

RESEARCH

Open Access



Comparison of gene expression between human and mouse iPSC-derived cardiomyocytes for stem cell therapies of cardiovascular defects via bioinformatic analysis

Ryan Bellman¹, Jishizhan Chen¹, Lidan Chen^{1,2}, Nikolitsa Nomikou³, Janice Tsui¹, George Hamilton¹ and Wenhui Song^{1*} 

Abstract

Background Preclinical studies have demonstrated the potential use of induced pluripotent stem cells (iPSCs) to treat cardiovascular disease (CVD). In vivo preclinical studies conducted on animal models (murine, porcine, guinea pig, etc.) have employed either syngeneic or human-derived iPSCs. However, no study has been carried out to investigate and report the key genetic differences between the human and animal-derived iPSCs. Our study analysed the gene expression profile and molecular pathway patterns underlying the differentiation of both human and mouse iPSCs to iPSC-cardiomyocytes (iPSC-CMs), and the differences between them via bioinformatic analysis.

Method Data sets were downloaded from the Gene Expression Omnibus (GEO) database and included both human and mouse models, and the data for undifferentiated iPSCs and iPSC-CMs were isolated from each. Differentially expressed genes (DEGs) were screened and then analysed. The website g:Profiler was used to obtain the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. Protein-protein interaction (PPI) networks of the DEGs were constructed using the Search Tool for the Retrieval of Interacting Genes (STRING) database and Cytoscape software. The subclusters were then extracted from the PPI network for further analysis.

Results iPSC-derived cardiomyocytes expressed many genes related to vascular, endothelial, and smooth muscle repair in the human iPSC-CMs, and prevention of calcification in the mouse iPSC-CMs with clear differences in gene expression, which will affect how iPSCs act in research. Especially in the human iPSC-CMs, and also prevention of calcification processes in the mouse data. The identified differences in gene expression of iPSCs derived from the two species suggests that in vivo studies using mouse iPSC-CMs may not reflect those in humans.

Conclusion The study provides new insights into the key genes related to the iPSCs, including genes related to angiogenesis, calcification, and striated muscle, endothelium, and bone formation. Moreover, the clear differences between both mouse and human-derived iPSCs have been identified, which could be used as new evidence and

*Correspondence:

Wenhui Song
w.song@ucl.ac.uk

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



© The Author(s) 2023. **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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

guidance for developing novel targeted therapy strategies to improve the therapeutic effects of iPSC treatment in cardiovascular defects.

Keywords Bioinformatics, iPSCs, Cardiomyocytes, Enrichment analysis, Cardiovascular disease

Background

Cardiovascular diseases (CVDs) are the number one cause of mortality around the world with 31% of all global deaths in 2016 [1]. Several heart and blood vessel diseases are encompassed by this. Aortic stenosis is the most prevalent heart valve disease and the third most prevalent CVD in developed countries. It is characterised by progressive narrowing of the aortic valve due to fibrosis and calcification [2], producing stiffness and disruption of normal flow of blood through the valve. This leads to insufficient outflow from the ventricle, leading to left ventricular hypertrophy to compensate which eventually becomes maladaptive [3]. A brief review of valvular heart disease, including the current treatments and unmet clinical outcomes, is summarised in Table S1 in Supplementary Information. Therefore, there is a need to find alternative therapeutic options, with one possibility being stem cell therapy or cell-based tissue engineered heart valve (TEHV).

Different parts of the cardiovascular system have different anatomical arrangements, including different cells which make up their structure as well as important markers which are used to identify them (Table S2 and Fig. S1 in Supplementary Information). These are important to understand what cells are needed to form fully functional structures to treat CVDs. In order to produce the necessary cells, cell-based TEHV can help to generate biomimetic and viable tissue either before (in vitro) or after (in situ) implantation. Induced pluripotent stem cells (iPSCs) are a cell type that has been used in cell therapy and modelling of TEHVs and other CVD diseases (Table S3 in Supplementary Information). The iPSCs have several advantages over other cell types, especially in disease modelling. When iPSCs differentiate the genetic makeup of the donor cells remains, meaning there is the potential to reproduce essential aspects of genetic diseases in vitro using donor cells from patients. However, iPSCs for this purpose still carry limitations hindering their clinical translation. Firstly, the potential for cell survival and proliferation and thus any long-term benefits are poor. Secondly, iPSCs display a more fetal CM phenotype [4] which not only inaccurately replicates the electrophysiological responses but also produces a risk of arrhythmias on transplantation. Thirdly, enrichment and maintenance processes often produce a heterogeneous population of iPSC-CMs [5] which exhibit phenotypic and genomic differences. Moreover, studies have shown that iPSC-CMs

