(^{30}\), we show that this hECM network is enriched for genes upregulated in DCM that progresses to HF. Compared with control heart, 54% of the hECM-network genes are upregulated in DCM \((P = 5.4 \times 10^{-59}\), FET) and 53% are upregulated in HF \((P = 2.4 \times 10^{-49}\), FET). These results in human and mouse DCM/HF suggest that the identified hECM-network recapitulates the maladaptive fibrotic remodeling that promotes HF and adverse cardiovascular outcomes.

Since the hECM network was enriched for genes involved in the TGF\(\beta\) signaling (Fig. 1e), we looked at whether the hECM network was downstream of TGF\(\beta\)-receptor activation. We analyzed RNA-seq data generated from primary cultures of human atrial cardiac (myo)fibroblasts exposed to TGF\(\beta1\) (24 h), TGF\(\beta2\) (24 h) or control media. TGF\(\beta1\) and TGF\(\beta2\) induced the differential expression of 46% and 49% of the genes in the hECM-network: 315 and 335 genes, respectively (Fig. 1g, Supplementary Data 3c). Given the key role of SMAD TFs downstream of TGF\(\beta\)-receptors activation\(^{31}\), we tested whether the hECM-network was enriched for SMAD target genes. Using published chromatin immunoprecipitation-sequencing data\(^{32}\), we found that the hECM network was significantly enriched for SMAD-regulated genes \((n = 291, 43\%\) of the network genes), especially for SMAD2, SMAD3, and SMAD4 (Fig. 1h, Supplementary Data 3d). Together, these analyses revealed a cross-species conserved gene network, which might recapitulate the pathological ECM remodeling undergoing downstream of TGF\(\beta\)/SMAD signaling activation in the diseased heart.

**WWP2 regulates the pro-fibrotic gene network in heart.** Gene-gene co-expression suggests coordinated gene regulation by one or more master regulator gene. To identify master regulators of the hECM-network, we used advanced Bayesian genetic mapping approaches\(^{33}\) to pinpoint genomic loci regulating the whole ECM-network. In this, we consider the expression of the rat (or human) genes in the ECM-network as a multivariate quantitative trait, and then we test if the joint expression levels of the network genes are associated with genome-wide genetic variants (i.e., single-nucleotide polymorphisms, SNPs). This Bayesian expression quantitative trait locus (eQTL) mapping (or network-eQTL mapping) method was developed by our group to discover trans-acting master genetic regulators of networks in disease, including type 1 diabetes\(^{34}\), epilepsy\(^{35}\) and inflammatory disease\(^{36}\).

Here, we first mapped the ECM-network in the rat, and identified a single locus on rat chromosome 19 regulating 219 genes of the network (median Bayes factor \([BF] = 181.7\), where the BF represents the strength of genetic regulation versus no genetic control of the network) (Fig. 2a). Then, we investigated whether the regulatory locus for the rat ECM-network was conserved in humans. In the DCM patient cohort where the network genes are strongly co-regulated (Fig. 1f), we tested whether the hECM-network genes were jointly mapping to the human locus syntenic to the regulatory locus found in the rat (the human syntenic locus is located on chromosome 16, Fig. 2b). Network-eQTL mapping in human DCM detected a single regulatory SNP (rs9936589) located within an intron of the WWP2 gene, an E3 ubiquitin ligase. This SNP was strongly associated with the expression of the hECM-network in DCM heart (Fig. 2c) (median BF = 2004 for the 683 hECM-network genes). This network-eQTL was not detectable in the heart from control organ donors (Supplementary Data 3a), suggesting that the genetic regulation of the hECM network is present (or is detectable) only in diseased heart. This regulatory SNP (rs9936589) and WWP2 have not been previously associated with any fibrotic or heart disease.

These systems genetics analyses revealed a coordinated pro-fibrotic ECM transcriptional program in the diseased heart, which is regulated by a genetic variant within the WWP2 locus (Fig. 2a-c). Therefore, we investigated whether the WWP2 gene was a potential regulator of the hECM network in human fibrotic heart disease. We correlated the expression levels of WWP2 to the expression levels of the 683 hECM-network genes in LV and RV fibrotic hearts (i.e., in DCM and rTOF), and control hearts separately. In both LV (DCM) and RV (rTOF) fibrotic hearts, we observed a positive and significant shift in the distribution of the correlations between WWP2 and the expression of the hECM-network genes (Fig. 2d, e), suggesting a positive association between WWP2 cardiac expression and the hECM-network genes in disease. However, WWP2 cardiac expression was only moderately increased in heart disease: human rTOF vs control, fold change \((FC) = 1.23\); human DCM vs control\(^{26}\), FC = 1.02; mouse HF vs WT control\(^{30}\), FC = 1.86. We also identified a core set of genes in the hECM network that were positively and consistently correlated with WWP2 expression in the heart (false discovery rate \([FDR]\) < 1%), irrespective of heart tissue of origin (i.e., LV or RV) or disease (i.e., rTOF or DCM) (Fig. 2f, Supplementary Data 3e). This core gene set comprises known regulators of the pathological ECM remodeling, including matrix metalloproteinases (e.g., MMP14, MMP2)\(^{37}\) and their tissue inhibitors (e.g., TIMP2)\(^{38}\), several collagen (COL1A1, COL1A2, COL5A1, COL6A2, COL8A2, and COL14A1) and their binding partners (e.g., TGF\(\beta3\))\(^{39}\), microfibrillar-associated proteins, and pro-fibrotic cytokines (e.g., TGF\(\beta3\))\(^{30}\). In summary, these findings show that increased expression of WWP2 is associated with an elevated pro-fibrotic gene expression program in the diseased heart.

**Analysis of the cardiac expression levels of hECM-network genes stratified by the genotypes of the regulatory SNP rs9936589 showed increased expression associated with the TT genotype (Fig. 2g).** To investigate whether the effect of regulatory SNP on the hECM network was mediated by WWP2, we tested whether WWP2 cardiac expression was similarly regulated by the same SNP. Three main WWP2 gene isoforms have been characterized containing different protein domains: full-length isoform (WWP2-FL, covering the entire gene), N-terminal isoform (WWP2-N, containing the 5’ end of the protein) and C-terminal isoform (WWP2-C, containing the 3’ end of the protein)\(^{40}\) (Fig. 2h). All three isoforms are expressed in the heart of DCM patients (Supplementary Data 3f). Isoform-specific cis-eQTL mapping for each WWP2 isoform showed that only the WWP2-N was regulated by the SNP rs9936589 in DCM heart (Fig. 2i, right). Increased WWP2-N cardiac expression was associated with the TT genotype (Fig. 2i, left), matching the hECM-network regulation by SNP rs9936589 in DCM heart (Fig. 2g). This concordance of genetic regulation suggests that the
WWP2-N isoform, containing the N-terminal C2 and WW1 domains of WWP2, can be a positive regulator of the hECM network in the diseased heart.

**WWP2 regulates cardiac fibrosis in vivo.** To identify the pathophysiological processes regulated by the N-terminal region of WWP2, we generated Wwp2 mutant mice (WWP2Mut/Mut) by introducing a 4-bp deletion in exon 2 of Wwp2, which would lead to disruption of WWP2-FL and WWP2-N isoforms (Fig. 3a, Supplementary Fig. 2a–c). Consistent with previously reported phenotype in Wwp2-null mice, the WWP2Mut/Mut mice generated in our study showed reduced body weight, abnormal craniofacial development and elongated teeth (Fig. 3b, c,
Supplementary Fig. 2d). This 4-bp deletion in Wwp2 resulted in ablation of WWP2 isoforms containing the Wwp2 N-terminal region (i.e., Wwp2-FL and Wwp2-N isoforms, detected at the mRNA level by isoform-specific primer pairs P1; Fig. 3a, d), and lack of WWP2-FL and WWP2-N proteins (Fig. 3e, Supplementary Fig. 2e). The cardiac expression of WWP2-C protein was not affected.

Our systems genetics analysis indicated that the WWP2-N isoform positively regulates a pro-fibrotic transcriptional network in diseased heart (Figs. 1 and 2). Therefore, we hypothesized that loss-of-function (LOF) of WWP2-N/FL protein isoforms might regulate the in vivo fibrogenic response. Upon chronic (4 weeks) AngII treatment, AngII infusion resulted in increased tissue fibrosis, ventricular remodeling, and worsened cardiac function (Supplementary Fig. 3a–d). We also detected increased WWP2 transcripts and proteins levels in LV tissue (Supplementary Fig. 3e–g). Compared with WT mice, WWP2 Mut/Mut mice showed a significant improvement of AngII-induced cardiac fibrosis as shown by the lower percentage of tissues fibrosis (Fig. 3f), which was accompanied by an attenuation of cardiac hypertrophy (Fig. 3g, Supplementary Fig. 4). Cardiac function was also improved in WWP2 Mut/Mut mice, which showed increased fractional shortening (FS%) and ejection fraction (EF%) as compared with AngII-treated WT mice (Fig. 3h).

Bulk RNA-seq analysis in AngII-treated WT (n = 8) and WWP2 Mut/Mut mice (n = 8) hearts revealed that the mouse orthologs of the hECM-network genes detected in DCM heart have a different co-expression pattern with increased gene co-expression in WT compared with WWP2 Mut/Mut (P = 0.003 by permutation test, Fig. 3i). This pattern of differential co-expression (WWP2 Mut/Mut vs WT) was remarkably similar to the pattern observed in humans (i.e., between DCM or rTOF and controls, Fig. 1f), where the WWP2 Mut/Mut mice showed a co-expression pattern similar to control heart. Consistent with human data (Fig. 1c, g), the hECM network was significantly enriched for differentially expressed (DE) genes between AngII-treated WT and WWP2 Mut/Mut mice (Supplementary Data 4b). Among others, we detected “TGFβ signaling” and “extracellular matrix” as two of the major downregulated pathways in WWP2 Mut/Mut mice upon AngII treatment (Fig. 3f, Supplementary Data 4a). WWP2 Mut/Mut mice had reduced levels of fibroblast activation and ECM protein markers, as shown by smooth muscle actin (ACTA2), collagen1 (COL1A1), fibronectin extracellular domain A (FN-EDA), and periostin (POSTN) abundance in the heart after AngII treatment (Fig. 3k–m, Supplementary Fig. 5).

We used a second in vivo fibrosis disease model (myocardial infarction [MI] model, see Methods) and tested the potential protective effect of the WWP2-N/FL LOF on cardiac fibrosis and function post-MI. Histological analysis showed less post-MI fibrotic remodeling in the WWP2 Mut/Mut hearts compared with WT (Fig. 3n). This was associated with reduced chamber dilation and greater preservation of contractile function in WWP2 Mut/Mut mice compared with WT (Fig. 3o, p). Taken together, these data suggest that LOF of the WWP2 isoforms containing N-terminal region reduced cardiac fibrosis and improved cardiac function following AngII treatment or MI. This is in keeping with a role for WWP2 as a positive regulator of fibrosis in the diseased heart.

**WWP2 regulates the TGFβ1-induced fibrotic response In vitro.** To investigate the regulation of WWP2 in cardiac cells, we imaged WWP2-expressing cell(s) in heart sections by immunofluorescence. WWP2-positive cells did not show the morphology typical of a sarcomere-containing cardiomyocyte, and some WWP2-positive cells expressed fibroblast-specific protein 1 (FSP1) (Fig. 4a, Supplementary Fig. 6a). Cultured (myo)fibroblasts isolated from the LV of WT mice showed co-expression of WWP2 and FSP1 (Supplementary Fig. 6b). Single-cell RNA-seq analysis in the WT heart following AngII treatment provided additional evidence of WWP2 expression in (myo)fibroblasts, but also in endothelial and immune cells (Fig. 4b). By contrast, and consistent with our immunofluorescence data, WWP2 was not expressed in cardiomyocytes.

**TGFβ1 stimulation of primary murine LV (myo)fibroblasts induced robust induction of Wwp2 transcription at 72 h of treatment (Supplementary Fig. 7).** We then investigated the impact of WWP2 LOF in the response to prolonged (72 h) TGFβ1 treatment in primary murine LV (myo)fibroblasts (Fig. 4c). In (myo)fibroblasts, TGFβ1 stimulation increased pro-fibrotic activity and ECM production (measured by ACTA2, COL1A1, and POSTN), but the TGFβ1-induced pro-fibrotic expression changes at both the mRNA and protein levels were largely prevented in WWP2 Mut/Mut (myo)fibroblasts (Fig. 4d, e). TGFβ1-stimulated WT (myo)fibroblasts presented a clear organization of ACTA2 into stress fibers, while WWP2 Mut/Mut-derived cells showed a diffuse expression of ACTA2 with rare incorporation into stress fibers (Fig. 4f). TGFβ1 mildly increased vimentin protein expression, which was reduced in WWP2 Mut/Mut cells (Supplementary Fig. 8a, b). However, we did not detect differences in the mRNA level of vimentin and Transcription Factor 21 (Tcf21) (Fig. 4d, Supplementary Fig. 8c). Upon TGFβ1 treatment we also observed induction of TGFβ3 receptors (Tgfbr1 and Tgfbr2) in cardiac (myo)fibroblasts from WWP2 Mut/Mut mice (Supplementary Fig. 8d), suggesting a potential compensatory effect of WPP2 on TGFβ-signaling activation. In addition, WWP2 Mut/Mut cardiac (myo)fibroblasts...
showed higher cell proliferation and migration compared to controls (Supplementary Fig. 9), which is consistent with reduced maturation in myofibroblasts. Our analysis of WWP2 in DCM heart suggested a role for the WWP2 isoforms containing N-terminal region in fibrosis (Fig. 2i).

In line with this, murine cardiac (myo)fibroblasts responded to TGFβ1 treatment with increased protein levels of the WWP2-FL/N isoforms, but not of WWP2-C (Fig. 4g). Such induction of WWP2-FL/N isoforms was consistent with the findings in mouse heart following AngII infusion in vivo (Supplementary Fig. 3a–c).
WWP2Mut/Mut mice lack WWP2-FL and WWP2-N protein isoforms (Fig. 3d, e, Supplementary Fig. 2e), and this is sufficient to alter the co-regulation of the hECM-network genes, reduce cardiac fibrosis in vivo (Fig. 3f–p), and decrease the maturation of myofibroblasts in vitro (Fig. 4c–f). These data combined suggest a primary role for WWP2 gene isoforms containing N-terminal region of the protein in regulating a transcriptional program associated with cardiac fibrosis.

To confirm the differential effects of the WWP2-N-terminal region on cardiac (myo)fibroblast activity, we designed siRNA sequences to match either the 5′- (siRNA-Wwp2-N′) or 3′-terminal (siRNA-Wwp2-C′) regions of the Wwp2 mRNA, targeting different transcripts (Fig. 4h, Supplementary Fig. 10a). We successfully decreased the expression of WWP2-FL/N and WWP2-FL/C isoforms in cardiac (myo)fibroblasts, respectively (Fig. 4i, Supplementary Fig. 10b). Compared with scrambled control, both siRNAs mitigated the expression of ACTA2 in WT primary cardiac (myo)fibroblasts treated with TGFβ1 (Fig. 4i, j). In keeping with the WWP2Mut/Mut cardiac (myo)fibroblasts data (Supplementary Fig. 8d), after WWP2 knockdown, the (myo)fibroblasts showed increased mRNA expression of Tgfr1 and Tgfbr2 following TGFβ1 treatment (Fig. 4i). In human primary cardiac (myo)fibroblasts, siRNA experiments targeting the 5′-terminal of WWP2 mRNA (siRNA-Wwp2-N′) confirmed that WWP2-N/FL knockdown is able to reduce pro-fibrotic gene expression (Fig. 4k). We then performed a rescue experiment in mice, by re-introducing separately each of the two isoforms containing WWP2 N-terminal region (i.e., WWP2-FL and WWP2-N) in primary cardiac (myo)fibroblasts from WWP2Mut/Mut mice (Fig. 4l). Both WWP2-FL and WWP2-N individually increased the expression of pro-fibrotic genes in the WWP2Mut/Mut (myo)fibroblasts treated with TGFβ1 (Fig. 4m, n), supporting the role of these isoforms in regulating the fibrotic response.

**WWP2 regulates the nucleocytoplasmic shuttling of SMAD2.** We investigated the mechanisms through which WWP2 regulates the fibrotic response downstream of TGFβ-signaling activation. WWP2 is present in both the cytoplasm and nuclei in untreated cardiac (myo)fibroblasts; however, upon TGFβ1 stimulation (16 h) we observed increased WWP2 expression localized predominantly to the nucleus (Fig. 5a). While the WWP2-FL and WWP2-N protein isoforms were increased in the nucleus after TGFβ1 stimulation, the WWP2-C isoform remained in the cytoplasm (Fig. 5b, Supplementary Fig. 11a).

This TGFβ1-induced nuclear localization of WWP2 isoforms suggest that WWP2 may be involved in regulating gene expression, possibly targeting TFs for ubiquitination as shown for other E3 ubiquitin ligases. We have showed that the pro-fibrotic transcriptional program (i.e., the hECM network) regulated by WWP2 is enriched for transcriptional targets of the TGFβ-signaling transducer SMAD Tfs (Fig. 1h). In keeping with this, the genes downregulated by WWP2 in vivo following AngII treatment were enriched for “SMAD binding” (FDR = 0.003), “SMAD protein signal” (FDR = 0.006), and “transcriptional activity of SMAD2/3/4 heterotrimer” (FDR = 0.015) (Supplementary Data 4). We did not observe any difference at the mRNA level of SMAD2 in WWP2Mut/Mut cardiac (myo)fibroblasts or in cells overexpressing or silencing the WWP2-N/FL isoforms (Supplementary Fig. 11b, c), suggesting that WWP2 function on cardiac fibrosis could be exerted through SMAD2 interaction and ubiquitination at the protein level. When ectopically expressed in NIH-3T3 mouse embryonic fibroblast cells, we found that Flag-tagged WWP2-FL and WWP2-N, but not WWP2-C, co-immunoprecipitated with SMAD2 protein (Supplementary Fig. 11d). Mouse WT cardiac (myo)fibroblasts responded to TGFβ1 stimulation with SMAD2 binding to WWP2-FL and WWP2-N, but not to WWP2-C (Supplementary Fig. 11e). These data are suggestive of an endogenous physiological interaction between SMAD2 and WWP2-FL/N protein isoforms. Further analysis of inhibitory SMAD7, a preferred substrate for WWP2-FL and WWP2-C, showed co-immunoprecipitation with WWP2-FL and WWP2-C following TGFβ1 stimulation (Supplementary Fig. 11d), which we confirmed in primary (myo)fibroblasts (Supplementary Fig. 11g, h). We also found that WWP2 directly interacts with p-SMAD2 (Supplementary Fig. 11f). Notably, the levels of SMAD2 and p-SMAD2 proteins were similar in WT and WWP2Mut/Mut (myo)fibroblasts treated with TGFβ1 (Fig. 5c). This would be consistent with the monoubiquitination of SMAD2 by WWP2, resulting in a post-translational modification that does not affect SMAD2 protein levels. Ubiquitination assays followed by SMAD2 immunoprecipitation detected monoubiquitinated SMAD2 within 16 h of TGFβ1 stimulation (Supplementary Fig. 11i). We confirmed SMAD2 monoubiquitination by WWP2 in mouse primary cardiac (myo)fibroblasts, and show that this was reduced in WWP2Mut/Mut cells (Fig. 5d).
Monoubiquitination of SMADs by E3 ubiquitin ligases reportedly affects their transcriptional activity. To test if WWP2 similarly affects SMAD2 transcriptional activity, we performed SMAD2-dependent luciferase reporter activity assay in primary cardiac (myo)fibroblasts from WT and WWP2\textsuperscript{Mut/Mut} mice. TGF\(\beta\)1-dependent SMAD2 reporter activity was significantly lower in WWP2\textsuperscript{Mut/Mut} cardiac (myo)fibroblasts compared to WT cells (Fig. 5e). TGF\(\beta\)-receptor activation promotes the