are associated with epigenetic pattern retention, producing differences in differentiation and physiological responses [6]. The potential of undifferentiated iPSCs also poses a tumorigenic risk, proving that these processes cannot account for the heterogeneity present [7].

The aim of this study was to conduct bioinformatic analysis for identifying key genetic factors and signalling pathways involved in the differentiation of iPSCs to cardiomyocytes (iPSC-CMs). We anticipate that identification of key genes and pathways will help lead to the understanding and optimisation of iPSC-based therapy for cardiovascular defects. Moreover, cells derived from different species have been widely used as replacements for studying human diseases. It has been shown that the key genes and the signaling pathways for a certain disease could vary between species, which is a common factor that confines much success to laboratories. Comparison is undoubtedly needed to minimise the inaccuracy when transforming achievements from the bench to clinical trials. We noticed the importance of this but found highly limited studies comparing at the gene level. Thus, we wish to compare the results from bioinformatic analyses between the human and mouse iPSC-CMs, and we hypothesise that the comparison will provide an understanding of iPSC-CM gene expression and guidance on potential different markers and therapeutic effects in in vivo mouse models and the human condition.

Database and method

Database searching

Gene Expression Omnibus (GEO) Database [8] is an international public archive with functional genomic data, and using this, two gene expression data sets which conducted transcriptomic profiling, GSE17579 [9, 10], and GSE18514 [11] were found and downloaded from the site. The data for undifferentiated iPSCs and iPSC-CMs were isolated from each data set. GSE17579 used a human model and was based on the GPL6947 (Illumina HumanHT-12 V3.0 expression beadchip), while GSE18514 used a murine model and was based on the GPL6997 platform (Illumina MouseWG-6 v2.0 expression beadchip). Ribonucleic acid (RNA) samples were prepared from three independent biological replicates for both undifferentiated iPSCs and iPSC-CMs for each dataset.

The human iPS cell line was derived by the data set GSE17579 from foreskin fibroblasts, and the human ES

cell line HES-2 were differentiated to CMs using the END2 co-culture system. They were experimented on day 18 of their development. The data set GSE18514 developed the mouse iPSC cell line using a transgenic murine iPSC cell line “which expresses puromycin resistance protein N-acetyltransferase and EGFP under the control of the cardiomyocyte-specific α -myosin heavy chain promoter (alphaMHC-Puro-IRES-GFP, aPiG)” [11]. Murine aPiG-iPS and aPiG-ES cells differentiated into spontaneously beating CMs and were analysed on day 16 of their development.

Identification of DEGs

Network analyst [12] is a website used for gene expression analysis which was used to identify differentially expressed genes (DEGs) in the samples collected. The data was firstly normalised in Network Analyst using a log₂ transformation, and an adjusted *p*-value of 0.01 and log₂ fold change of 2.0 were used as thresholds to identify DEGs. Using the matrices downloaded from GEO, significant gene matrices of the DEGs were produced in Microsoft Excel. The DEGs were then converted into their equivalent official gene symbols, and their log₂ fold changes calculated from their count values so heatmaps could be produced using a program called TBtools [13].

GO and KEGG pathway enrichment analysis of DEGs

g:Profiler [14] was used to perform pathway enrichment analysis of the DEGs using their official gene symbols. Gene ontology (GO) is a data source which focuses on the functions of genes and the GO terms are divided into molecular function (MF), biological processes (BP) and cellular component (CC). Moreover, Kyoto Encyclopedia of Genes and Genomes (KEGG) terms focus on the biological pathways of the DEGs. Together they can provide overall functional profiling of the DEGs.

The input parameter of either *homo sapiens* or *mus musculus* could be chosen depending on the data set analysed. g:Profiler has a few threshold algorithms to enrich terms, the one used was the g:SCS method to correct for the multiple testing problem. This occurs in pathway enrichment analysis of genes due to the high numbers of alternative hypotheses being tested at once (whether the GO and KEGG terms are significant), causing a higher chance of false positive results. Therefore, this algorithm was used as it provides a more accurate threshold between significance and non-significance [14]. A *p*-value threshold of < 0.05 was used to screen for the statistically significant results.

PPI network construction and functional subcluster analysis

Protein-protein interaction (PPI) networks were produced using a website called Search Tool for the Retrieval of Interacting Genes (STRING) [15] and a program called Cytoscape (Version 3.8.2) [16], a software for visualising molecular interactions and biological pathways. The PPI networks were created using STRING and imported into Cytoscape, where they were annotated and edited as appropriate. The add-on MCODE was also downloaded through Cytoscape, and this was used to produce subclusters of the densely connected parts of the networks. The genes in these subclusters also underwent pathway enrichment analysis to find their GO and KEGG terms.

Analysing the GO and KEGG terms and up- or down-regulation of genes

Using g:Profiler, the genes related to the GO and KEGG terms could be retrieved. To confirm if they were up- or down-regulated, this was done manually by locating these genes from the significant gene's matrices, produced from the GEO matrices. Their log₂ fold changes were then calculated to find if they were up- or down-regulated in the iPSC-CMs compared to the undifferentiated iPSCs. Using this data, they were then ranked to show which genes had the highest and lowest log₂ fold changes, and which genes were up- or down-regulated.

Results

Identification of DEGs

The normalisation box plots after log₂ normalisation for both the human and mouse data sets show that the data has been correctly normalised with an appropriate range of values for data analysis, as shown in Fig. 1.

The heatmaps produced by TBtools are shown in Fig. 2. By undergoing differential expression analysis, in GSE17579 (human model) 471 DEGs were identified (Fig. 2a), and in GSE18514 (mouse model) 972 DEGs were identified (Fig. 2b). They show the DEGs which are significant at the threshold of a log₂ fold change of 2.0. These verify that the undifferentiated iPSCs and the iPSC-CMs have different gene expression profile compared to one another for the DEGs. The volcano plots (Fig. 3) also show the non-significant genes and significant up- and down-regulated genes.

GO and KEGG pathway enrichment analysis of DEGs

Table 1 shows the top 10 ranked GO and KEGG terms for the DEGs, for both the human and mouse models, produced in the pathway enrichment analysis from g:Profiler. Looking at the human data set (Table 1a), enrichment analysis showed that: for MF, DEGs were most enriched