![Image](https://example.com/image1.png)

![Image](https://example.com/image2.png)

![Image](https://example.com/image3.png)

![Image](https://example.com/image4.png)

![Image](https://example.com/image5.png)

![Image](https://example.com/image6.png)

![Image](https://example.com/image7.png)

![Image](https://example.com/image8.png)

![Image](https://example.com/image9.png)

![Image](https://example.com/image10.png)

![Image](https://example.com/image11.png)

![Image](https://example.com/image12.png)

![Image](https://example.com/image13.png)

![Image](https://example.com/image14.png)

![Image](https://example.com/image15.png)

![Image](https://example.com/image16.png)

![Image](https://example.com/image17.png)

![Image](https://example.com/image18.png)

![Image](https://example.com/image19.png)

![Image](https://example.com/image20.png)
WWP2 regulates pro-fibrotic gene activity in primary cardiac fibroblasts. a Immunofluorescence images of LV-section staining from WT mice after AngII infusion, showing WWP2 expression (green, arrow) in non-myocytes (Left); WWP2 co-localization with FSP1-positive cells (red, arrow; Right). Scale bar: 40 μm. b t-SNE displaying LV single-cell RNA-seq data (508 cells detected) in WT mice after AngII infusion. Each dot corresponds to a single-cell, colored by Wwp2 expression level (Log2, normalized counts). Different cell subpopulations detected in mouse heart are indicated. c Primary cardiac (myo) fibroblasts were taken from LV in WT and WWP2Mut/Mut mice and modeled in vitro by incubation with TGFβI (5 ng/ml, 72 h). d Relative mRNA expression (normalized to 18S) of ECM and pro-fibrotic marker genes (Acta2, Col1a1, Tgfβ1, Ctgf, Postn, and EDN-1) in primary cardiac (myo)fibroblasts (n = 5–6 for each group). e WB of ACTA2 and POSTN protein expression in WT and WWP2Mut/Mut cardiac (myo)fibroblasts. f Representative microscopy images (left) and quantification analysis (right) with immunostaining for ACTA2 and COL1A1 after TGFβI stimulation in WT and WWP2Mut/Mut cardiac (myo)fibroblasts (five biological replicates, nine images each; box-and-whisker plots). g Representative WB of WWP2-FL, WWP2-N and WWP2-C isoforms expression in cardiac (myo)fibroblasts after 72 h TGFβI stimulation. h WT cardiac (myo)fibroblasts were incubated with TGFβI (72 h) and siRNA pools (SiRNA-Wwp2-N′ and SiRNA-Wwp2-C′) against 5′- or 3′-region of Wwp2 mRNA. i WT (myo)fibroblasts, SiRNAs-Wwp2-N′/C′, attenuated TGFβI-induced pro-fibrotic responses, as shown by qPCR analysis of Acta2, Col1a1, and TGFβI receptors (n = 5–6 for each group). j Representative WB showing the effect of siRNA-Wwp2-N′ and siRNA-Wwp2-C′ on each Wwp2 isoform expression in WT (myo)fibroblasts and decrease of TGFβI-induced ACTA2 protein level by siRNA-Wwp2-N′. k Cultured human RV primary cardiac (myo)fibroblasts: siRNA-mediated knockdown of WWP2 reduced the expression of pro-fibrotic genes (COL1A1, COL1A2, and LUM). P1 tags both WWP2-N and WWP2-FL, P2 tags WWP2-FL only (n = 3 for each group). l Cardiac (myo) fibroblasts from WWP2Mut/Mut (myo)fibroblasts (n = 5 for each group) and ACTA2 protein levels. These fibrogenic markers were enhanced by either Wwp2-FL or Wwp2-N transfection after 72 h TGFβI treatment. P values calculated by Mann–Whitney U test; *P < 0.05, **P < 0.01, NS not significant; data reported as means ± SD.

nuclear accumulation of SMAD2/3 and this process is not necessarily accompanied by SMAD degradation in the nucleus as SMADs are exported out of the nucleus upon dephosphorylation and dissociation of the SMAD complexes. We obtained nuclear and cytoplasmic fractions from cardiac (myo) fibroblasts, and show the nuclear accumulation of SMAD2 upon TGFβ1 stimulation (<16 h). Compared to WT cells, SMAD2 is more localized to the nucleus in WWP2Mut/Mut cells (Fig. 5f). SMAD4, which forms a heteromeric complex with SMAD2 after TGFβ1 activation, showed similar protein level and subcellular localization in WT and WWP2Mut/Mut (myo)fibroblasts. Thus, despite the fact the SMAD2 protein is more abundant in the nucleus of WWP2Mut/Mut (myo)fibroblasts, lack of WWP2-N/FL was associated with a reduced transcriptional activity of SMAD2 downstream of TGFβ-receptor activation (Fig. 5e).

Monoubiquitination is also important for the proper subcellular localization of SMADs, which in turn might regulate their transcriptional activity in the nucleus. We used SB431542, a selective inhibitor of TGFβ superfamily type I activin receptor-like kinase (ALK) receptors, to study the nucleocytoplasmic shuttling of SMAD2 (ref. 49). Using primary (myo)fibroblasts from WWP2Mut/Mut and WT mice, we observed a delay in the nuclear export of SMAD2 in WWP2Mut/Mut cells (Fig. 5g). Compared with WT cells, WWP2Mut/Mut cardiac (myo)fibroblasts maintained sizeable levels of p-SMAD2 at 3 h treatment with SB431542 (Supplementary Fig. 12). Taken together, these data show that WWP2 interacts with SMAD2, promoting its monoubiquitination, and modulates the nucleocytoplasmic shuttling and transcriptional activity of SMAD2 downstream of TGFβ-signaling activation (Fig. 5h).

**Discussion**

The WW domain containing E3 ubiquitin protein ligase 2 (WWP2) gene, also known as atrophin-1-interacting protein 2 (AIP-2), is a multifunctional ubiquitin E3 ligase. It has been previously reported to regulate various processes, including palatogenesis, craniofacial development, TLR3-mediated innate immune and inflammatory responses, cell death and tumorigenesis, and tumor growth. Common genetic variants within the WWP2 locus have been associated with plantar fasciitis and osteoarthritis, supporting the heterogeneous functions associated with WWP2. Working in cancer cell lines, Soond et al. implicated WWP2 in oncogenic TGFβ-induced epithelial–mesenchymal transition (EMT), reporting the selective and TGFβ-dependent targeting of SMADs by specific ectopic WWP2 isoforms. The WWP2-N isoform was shown to enhance the activity of WWP2-FL, which in turn degrades SMAD2/3. Given the primary role of TGFβ-SMAD2/3 signaling in activating tissue-resident cardiac fibroblasts and the fibrotic response, the data by Soond et al. might suggest that increased WWP2-N protein isoform expression with concurrent increased activity of WWP2-FL would lead to reduced fibrogenesis. However, any possible fibrotic or anti-fibrotic function of WWP2 has never been investigated, especially in cardiac pathophysiology. Here, starting from the identification of a pro-fibrotic gene network conserved in rat and human heart disease characterized by diffuse myocardial remodeling and fibrosis, we identified WWP2 as a regulator of pathological cardiac fibrosis. The WWP2-regulated pro-fibrotic gene network was also conserved across different heart fibrotic diseases and cardiac tissues, and was upregulated in DCM that progresses to HF in mice. The lack of regulation of the pro-fibrotic gene network by WWP2 in human control heart suggests that WWP2 exerts its regulatory role on cardiac fibrosis upon disease. Interestingly, the hECM network regulated by WWP2 in DCM heart is enriched for genes highly expressed in fibroblasts (see Supplementary Fig. 13, Supplementary Methods). This might suggest that the power to detect genetic regulation of the pro-fibrotic network in DCM was enhanced because the heart tissues underwent strong remodeling with concurrent changes in relative cellular composition (i.e., more fibroblasts). WWP2 mRNA expression was only marginally increased the fibrotic heart disease (less than 2 folds in either human DCM, rTOF, or mouse HF). This might explain why WWP2 passed undetected to GWAS, eQTL mapping, and genetic screening studies of fibrotic diseases. Detailed systems genetics analyses in the DCM heart allowed us to hypothesize that increased expression of the WWP2-N-terminal isoform was associated with the activation of a pro-fibrotic gene program downstream of TGFβ/SMAD signaling activation (Figs. 1 and 2). We tested and corroborated this hypothesis in primary cardiac fibroblasts and in two preclinical models of fibrotic heart disease. Our results provide the first indication of a role for WWP2 in regulating pathophysiological processes in the heart. In detail, we have demonstrated that WWP2-N/FL LOF improved cardiac function and reduced myocardial fibrosis in vivo (Fig. 3). Upon