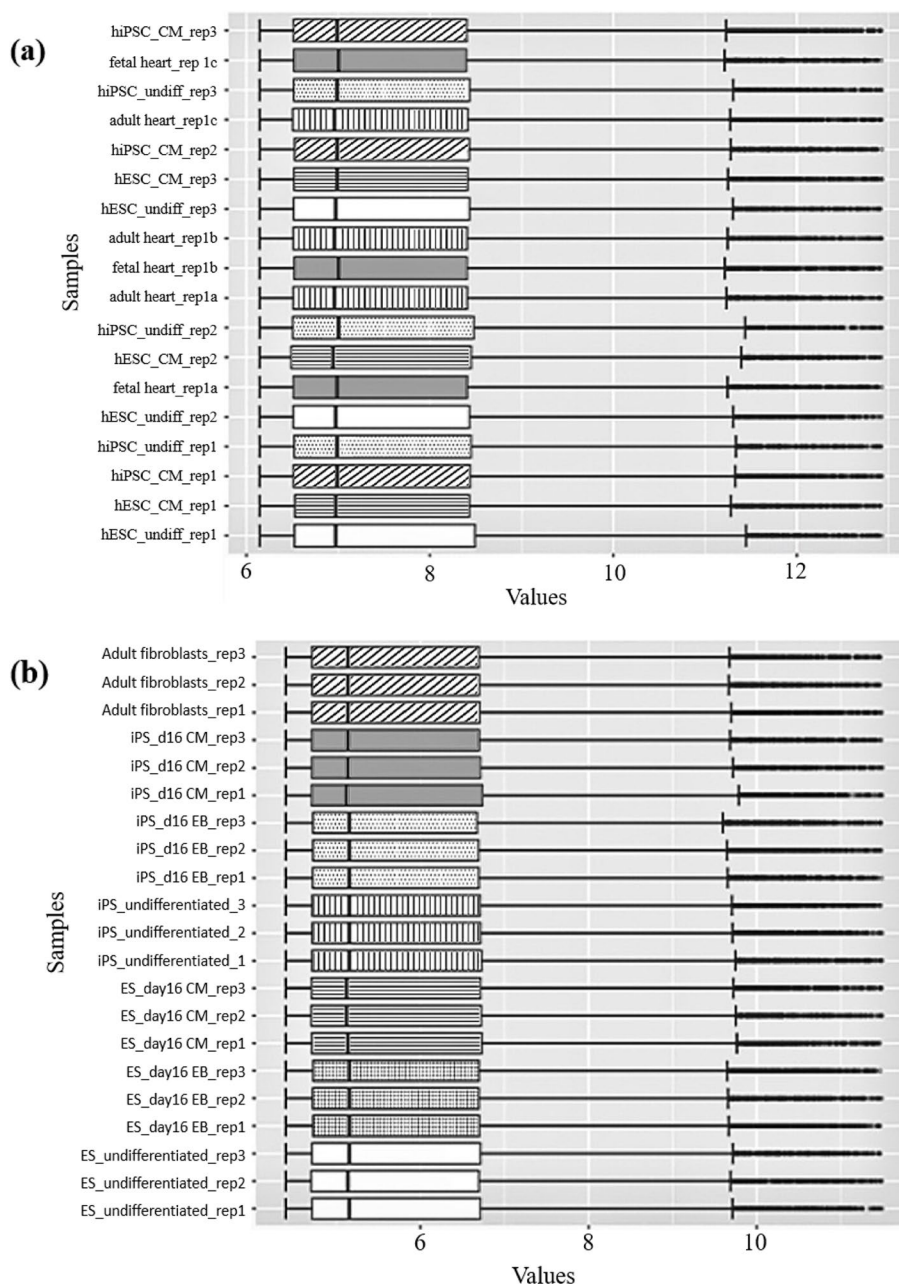


Fig. 1 Normalisation box plots- Box plots of normalised data sets of iPSCs and iPSC-CMs produced after the data was normalised in Network Analyst using a log2 transformation. Black lines in the boxes represent the median values. **a** the human data set, GSE17579. **b** mouse data set, GSE18514

in relation to extracellular matrix (ECM) and muscle structure, and binding of proteins, growth factors, signalling receptors, platelet-derived growth factor, actin, and glycosaminoglycans; for BP, DEGs were most enriched in relation to the development of the circulatory system, heart, and other tissues and organs, and developmental and system processes; for CC, DEGs were most enriched in relation to structural components of the ECM and

muscle; for KEGG, DEGs were most enriched in relation to ECM-receptor interactions, focal adhesion, signalling pathways of TGF-beta, PI3K-Akt, regulation of pluripotency of stem cells, and cardiac muscle contraction.

In comparison, pathway enrichment analysis in the mouse data set in Table 1b showed that: for ME, DEGs were most enriched in relation to exo-alpha- and alpha-sialidase activity and protein tyrosine/serine/threonine

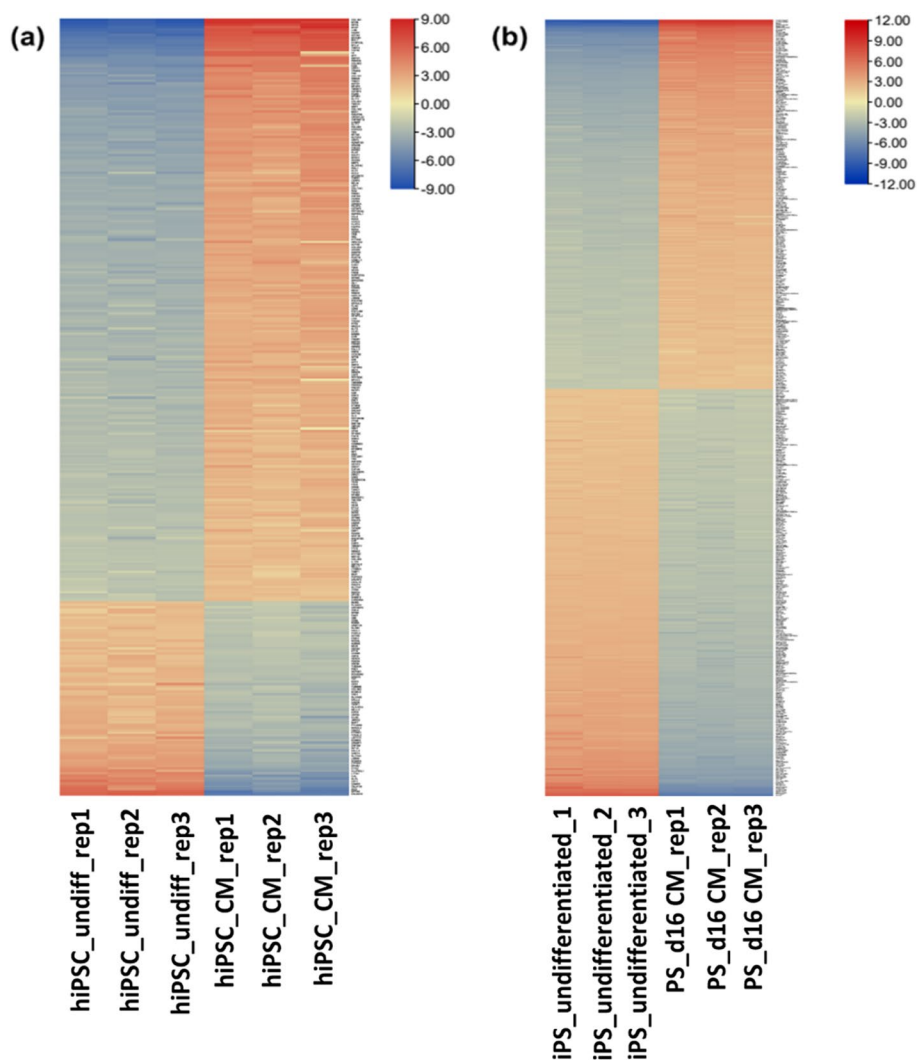


Fig. 2 Heatmaps- Heatmaps which show hierarchical clustering analysis results of the DEGs from both the mouse and human data matrices and show those genes which are significant at a threshold of a \log_2 fold change of 2.0. Each column represents a sample, and each row represents a DEG. Blue shows a lower value for gene expression while red shows a higher value for gene expression. **a** the human data set, GSE17579. **b** mouse data set, GSE18514

phosphatase activity; for BPs, DEGs were most enriched in relation to ganglioside catabolic, oligosaccharide catabolic, and nervous system processes; for CC, DEGs were enriched in relation to intermediate filaments and the intermediate filament cytoskeleton; for KEGG, DEGs were enriched in relation to the estrogen signaling pathway.

Table 2 shows a table of the shared DEGs between the two data sets, allowing them to be compared with one another. There are genes which are involved: in cardiomyocyte formation, including *TMOD1*, which encodes a member of the tropomodulin family and has roles in regulating the organisation of actin filaments; in angiogenesis, including *VEGFC* and *VASH2* (as well as *VEGFC*

being involved in endothelial formation); bone formation and remodelling, including *FRZB* and *SPP1* respectively; oxidation of low-density lipoproteins, including *OLR1*; stem cell proliferation and renewal, including *NANOG*.