Fig. 4 WWP2 regulates pro-fibrotic gene activity in primary cardiac fibroblasts. a Immunofluorescence images of LV-section staining from WT mice after AngII infusion, showing WWP2 expression (green, arrow) in non-myocytes (Left); WWP2 co-localization with FSP1-positive cells (red, arrow; Right). Scale bar: 40 μm. b t-SNE displaying LV single-cell RNA-seq data (508 cells detected) in WT mice after AngII infusion. Each dot corresponds to a single-cell, colored by Wwp2 expression level (Log2, normalized counts). Different cell subpopulations detected in mouse heart are indicated. c Primary cardiac (myo) fibroblasts were taken from LV in WT and WWP2Mut/Mut mice and modeled in vitro by incubation with TGFβI (5 ng/ml, 72 h). d Relative mRNA expression (normalized to 18S) of ECM and pro-fibrotic marker genes (Acta2, Col1a1, Tgfβ1, Ctgf, Postn, and EDN-1) in primary cardiac (myo)fibroblasts (n = 5–6 for each group). e WB of ACTA2 and POSTN protein expression in WT and WWP2Mut/Mut cardiac (myo)fibroblasts. f Representative microscopy images (left) and quantification analysis (right) with immunostaining for ACTA2 and COL1A1 after TGFβI stimulation in WT and WWP2Mut/Mut cardiac (myo)fibroblasts (five biological replicates, nine images each; box-and-whisker plots). g Representative WB of WWP2-FL, WWP2-N and WWP2-C isoforms expression in cardiac (myo)fibroblasts after 72 h TGFβI stimulation. h WT cardiac (myo)fibroblasts were incubated with TGFβI (72 h) and siRNA pools (SiRNA-Wwp2-N′ and SiRNA-Wwp2-C′) against 5′- or 3′-region of Wwp2 mRNA. i WT (myo)fibroblasts, SiRNAs-Wwp2-N′/C′, attenuated TGFβI-induced pro-fibrotic responses, as shown by qPCR analysis of Acta2, Col1a1, and TGFβI receptors (n = 5–6 for each group). j Representative WB showing the effect of siRNA-Wwp2-N′ and siRNA-Wwp2-C′ on each Wwp2 isoform expression in WT (myo)fibroblasts and decrease of TGFβI-induced ACTA2 protein level by siRNA-Wwp2-N′. k Cultured human RV primary cardiac (myo)fibroblasts: siRNA-mediated knockdown of WWP2 reduced the expression of pro-fibrotic genes (COL1A1, COL1A2, and LUM). P1 tags both WWP2-N and WWP2-FL, P2 tags WWP2-FL only (n = 3 for each group). l Cardiac (myo) fibroblasts from WWP2Mut/Mut (myo)fibroblasts (n = 5 for each group) and ACTA2 protein levels. These fibrogenic markers were enhanced by either Wwp2-FL or Wwp2-N transfection after 72 h TGFβI treatment. P values calculated by Mann–Whitney U test; *P < 0.05, **P < 0.01, NS not significant; data reported as means ± SD.
AngII treatment, a typical promoter of cardiac fibrosis, the WWP2Mut/Mut mouse showed increased oxidative phosphorylation, and exacerbated cell cycle and interferon alpha and gamma response in the heart. This suggests that depletion of the N-terminal part of WWP2 might have a wider role in myocardial fibrosis, preventing the glycolytic metabolic reprogramming required for myofibroblast maturation, and potentially blocking the cell cycle arrest that has been shown to be a feature of fibrosis. We have also shown that WWP2 positively regulates the expression of established pro-fibrotic markers in primary (myo)fibroblasts cultures (Fig. 4). However, one limitation of our study relates to the maintenance of a quiescent cardiac fibroblast.
phenotype using common cell culture methods, and the cultured fibroblasts are prone to be converted to myofibroblasts before TGFβ1 treatment. In our study, we notice that the cultured cells presented a pro-fibrotic phenotype without TGFβ1 treatment, which also explains why the pro-fibrotic response to TGFβ1 was relatively modest. Therefore, we referred to the in vitro cultured cardiac fibroblasts as “(myo)fibroblasts”. However, TGFβ1 stimulation increased (myo)fibroblasts maturation, inducing the expression of pro-fibrotic markers and ECM genes, which was significantly reduced in WWP2Mut/Mut cells. The differential fibrotic response between WT and WWP2Mut/Mut cardiac (myo)fibroblasts was corroborated by WWP2 siRNA experiments and gain-of-function rescue experiments. Poly-ubiquitination and proteasome-mediated degradation of nuclear SMAD2 have been documented. In addition, mono-ubiquitination can regulate nuclear accumulation and the nucleocytoplasmic shuttling of SMAD complexes, which are crucial for transduction of TGFβ-superfamily signals and there is mounting evidence supporting the role of monoubiquitination of SMADs by E3 ubiquitin ligases in this context. Here, we reveal that TGFβ1 stimulation promotes the translocation of WWP2 to the nucleus, where it interacts with SMAD2, possibly promoting its monoubiquitination, as shown for other E3 ubiquitin ligases. The regulatory consequences of E3 ligase-mediated monoubiquitination appear to be complex and context specific. Monoubiquitination can regulate protein localization, activity, and protein interactions with binding partners. Monoubiquitination is also required for the activity and the intrinsic nuclear import of target TFs, and to disrupt specific TFs interactions and their transcriptional activity. In our studies, we did not observe SMAD2/3 degradation by endogenous WWP2 in primary (myo)fibroblasts, suggesting a mechanism of degradation-independent repression of SMAD2 activity downstream of TGFβ-signaling activation. Our data suggest that blocking WWP2 function can delay the TGFβ1-induced nucleocytoplasmic shuttling of SMAD2 (Fig. 5h). However, further studies are required to prove whether WWP2-mediated mono-ubiquitination of SMAD2 disrupts or directly regulates the activity SMAD complex, and to explain mechanistically how this WWP2 mono-ubiquitination affects the nuclear export and regulation of SMAD2 (ref. 67).

Ubiquitin ligases are involved in a wide range of diseases, and targeting E3 ubiquitin ligases is thought to yield higher specificity and less toxicity than other ubiquitins. E3 ubiquitin ligases can both degrade and activate specific substrates, and so may allow for a specific regulation of currently “undruggable” protein targets. To date, only a handful of small molecules for cancer therapy (e.g., targeting MDM2 (ref. 70) and SKP2 (ref. 71)) have been successfully developed using this type of enzymes. However, several E3 ubiquitin ligases such as SMURF1, SMURF2, ARKADIA, SYNOVIOLIN, NEDD4, and PELLINO1 (reviewed in ref. 72) have been proposed as regulators of tissue fibrosis, through several mechanisms including regulation of TGFβ-signaling. This study adds to our understanding of the regulation of cardiac fibrosis and uncovers a specific E3 ubiquitin ligase, WWP2, as a regulator of ECM accumulation in the diseased heart downstream of TGFβ/SMAD signaling. Since cardiac fibrosis has been proposed as an important therapeutic target in HF patients, the identification of druggable targets that regulate pathological fibrosis may provide new avenues to control the progression to HF. Here, using our systems genetics approach followed by functional validation studies in several cellular and preclinical animal models, we have first identified and then mechanistically explained that WWP2 is a novel and druggable therapeutic target with the potential to control fibrosis in pathological cardiac remodeling.