PPI network construction and functional subcluster analysis

Figure 4 shows the PPI networks of iPSC-CMs produced from Cytoscape, and Table 3 shows the top 10 hub genes present in each of the PPI networks. The PPI network produced from the human data set (Fig. 4a) is composed of 398 nodes and 2271 edges, and the PPI network from the mouse data set (Fig. 4b) is composed of 663 nodes and 1780 edges.

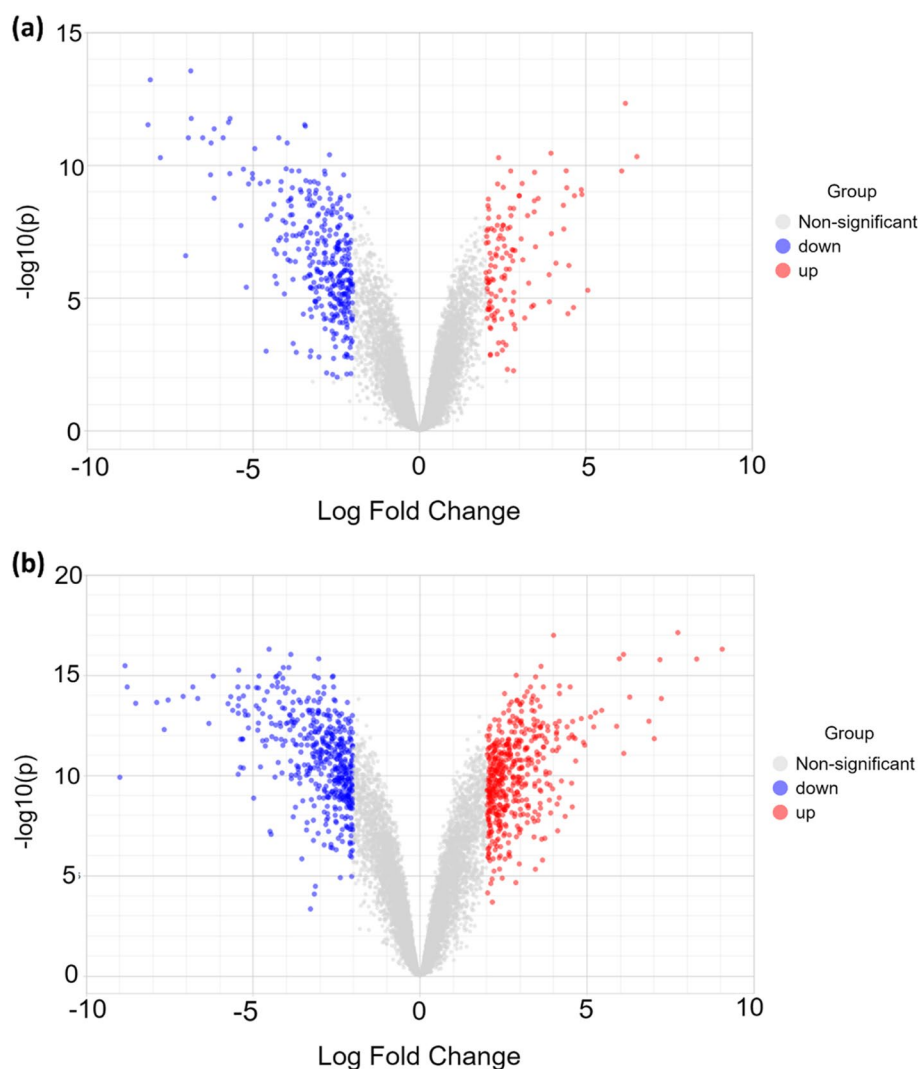


Fig. 3 Volcano plots: Volcano plots of all genes in the data sets, with each dot representing an individual gene. Blue dots show the genes which are downregulated, red dots show the genes which are upregulated, and the grey dots are genes which are non-significant. The X-axis is the log₂-base fold change, and the Y-axis is the -log₁₀-base adjusted *P*-value. **a** the human data set, GSE17579. **b** mouse data set, GSE18514

The top 3 subclusters for both the human and mouse networks, which included hub genes, were retrieved, and can be seen in Fig. 5. For each subcluster, their top 5 genes with the highest coexpressions are shown in Table 4. Pathway enrichment analysis was also undertaken on the genes in these subclusters to find their GO and KEGG terms, and these can be seen in Table 5.