Methods

data generation and processing in the rat. This cardiac expression data set consists of RNA-sequencing in the heart (LV) from the 30 rat R1 strains, which were previously published. Data were processed with TopHat 1.2.0 (http://ccb.jhu.edu/software/tophat/), which was used to map the reads to the Brown Norway (BN) reference genome RGS3.3 (known splice junctions in the Ensembl reference database were supplied, de novo splice junction detection was also enabled). After mapping the reads to the reference genome, Cufflinks 1.2.1 (http://cole-trapnell-lab.github.io/cufflinks/releases/v1.2.1/) was used to assemble the aligned RNA-Seq reads into transcripts by using the Ensembl annotation of the rat transcriptome. Cufflinks is able to reconstruct the set of transcripts that “explains” the reads observed in our RNA-seq experiment, in our case, by using a transcriptome reference annotation. The reason for using a reference-based transcriptome assembly approach is that, comparative transcriptome assembly analyses have shown that, overall, reference-based transcriptome assembly approaches have a better performance than de novo approaches, in both sensitivity and the ability to discover splicing patterns. Cufflinks measures transcripts abundances in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). One of the assumptions of reference-based transcriptome assembly approaches is that all the isoforms of all the genes are known. However, transcriptomic reference sets are still incomplete, more specially in the case of the rat transcriptome. Hence, the transcriptomic assembly is not very accurate in rat. Therefore, all the RNA-seq expression data analyses in the rat have been carried out by summarizing expression at the gene level. Six steps followed for gene level FPKM computation and filtering of the data, as follows. (1) Replace the values of all the isoforms with isoform quantification status not OK by missing values. The isoform quantification status is part of the output from Cufflinks and it measures the successful deconvolution of each isoform. (2) Remove all isoforms with more than 10 missing values across all samples. (3) Impute all the missing values by using R function pcaMethods R package 1.5.60 (https://biocomputator.riken.jp/packages/pcaMethods/bioc/html/pcaMethods.html). The initial number of isoforms was 39,492. After removing those with more than 10 missing values across all samples, the number of isoforms got reduced to 39,181. (4) Sum up all the FPKM values of all the isoforms of each gene to get the FPKM value at the gene level (number of genes: 29,469). (5) Filter out genes that did not have FPKM >1 in at least a 5% of the samples (i.e., in at least two samples). (6) Replace 0s in the data by the minimum number in the data set higher than 0. Afterwards, Log2 transform the data. By following these steps, we ended up with a filtered data set of 12,061 genes. We also inspected the presence of unmeasured confounding factors (i.e., potential sources of expression variation that are not being measured in the study and they may be affecting gene expression).
We computed the variability explained by each of the principal components (PCs) of the transcriptomic data. We found out that the first PC was explaining a large proportion of the total variance, and was associated with the presence of collagen fibres correlated to the library concentration of the Bioanalyser (a library quantification measure, Spearman’s rank-order correlation = 0.70, $P = 1.5 \times 10^{-7}$). To account for possible confounding effects introduced by the differences in library concentration, we adjusted the gene expression data for this first PC using a linear regression model. Then, the healthy controls which were also adjusted from which the tissue had been collected by adding this information as an additional covariate. (5) Outlier samples removal (samples QC): Outlier samples were inspected by clustering the samples’ gene expression levels. The DCM and control cohorts were independently clustered by hierarchical clustering function flashClust from flashClust 1.0.1R2 package (https://rdr.io/cran/flashClust/). The agglomeration method used was set to “average”. The obtained dendrograms were cut at the 99th percentile of the dendrogram height distribution including all the samples assigned to small clusters: 2 DCM patients and 14 healthy donors. By following these steps, we ended up with two LV adult cohorts of 126 DCM patients (cases) and 92 healthy donors (controls). Within the 126 DCM patients, 106 were males and 20 were females. In the healthy donors group there were 55 males and 37 females. The mean age in DCM patients was 41.3 ± 13.3 years (16–64); mean age in the controls was 42.6 ± 13.6 years (17–72). A subset of the DCM and control LV samples from the EU population (96 DCM samples and 91 control samples) were typed by using genotyping arrays. After imputation and quality control, the final number of SNPs in the DCM patients and healthy controls LV samples was 1,399,892 SNPs. In order to run genetic mapping, we carried out a pairwise linkage disequilibrium (LD) filtering (with an $R^2$ threshold of 0.8) based on the 1000 Genomes pilot 1 CEU population. Within this population, SNP LD was computed using SNAP 2.2 (http://www.broad.mit.edu/mpg/snap/) using the SNP identifiers as input. The SNP data set chosen to run SNAP was the 1000 Genomes pilot 1 in the CEU reference population with an $R^2$ threshold of 0.8 and a distance limit of 500 bp. From SNAP output, SNPs returning a warning message as not present in the SNP database were removed and not considered. Further analyses (the number of missing SNPs was 101,070 (SNPs)). SNAP web tool outputs the input SNPs clustered in LD blocks (a total number of 126,922 LD blocks were returned by SNAP). To obtain a genomic map filtered by LD, we selected one single SNP for each of the LD blocks ($R^2 = 0.8$). SNPs that were not included in any of the LD blocks (not clustered by SNAP) were added at random. For each LD block we followed these steps: (1) Get all the SNPs in the LD block and compute their pairwise Spearman’s rank correlation. (2) As representative of the LD block, take the SNP that has the highest average Spearman’s rank correlation with the rest of SNPs included in the LD block. With this procedure we obtained a set of 126,922 SNPs, each of them representing a LD block. Finally, we added to these SNPs the ones that were not included by SNAP in any of the LD blocks (41,315 SNPs). Then, the final number of SNPs in our LD-pruned ($R^2 = 0.8$) genetic map was 168,237 SNPs. The rTOF cohort consists of 27 patients with rTOF (mean age 32.0 ± 10.6 years, 78% male). Summary of the clinical information of these patients is included in Supplementary Table 1. All the patients aged ≥16 years old and were scheduled for elective pulmonary valve replacement. They were also under the care of the Adult Congenital Heart Disease service at the Royal Brompton Hospital, UK. The rTOF patients, the clinical data from non-invasive investigations (performed as part of the surgical work-up) were available. This clinical information included electrocardiography, chest radiography, cardiac magnetic resonance imaging (CMR) and echocardiography. Furthermore, myocardial biopsy samples from the 27 patients were snap-frozen in liquid nitrogen at the time of tissue sampling intra-operatively. RV tissue from 11 structurally normal hearts donated for cardiac transplantation was also collected (age donors: 34.0 ± 13.0, sex, male/female: 6/5). RV myocardial biopsies were made available via the Cardiovascular Biomedical Research Unit Biobank of the Royal Brompton & Harefield NHS Foundation Trust. Donor RV control tissue was collected and stored at the time of surgery following similar procedures as with the TOF tissue samples. rTOF RV expression data are a cohort of RV RNA-seq samples with rTOF patient (fibrotic RV) and control samples. TOF is a congenital heart disease (i.e., problem in the structure of the heart) characterized by pulmonary artery stenosis and ventricular septal defect (a hole between the LV and RV and an overriding aorta, which allows blood from both ventricles to enter the aorta leading to cyanosis and RV hypertrophy). TOF needs to be treated surgically in the first few weeks of life to improve the function of the pulmonary valve and arteries and repairing the septal defect. All eligible patients were under the care of the Adult Congenital Heart Disease service at the Royal Brompton Hospital, UK. The Royal Brompton & Harefield NHS Trust and National Heart & Lung Institute Ethics Committee approved the study (NRES ref: 2011/04/20). The samples were collected with informed written consent from all study participants. Tissue studies were compliant with UK Human Tissue Act regulations. Patients provided specific signed permission for RV myocardial biopsies to be taken intra-operatively under direct vision during open-heart surgery when judged technically feasible by the operating surgeon. However, surgical repaired TOF is usually followed by cardiac fibrosis in both ventricles. A number of previous cases have shown several times damaged, TOF cases are therefore not suitable for basic research development because the prosthetic material used at surgeries is unable to grow with the growing of the heart and valve. Therefore, surgery leftover samples can be collected at different ages of the patients.
TIRiloz (Life Technologies) was used for total RNA extraction from the frozen samples by following the manufacturer’s protocol. RNA was quantified by ultraviolet spectrophotometry and RNA quality was assessed on the Agilent 2100 bioanalyzer. RNAs ranged from 6.3 to 9.1 (mean 8.2 ± 0.6). One microgram of total RNA was used to prepare the RNA-Seq libraries. RNA-Seq libraries were prepared with Illumina TruSeq RNA sample preparation kits by using the protocol for poly-A enriched mRNA. To avoid batch effects, samples were pooled (4–5 samples per pool, 2 pools per array) and RNA samples were tested for quality with the Bioanalyzer 2100 (Agilent Technologies). The Illumina HiSeq platform was used for sequencing (mean sequencing depth of 196M reads). TopHat 2.0.12 (https://ccb.jhu.edu/software/tophat/) with Bowtie 2 2.2.3 (http://bowtie-bio.sourceforge.net/bowtie2/) and Samtools 0.1.18 (https://sourceforge.net/projects/samtools/) was run by default for mapping to GRCh38 (hg38.78) reference genome. RNA-Seq read counts were computed with HTSeq 0.6.1 (https://htseq.readthedocs.io/). The percentage of reads mapping to the human genome was higher that 80% (above 70%) is considered an acceptable mapping percentage for paired-end sequenced reads. Steps followed for the normalization, filtering, and adjustment of the data: (1) Gene selection: From HTSeq output, only “protein coding” genes with status “known” in Ensemble genes (GRCh37.p3.d13 data set) were selected (18,964 genes). (2) Filtering of low expressed genes: FPKM values were computed with DESeq2 1.6.3 package using the average transcript length of each gene, which were retrieved from Ensembl Biomart (GRCh37 version). We used a FPKM > 1 filtering criterion by keeping only those genes with a value of FPKM > 1 in at least 5% of the samples (in this case 2 samples). Following these criteria, the number of genes got reduced from 18,964 genes to 13,936 genes. (3) Data transformation: Size factors normalization and VST was applied to the raw gene counts by using DESeq2 package 1.6.3. Then, the gene counts that passed the filtering criteria described in the previous step were selected. (4) Covariates adjustment: After normalization and filtering, the data were split into TOF patients and controls. Gene expression counts of TOF patients were adjusted for: (1) age at which the tissue was collected (age of operation) and (2) sex. This adjustment was performed by taking the residuals of a multivariate linear model, in which both age and sex were added as predictors. Among the control samples, some had missing age value for age. More specifically in those samples the sex of the sample with missing value was imputed by clustering the expression levels of selected sex specific genes (as described in ref. 79). Therefore, the gene expression counts of the control samples were adjusted only for sex by taking the residuals of a linear model in which sex was added as a predictor. Differentially expressed genes between rTOF patients (n = 27) and controls samples (n = 11) were computed using the R package DESeq2 1.6.3 and adding sex as covariate in the model (age was not added as there were several RV control samples with missing age values). CooksCutoff DESeq2 parameter was set to False. The rest of parameters were left to default.

**Gene co-expression network analysis.** Co-expression networks were inferred by using weighted gene correlation network analysis (WGCNA) 1.42 (https://northeastgenetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/). Two inde-pendent WGCNA runs were carried out in the processed RNA-seq LV data (as described above): one in rat LV samples and one in rat LV samples with missing age values for age. More specifically in those samples the sex of the sample with missing value was imputed by clustering the expression levels of selected sex specific genes (as described in ref. 78). Therefore, the gene expression counts of the control samples were adjusted only for sex by taking the residuals of a linear model in which sex was added as a predictor. Differentially expressed genes between rTOF patients (n = 27) and controls samples (n = 11) were computed using the R package DESeq2 1.6.3 and adding sex as covariate in the model (age was not added as there were several RV control samples with missing age values). CooksCutoff DESeq2 parameter was set to False. The rest of parameters were left to default.

**Genome-wide gene expression levels in the RI strains LV were correlated with both interstitial and perivascular fibrosis using Spearman’s ranked correlation procedure (rho); P < 0.05 (significance threshold).** For each gene, the results for the top significant GO (BP) only and KEGG terms were used for enrichment analysis using the R DAVID Web Service 1.4.0 (https://david.abcc.africau.edu/server.jsp?packages=3.0/bioc/html/RDAVIDWebService.html).

Rat networks were queried independently from human networks for Gene Ontology (Biological Process, BP, Molecular Function, MF, and Cellular Component, CC) and KEGG pathways. Gene Ontology (GO) and KEGG annotations categories. The gene reference background was set to the group of genes from which the gene co-expression networks were inferred. In the rat data this background was constituted by the genes robustly expressed in the rat LV data (i.e., after filtering procedure described in previous sections, n = 12,061), whereas in the human data the background used was the set of genes robustly expressed in the human LV RNA-seq cohort (i.e., after the filtering procedure previously described, n = 14,281). In this analysis results were deemed significant if DAVID FDR score < 0.05 (i.e., % FDR). For each network, the results for the top significant GO (BP) only and KEGG terms can be seen in Supplementary Data 1a (rat networks) and Supplementary Data 1b (human DCM network). The full list of KEGG enriched in the Hs-M47 (human ECM network, hECM-network) is found in Fig. 1e.

**Bayesian network-expression QTL (network-eQTL) analysis.** Co-expression networks suggest coordinated genetic regulation, which can be exploited to uncover
genetic regulators of these transcriptional programs. Moreover, conserved genetic regulation can be driving fundamental biological mechanisms. In keeping with previous studies in the rat, where the B6X129 F1 rat panel yielded increased power to carry out genetic mapping of gene networks34, we used multivariate Bayesian genetic mapping approaches to map the rat and human ECM-networks to the rat and human genomes. We first considered the expression of the rat ECM-network genes as a multivariate quantitative trait and jointly mapped this to the rat genome. Then, we inspected whether the regulatory loci identified in the rat was independently replicated in human DCM heart by joint mapping of the genes in the ortholog human network (e.g., hECM network) to the human locus that is syntenic to the rat regulatory loci. This two-step strategy (mapping in rats first followed by mapping in humans) has been previously used to identify trans-acting genetic regulators of transcriptional networks underlying complex disease34. The mapping of the rat and human networks was carried out by using HESS33. HESS is a sparse Bayesian multiple linear regression method in which mRNA expression levels for multiple genes are regressed against all SNPs to identify the minimum (non-redundant) set of SNPs that predicts the mRNA expression variability. This method has the following features: (1) it takes into account the LD structure of the genotype data (the dependence of the genetic determinants or predictors), allowing to reduce the number of tests to be carried out and pinpoint the putative causal genetic variant; (2) it makes possible to map several responses in one single test; therefore, with HESS is possible to map to the genome (i.e., all genome-wide genetic markers) expression levels of several genes jointly (for instance, map the expression levels of genes included in a co-expression network, without having to summarize their variability by PCA analysis); and (3) it exploits multidimensional dependencies within the responses (i.e., correlation of the gene expression levels). This can be used to boost detection of moderate trans-acting eQTL effects. The output of this strategy is the marginal posterior probability of inclusion (MPPI) for each gene-SNP pair tested, which represents the posterior probability of association of each SNP given the data. From this MPPI, the BF can be computed. BF represents the evidence of genetic regulation versus no genetic control and it is defined as the ratio between the posterior odds and the prior odds or ratio between the strength of evidence for genetic regulation and the prior probability that there is no genetic regulation. In our case of interest, for each gene, an SNP associated with the gth gene is defined as: $\pi = P(g|p)$, where $\pi$ is the number of SNPs we are testing and $P(g|p)$ is the prior expected number of control points for the gth gene, in our case we fix $E(p) = 2$. For instance, in the case of the rat RI strains, as the number of genome-wide SNPs is $p = 1384$ and the prior probability becomes $\pi = 1.4 \times 10^{-3}$, The BF is defined as the ratio between the posterior odds and the prior odds: $BF = \frac{P(g|p)}{P(p|g)}$, where MPPI represents the marginal posterior probability of inclusion for the gth gene and the jth SNP. By using this BF formula, we can compute the BF for each response-predictor pair (i.e., gene and SNP under testing) from the output MPPIs of HESS. Genome-wide mapping of the rat networks in the RI strains. The gene expression levels of the genes included in each of the rat co-expression networks built in the LV of interest were jointly mapped to the rat genome with HESS, i.e., rat networks conserved in human, overrepresented for genes correlating with expression not present in human control LV tissue: M1, M2, and M12 rat networks. The expression data used was the RNA-seq data in the RI strains (log2 transformed RPKM values) for the first 27 cases, where the expression was correlated to WWP2 was extracted by taking all the genes with correlation $r > 0.18$ isoforms. Here, we select the expression of the rat ECM-network genes, which are encoded by the three different gene isoforms. Here, a reading frame shift mutation s was used to generate mutation-segregated heterozygous F1 animals by CRISPR/Cas9 technology based on C57BL/6J strain. The list of the targeted genes and each gene (the average correlation in DCM > 0.18 isoforms).

**Expression QTL (eQTL) mapping of WWP2 isoforms in DCM.** WWP2 isoform expression levels were carried out in the DCM patients from the TopHat output with Sailfish 0.6.3 (http://www.cs.cmu.edu/~ckingsf/software/sailfish/). Sailfish was run with default parameterizations. The expression levels in DCM patients of all isoforms were analyzed with DEseq Reads Per Kilobase Million, (RPKM) are included in Supplementary Data 3. These data were adjusted for relevant technical ("RIN" score and "library preparation preparation") and clinical ("sex" and "age at tissue collection") covariates by using a multivariate linear model and then mapped to the regulatory locus identified in human chromosome 16 (1 Mb centered around the human ortholog gene. (4) In humans (human Ensembl version GRCh37), compute the positions of the rat haplotype that contained the regulatory SNP (rat Ensembl version GRCh38) and each gene tagged by this SNP (using the tool string), with the help of a genome visualization graph (see Fig. 2f, only the largest connected component is displayed, 40 genes). To obtain this network graph from STRING database, we only considered experimental connections and co-expression interactions with a minimum interaction score of 0.4. In the network graph, each node corresponds to a gene and color was mapped to the correlation between WWP2 and each gene (the average correlation in DCM and RTOF patients). Genes annotated in STRING with Gene Ontology Cellular Component (CC) term "extracellular matrix" were also retrieved and highlighted with thicker node border.

**Animal studies.** Mice were bred and maintained in a specific pathogen-free (SPF) environment. We complied with all relevant ethical regulations according to the guidelines Advised by the National Advisory Committee on Animal Research. Protocol with IACUC number 2016/SHS/1170 was approved by Institutional Animal Care and Use Committee of National University of Singapore, Duke-NUS Medical School. Steps were taken to minimize animal suffering. WWP2Mut/wt mouse generation: WWP2Mut/wt mice were generated by CRISPR/Cas9 genome editing using the CRISPR_PCRP kit (Qiagen) to insert a PGKNeo cassette in the 3’UTR of the WWP2 gene. Transgenic founder males were obtained by mating WWP2Mut/wt mice to CAST/Ei females. Targeting of the first exon of the WWP2 gene was confirmed by PCR amplification with the TGWT-WT and the TGWT-MUT primers. The presence of a 105 bp amplicon indicated successful targeting. All founder male mice were bred to C57BL/6J mice to establish a WWP2Mut/wt colony. Mice were weaned at 3 weeks of age and housed in standard conditions. Housing consisted of four to six mice per cage with food and water available ad libitum. All procedures were carried out in accordance with institutional guidelines and with approval from the Animal Care and Use Committee of National University of Singapore. Brieﬂy, mutant animals were generated by co-injection of Cas9 mRNA and individual gRNAs into one-cell mouse embryos. Founder animals carrying the indel mutations were identiﬁed ﬁrst by PCR and T7 endonuclease I assay, and then by deep sequencing of the PCR products. Founders carrying the desired reading frame shift mutations were used to generate mutation-segregated heterozygous F1 animals by crossing with the wild-type animals. Homozygous mutant animals were generated by heterozygote crossing and used for experiments in comparison with the wild-type littermates.

A customized script was written to look into the specific function of individual Wwp2 isoforms, three gRNAs were designed to target coding exon 2, aiming to introduce mutations in individual domains of the protein, which in humans have different functions and are encoded by the three different gene isoforms. Here, a reading frame shift mutation in exon 2 would render Wwp2-C function (see Supplementary Fig. 2). Wwp2Mut/wt mice were crossed to generate Wwp2Mut/Mut and Wwp2Mut/wt (WT) mice in vivarium at Agency for Science, Technology and Research (A*STAR), Singapore. Brieﬂy, mutant animals were generated by co-injection of Cas9 mRNA and individual gRNAs into one-cell mouse embryos. Founder animals carrying the indel mutations were identiﬁed ﬁrst by PCR and T7 endonuclease I assay, and then by deep sequencing of the PCR products. Founders carrying the desired reading frame shift mutations were used to generate mutation-segregated heterozygous F1 animals by crossing with the wild-type animals. Homozygous mutant animals were generated by heterozygote crossing and used for experiments in comparison with the wild-type littermates.

**Correlation analyses between WWP2 and hECM-network genes.** WWP2 transcript levels in the LV cohorts (DCM/control) and RV cohorts (RTOF/control) were compared with the patients and control groups and hECM-network. We computed the Spearman’s ranked correlation and P value using the WGCNA R package function corAndPvalue WGCNA 1.4. The resulting correlation distributions were plotted as a density plot and tested for differences by using a two-sample non-parametric Mann–Whitney U test by using the R function wilcox.test. The density plot of the obtained P values is presented in Fig. 3. The correlation between WWP2 and all the genes in the hECM-network in LV DCM and RV rTOF patients is included in Supplementary Data 3e. Nominal WWP2–gene correlation P values were corrected for multiple testing by using the R function p.adjust (method "fdr"). In this multiple testing correction, we corrected for the number of genes (563) on which the study was performed. The resulting correlation to WWP2 was extracted by taking all the genes with correlation FDR < 0.01 in both DCM and RTOF patients (326 genes). STRING protein–protein interaction database 10.0 (https://string-db.org) was queried with this set of genes to construct a protein–protein interaction network, which was retrieved and visualized using Cytoscape 2.8.1 (https://cytoscape.org/) to generate a network visualization graph (see Fig. 2f, only the largest connected component is displayed, 40 genes). To obtain this network graph from STRING database, we only considered experimental connections and co-expression interactions with a minimum interaction score of 0.4. In the network graph, each node corresponds to a gene and color was mapped to the correlation between WWP2 and each gene (the average correlation in DCM and RTOF patients). Genes annotated in STRING with Gene Ontology Cellular Component (CC) term “extracellular matrix” were also retrieved and highlighted with thicker node border.
closed with 6-0 coated vicryl suture and mice were followed for 4 weeks after surgery.

For both models, mice were sacrificed after weighing them at indicated time points. Hearts were harvested for weight measurement, histological studies, collagen determination, and molecular biology analyses. Samples sizes were determined by power analysis and n ≥ 8 mice per group were used to account for the inherent variability in the fibrotic response of mice. Mice that died undergoing surgery before the sample collection were excluded from statistical analysis. Data from the animal studies were collected in a blinded manner.

**Cell culture experiments.** Murine cardiac fibroblasts were taken from the LV of mice. Minced LV pieces (1–3 mm³) were placed in 6 cm dishes with DME/F12 supplemented with 20% fetal bovine serum for less than 10 days to generate mouse cardiac fibroblasts (P0) and passed to P1 and P2 DME/F12 supplemented with 10% fetal bovine serum in 24-well plates. In each experiment, all the cells from WWP2Mut/Mut and WWP2wt/wt (WT) harvested at the same time with same generation. Human cardiac fibroblasts were isolated from the right atrium (RA) appendage obtained from patients on cardiopulmonary-bypass during cardiac surgery operations by digesting the tissue with Collagenase II. Cardiac fibroblasts are obtained by growing the homogenized tissue suspended in DME/F12 supplemented with 20% fetal bovine serum in a humidified atmosphere. However, fibroblast-to-myofibroblast conversion occurs with each cell passage using a common cell culture method. Our primary cardiac cells isolated and cultured presented myofibroblasts features, and thus we refer to them as “(myo)fibroblasts”.

C2C12 mouse myoblast cell line, gifted by Dr. Lisa Tucker Kelloggs laboratory, Duke-NUS Medical School, Singapore, were grown in DME/F12 supplemented with 10% fetal bovine serum. The cells were passaged twice before being used for experiments.

To mimic the in vivo cardiac (myo)fibroblast activity, cells were treated with TGFβ1 (Sigma Aldrich, #77091) at a concentration of 5 ng/ml and 16–72 h. For siRNA and plasmid transfection, primary cardiac (myo)fibroblasts were seeded on a six-well plate (~70%) and were transiently transfected with siRNA duplexes (20 nM) designed for targeting β1 human (Sigma Aldrich, #77091) at a concentration of 5 ng/ml and 16–72 h. For siRNA and plasmid transfection, primary cardiac (myo)fibroblasts were seeded on a six-well plate (~70%) and were transiently transfected with siRNA duplexes (20 nM) designed for targeting 5’ or 3’ in WWP2 mRNA (Qiagen) using Lipofectamine RNAiMAX (Life Technologies holdings, #13778075) in a serum-free medium for 72 h according to the manufacturer's instructions. In parallel, WWP2+/+ and WWP2−/− mice were transiently treated with Lipofectamine 2000 (Life Technologies Holdings, #11668019) for 48–72 h according to the manufacturer’s instructions. For siRNA transfection in human cardiac (myo)fibroblasts, primary human cardiac (myo)fibroblasts were seeded on a 12-well plate (70,000 cells/well) and were transiently transfected with siRNA duplexes (20 nM) designed for targeting 5’ or 3’ in WWP2 mRNA (Qiagen) using Lipofectamine RNAiMAX (Life Technologies holdings, #13778075) in a serum-free medium for 24 h according to the manufacturer’s instructions.

For ubiquitination analysis, cells were treated with proteasomal inhibitor MG132 (Sigma Aldrich, #M7449) following stimulation with TGFβ1 for 4 h before harvesting. SB431452 ( Stem Cell Technologies, #72234) was used to inhibit TGFβ1 effect.

**Echocardiography analysis.** Transthoracic echocardiography was performed on day 28 after AngII infusion and MI model using Vevo 2100 (VisualSonics, VSI, Toronto, Canada) and a M5400 linear array transducer, 18- to 30-MHz for anesthetized condition. An average of 10 cardiac cycles of standard two dimension (2D) was acquired and stored for subsequent analysis using Vevo Imaging Workstation version 11.1 (VisualSonics, VSI, Toronto, Canada), VisualSonics, VSI, Toronto, Canada. VTI, area, and ejection fraction were measured using EchoPAC, GE Healthcare. All measurements acquisition and analysis were performed by a blinded operator. For the AngII infusion model, 2-D guided M-mode of parasternal short-axis short (middle) were selected for visualization of the papillary muscle during end systole and end diastole. For the MI model, the parasternal long axis was analyzed at three levels selected for visualization of the papillary muscle during end systole and end diastole and LVIDes is left ventricular internal diameter at end systole.

**Single-cell RNA-sequencing analysis in mouse heart.** Single-cell suspension was prepared from the adult LV of one mouse with Angiotensin II infusion for 28 days. After removal of dead cells with MACS Strepavid (Miltenyi Biotec, #130-090-101), cells were lysed and subsequently RNA was reverse-transcribed and converted into cDNA libraries for RNA-seq analysis using a Chromium Controller and a Chromium Single Cell 3’ v2 Reagent kit (Genomics 10x) following the manufacturer’s protocol. The library was sequenced using the Illumina HiSeq3000 sequencing platform.

The reads were mapped to the mouse genome (mm38, Ensembl version 89) and quantified using Cell Ranger 2.1.10 (10X Genomics). We provided to Cell Ranger a custom built reference transcriptional set generated by filtering the Ensemble transcriptome (Ensembl file: Mus_musculus.GRCm38.89.gtf) for the gene prototypes: lower case, Idiotype, and Antigen. Genebody was run with the number of cells parameter (expect-cells) set to 3000. Cell Ranger output filtered matrices (i.e., genes.tsv and barcodes.tsv) were then input into R and genes with zero counts in all cells were discarded. Three cell quality control filtering steps were implemented. We removed: (a) cells with less than the 50th percentile of the distribution of total of number of detected genes, (c) cells with more than 50% of their total gene count coming from mitochondrial genes. This resulted in a final number of 508 cells. After applying these cell filtering steps, we carried out additional gene quality control steps on the remaining cells: (a) we only kept detection signals derived from more than one transcript in at least two cells, (b) we removed genes with low average expression in the data (i.e., genes with an average expression below 0.01, this cutoff was set based on the total distribution of average gene expression across all cells and all genes), (c) we removed genes encoded on the mitochondrial genome and the gene “Mala1” as it was an outlier in the gene expression distribution.

After all the gene quality control steps, the resulting number of genes was 6,728. Gene counts were normalized with scanr 1.8.4R package [https://bioconductor.org/packages/3.4/bioc/html/scanr.html]. Scanr size factors were computed from cell pools by doing a pre-clustering of the data with the quickCluster function (the output object of this function was provided to the computeSizeFactors function and then run the normalization function was run, all of them with default parameterizations). t-SNE was computed by providing the Log2, scanr-normalized data to the function plotTSNE from scater 1.8.4R package (https://pdbeontology.rkenj.org/packages/3.4/bioc/html/scater.html). Ribosomal genes, genes with no gene body and mitochondrial genes were removed (391 genes). Gene counts were rounded using the R function round and differential expression analysis was performed with DESeq2 1.14.1 with a pre-filtering step in which we considered only genes with more than 1 count when summing up across all samples. DESeq2 was run pairwise comparing WWP2Mut/Mut with Angiotensin II treated WWP2Mut/Mut and WWP2wt/wt mice with the Angiotensin II treated WWP2wt/wt mice and the Wald test, with a false discovery rate correction parameter cooksCutoff set to false (default parameterizations for the rest of parameters). In the DESeq2 model, we added RNA concentration and sequencing lane as covariates. Functional enrichment analysis of the differential expression results was performed with GSEA software v 2.2.2.2. From all genes included in DESeq2 output, we selected those with one-to-one mouse–human ortholog relationships (as downloaded from Biomart, 14,820 genes) and then we mapped them to human gene symbols. Then ranked all the genes by the corresponding DESeq output Wald statistic (i.e., the estimate of the log2 fold change divided by its standard error). GSEA was run to assess the overrepresentation of one-on-one human infused pathways derived from the 391-gene sets Database gene sets 5.1 [http://software.broadinstitute.org/gsea/msigdb/collections.jsp] (gene sets were queried using gene symbols): Hallmark gene sets (i.e., coherently gene expressed signatures derived from the aggregation of many MSigDB gene sets to represent well-defined biological states or processes), Gene Ontology, and Reactome gene sets in classic pre-ranked mode. Two additional gene sets Database gene sets 5.1 were also included in the initial set of human genes considered for network inference in the human data. This background reduction was applied to both the DESeq2 output mouse and in the initial set of human genes considered for network inference in the human data. This background reduction was applied to both the ranked list and the human networks, resulting in a total number of 10,449 genes. Results of this test are included in Supplementary Data 4.

Analysis of WWP2Mut/Mut mouse RNA-seq data. Total RNA from tissue was isolated from LVs of nine WWP2Mut/Mut and nine WT mice with 28 days of AngII infusion. mRNA libraries were constructed from poly(A)+ selected RNA using the NEBNext Ultra Directional RNA library prep kit (Illumina, New England BioLabs) and sequenced on Illumina Hi-Seq3000 sequencing (150 × 2 bp, seven samples were loaded per lane).

RNA-seq reads were assessed for quality, aligned to m38 (Ensembl Gene annotation build 89) using STAR 2.5.2a [https://github.com/alexdobin/STAR] and quantified with RSEM 1.2.31 (https://github.com/deweylab/RSEM/). The average mapping rate (unique and multimapping) was 94.5%. Gene annotation was retrieved from Ensembl version 89 (m38) using the R library bioRama 2.30.0 [https://bioconductor.rkenj.org/packages/3.4/bioc/html/biorama.html]. Ribosomal gene and one-to-one mouse–human ortholog relationships (as downloaded from Biomart, 14,820 genes) and then we mapped them to human gene symbols. Then ranked all the genes by the corresponding DESeq2 output Wald statistic (i.e., the estimate of the log2 fold change divided by its standard error). GSEA was run to assess the overrepresentation of one-on-one human infused pathways derived from the 391-gene sets Database gene sets 5.1 [http://software.broadinstitute.org/gsea/msigdb/collections.jsp] (gene sets were queried using gene symbols): Hallmark gene sets (i.e., coherently gene expressed signatures derived from the aggregation of many MSigDB gene sets to represent well-defined biological states or processes), Gene Ontology, and Reactome gene sets in classic pre-ranked mode. Two additional gene sets Database gene sets 5.1 were also included in the initial set of human genes considered for network inference in the human data. This background reduction was applied to both the DESeq2 output mouse and in the initial set of human genes considered for network inference in the human data. This background reduction was applied to both the ranked list and the human networks, resulting in a total number of 10,449 genes. Results of this test are included in Supplementary Data 4.
previously did to compute differential co-expression of the hECM-network genes in the two human cohorts (DCM/controls and rTOF/controls, see previous sections). After transfection and removal of nonspecific binding step and transfection efficiency using Quanti-Fluor (Molecular Probes), cells were treated with vehicle or TGF-β1 (Sigma-Aldrich, #H8980) or TGF-β1 (Yeasen, #11557ES03) as a normalization control. Thirty hours after transfection, cells were evaluated for luciferase expression using a Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer’s protocol. Western blotting. Protein extracts were isolated from heart tissue and cells using RIPA buffer (ThermoFisher, #89900) supplemented with protease (Sigma-Aldrich, #A238410001) and phosphatase inhibitors cocktails (ROCHE, #PHOS-RO) according to the manufacturer’s instructions. Co-immunoprecipitation was performed with the cell lysates subjected to different treatment conditions with Pierce Direct Magnetic IP/CO-IP kit (Pierce, #88828) according to the manufacturer’s protocol. Immunoprecipitates were washed from conjugated beads and boiled in 5× SDS-PAGE buffer for further WB analysis. After quantification with the Bradford method, protein lysates were loaded onto a 4-12% acrylamide gel subjected to SDS-PAGE and then transferred onto a nitrocellulose membrane. After blocking in 5% non-fat dry milk, blotting was performed with antibodies against target antigens including anti-FLAG (Sigma, #F7425, 1:1000), anti-p-SMAD2 (CST, #9539, 1:500), anti-SMAD2/3 (CST, #9550, 1:500), anti-Acta2 (Santa Cruz Biotechnology, #sc-7071, 1:500), anti-GAPDH (Abcam, #ab18161, 1:10,000), anti-Periostin (Novus Bio, #NPB1-3001, 1:200), anti-Vimentin (Abcam, #ab9245, 1:5000), anti-Lamin A/C (Abcam, #ab89841, 1:5000) and anti-β-tubulin (Abcam, #ab60791, 1:5000) were used as nuclear controls. Blots were visualized by labeling with anti-Rabbit HRP (Bethyl Laboratories, #A200-11P, 1:5000 or Thermo Fisher #101023, 1:10000) and anti-Mouse HRP (Bethyl laboratories, #A90-116P, 1:5000) and developed using a ChemiDoc imaging system (Bio-Rad). Densitometry was performed using ImageJ software with the Fiji package. The unprocessed scans of the western blots are shown in Supplementary Fig. 18.

**Histology and immunofluorescence.** LVs harvested from the mice were fixed in 10% neutral buffered formalin (NBF) for 24 h at RT, processed with a Leica automatic tissue processor, paraffin-embedded and sectioned with thickness of 5 μm. After dewaxing and rehydration, slides were stained with Sirius Red Collagen kit (Chondrex, Inc., #9046) and Masson’s Trichrome staining kit (Sigma-Aldrich, #HT1J5) as per the manufacturer’s instructions. Sections were stained using anti-ACTA2 (1:100) and anti-S100A4 (1:100) to identify cell and biochemical features. Bovine Anti Rabbit IgG-CFL 488 (Santa Cruz Biotechnology, #sc-362256) and Bovine Anti Mouse IgG-CFL 488 (Santa Cruz Biotechnology, #sc-362256) were used as secondary antibodies for immunofluorescence. Rhodamine Wheat Germ Agglutinin (WGA, Vector laboratories, #RL-1022) was used to stain the myocytes. Cells were grown in an eight-well chamber slide with a removable silicone chamber (ibidi). After fixation with 4% paraformaldehyde in PBS and blocking with 1% BSA for 30 min at RT, the slides were incubated with primary antibodies against WW2P2-FL/N (1:100), anti-ACTA2 (1:100), anti-S100A4 (1:100), anti-Vimentin (1:100), and anti-Flag (1:100) overnight at 4 °C. Following washing steps, the slides were incubated with Bovine Anti Rabbit IgG-CFL 488 (Santa Cruz Biotechnology, #sc-362256) for 2 h at RT. VectaShield Mounting Medium (Vector laboratories, #H-1200) with DAPI was used to stain the nuclei and the slides were covered by coverslip. Slides were imaged on a Leica fluorescence microscope and image was processed using ImageJ software with the Fiji package. Sections were stained using anti-ACTA2 (1:100) and anti-S100A4 (1:100) to identify cell and biochemical features. After dewaxing and rehydration, slides were stained with Sirius Red Collagen kit (Chondrex, Inc., #9046) and Masson’s Trichrome staining kit (Sigma-Aldrich, #HT1J5) as per the manufacturer’s instructions. Sections were stained using anti-ACTA2 (1:100) and anti-S100A4 (1:100) to identify cell and biochemical features. Bovine Anti Rabbit IgG-CFL 488 (Santa Cruz Biotechnology, #sc-362256) and Bovine Anti Mouse IgG-CFL 488 (Santa Cruz Biotechnology, #sc-362256) were used as secondary antibodies for immunofluorescence. Rhodamine Wheat Germ Agglutinin (WGA, Vector laboratories, #RL-1022) was used to stain the myocytes. Cells were grown in an eight-well chamber slide with a removable silicone chamber (ibidi). After fixation with 4% paraformaldehyde in PBS and blocking with 1% BSA for 30 min at RT, the slides were incubated with primary antibodies against WW2P2-FL/N (1:100), anti-ACTA2 (1:100), anti-S100A4 (1:100), anti-Vimentin (1:100), and anti-Flag (1:100) overnight at 4 °C. Following washing steps, the slides were incubated with Bovine Anti Rabbit IgG-CFL 488 (Santa Cruz Biotechnology, #sc-362256) for 2 h at RT. VectaShield Mounting Medium (Vector laboratories, #H-1200) with DAPI was used to stain the nuclei and the slides were covered by coverslip. Slides were imaged on a Leica fluorescence microscope and image was processed using ImageJ software with the Fiji package. With merge function, the positive Sirius red staining in the whole section of the LV was quantified using custom semi-automated image analysis routine. To measure the fluorescence intensity of ACTA2 and COL1A1 in different cellular sections, sections were taken using ×20 Plan Fluor objective. Fluorescence intensity was measured by taking the integrated intensity of a region of interest and subtracting the background intensity, and normalized to cell number. For each group, at least six fields were analyzed per section. To measure fluorescence intensities of SMAD2, 1-μm Z-stacks through cells of fields interested were acquired. A region was drawn around each cell and nucleus to be measured, and background without fluorescence was subtracted. The nuclear/cellular fluorescence intensity ratio was calculated. Each field represented around 8–10 cells and at least four fields were analyzed for each section.

**Hydroxyproline assay.** The amount of total collagen in the LV was quantified using the Quicksilver Total Collagen assay kit (Quicksilver Biosciences). The assays were performed according to the manufacturer’s protocol.

**Luciferase assay.** Cells were transfected with a luciferase reporter gene plasmid with SMAD binding sites (Yeasen, #11543ES03) and co-transfected with SGMLR-TK (Yeasen, #11557ES03) as a normalization control. Thirty hours after transfection, cells were treated with vehicle or TGFβ1 for 16 h and harvested. Luciferase assays were performed using the Dual-Luciferase Reporter Assay System (Promega, #11402ES60).

**Cell proliferation and migration assays.** Cell proliferation was quantified by MTS assay (Promega) according to the manufacturers’ protocol. For migration assay, cells were seeded at a density of 10,000 cells/well in a 96-well plate. A uniform, 5 μm thick cell-free collagen matrix was created using Incucyte, Essen Biosciences (USA). The 96-well plate was placed in the Incucyte ZOOM apparatus and the images of cell migration was captured every 2 h for up to a total of 48 h.

**RT-qPCR analysis.** Total RNA was extracted from snap-frozen fibrotic cardiac tissue and primary cardiac myocytes using the RNeasy mini kit (#Qiagen, #74106) and cDNA was prepared using the iScript cDNA synthesis kit (primeScript, BIORAD, #170-8897) according to the manufacturer’s instructions. Fast SYBR-Green master mix (BIORAD, #170-8880AP) was used for the analysis of gene expression using the BIORAD CFX RT-PCR system. The primers used in the experiment are listed in Supplementary Table 2. 18S was used to normalize the relative gene expression and the 2ΔΔCt method was used to measure the fold change.

**References.**

1. Travers, J. G., Kamal, F. A., Robbins, J., Yutzey, K. E. & Blaxall, B. C. Cardiac fibrosis: the fibroblast awakens. *Circ. Res.* 118, 1021–1040 (2016).

2. Gulati, A. et al. Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy. *JAMA* 309, 896–908 (2013).

3. Garfinkel, A. C., Seidman, J. G. & Seidman, C. E. Genetic pathogenesis of hypertrophic and dilated cardiomyopathy. *Heart Fail. Clin.* 14, 139–146 (2018).

4. Harvey, P. A. & Leinwand, L. A. The cell biology of disease: cellular mechanisms of cardiomyopathy. *J. Cell Biol.* 194, 355–365 (2011).

**Data availability.** All the data generated in this study supporting the main findings have been deposited to NCBI’s Gene Expression Omnibus (GEO) and accessible through GEO Series accession number GSE33017, including GSE30468 (bulk RNA-seq data from mouse heart) and GSE13105 (single-cell RNA-seq data from mouse heart). The rest of the data are available from the authors on reasonable request, please refer to author contributions for specific data sets.

**Code availability.** We used published algorithms. See Methods section for full description of each analysis including input data, library, and algorithm version used, for which we provide corresponding web links.

Received: 10 October 2018 Accepted: 19 July 2019

Published online: 09 August 2019
Acknowledgements

The research was primarily supported by National Medical Research Council (NMRC) grant 1103840054 (to E.P.) and the British Heart Foundation (BHF) PhD Studentship grant FS/11/25/28740 (to E.P.). We acknowledge additional funding support from European Union FP7 CardioNest-ITN-289600 (to E.L.-P., S.A.C., and P.J.R.B.), Heart Research UK (to P.J.R.B.), NIHR CV BRU of Royal Brompton and Harefield, NHS Foundation Trust (to S.A.C. and P.J.R.B.), BHF (to S.A.C.), Leducq – e40 (2012).

Peer review information: Nature Communications thanks Ian Dixon, Bin Zhang and other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Copyright and reprint information is available online at http://npg.nature.com/reprintsandpermissions/

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

OPEN ACCESS This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019