The top 3 human subclusters contain: Human a (Ha)- 40 nodes and 285 edges, human b (Hb)- 26 nodes and 57 edges, human c (Hc)- 10 nodes and 19 edges (Fig. 6a). Pathway enrichment analysis in these subclusters showed that enriched genes were related to: Ha- ECM structure, muscle contraction, and myopathies; Hb- structural,

signalling receptor and growth factor binding, developmental processes, platelets, diabetes pathway, and focal adhesion; Hc- protein ubiquitination, sarcomeres, myofibrils, and contractile fibres. The top 3 mouse subclusters contain: Mouse a (Ma)- 13 nodes and 78 edges; mouse b (Mb)- 14 nodes, and 44 edges, mouse c (Mc)- 55 nodes and 178 edges (Fig. 5b). Pathway enrichment analysis showed that the enrichments were related to: Ma- GPCR and transmembrane receptor activity, calcium and chemical ion homeostasis, and neuron/axon/ligand-receptor interactions; Mb- RNA/nucleic acids, RNA and mRNA processing and spliceosomes and RNA splicing; Mc- DNA repair, cell cycle and chromosomes.

Table 1 Top 10 ranked GO and KEGG terms for DEGs for a) human or b) mouse data sets

Category	GO ID	Description	Gene Count	P-value
a. Human top 10 ranked GO and KEGG terms				
GO:MF	GO:0005201	Extracellular matrix structural constituent	31	2.934×10^{-16}
GO:MF	GO:0005198	Structural molecule activity	56	5.755×10^{-13}
GO:MF	GO:0005515	Protein binding	405	3.072×10^{-10}
GO:MF	GO:0019838	Growth factor binding	22	6.965×10^{-10}
GO:MF	GO:0005102	Signalling Receptor binding	85	6.875×10^{-9}
GO:MF	GO:0008307	Structural constituent of muscle	12	1.147×10^{-7}
GO:MF	GO:0048407	Platelet-derived growth factor binding	7	6.545×10^{-7}
GO:MF	GO:0003779	Actin binding	33	4.224×10^{-6}
GO:MF	GO:0030020	Extracellular matrix structural constituent conferring tensile strength	10	1.576×10^{-5}
GO:MF	GO:0005539	Glycosaminoglycan binding	22	2.948×10^{-5}
GO:BP	GO:0072359	Circulatory system development	125	4.102×10^{-42}
GO:BP	GO:0009888	Tissue development	164	3.639×10^{-40}
GO:BP	GO:0009653	Anatomical structure morphogenesis	190	1.027×10^{-39}
GO:BP	GO:0007507	Heart development	87	1.102×10^{-39}
GO:BP	GO:0048856	Anatomical structure development	286	3.192×10^{-36}
GO:BP	GO:0048513	Animal organ development	213	4.501×10^{-35}
GO:BP	GO:0007275	Multicellular organism development	269	3.918×10^{-34}
GO:BP	GO:0032502	Developmental process	295	1.495×10^{-33}
GO:BP	GO:0048731	System development	251	5.654×10^{-33}
GO:BP	GO:0032501	Multicellular organismal process	325	1.979×10^{-32}
GO:CC	GO:0031012	Extracellular matrix	65	4.532×10^{-24}
GO:CC	GO:0062023	Collagen-containing extracellular matrix	52	4.427×10^{-20}
GO:CC	GO:0043292	Contractile fiber	40	7.417×10^{-20}
GO:CC	GO:0030016	Myofibril	36	7.681×10^{-17}
GO:CC	GO:0030017	Sarcomere	34	1.992×10^{-16}
GO:CC	GO:0099081	Supramolecular polymer	69	6.482×10^{-13}
GO:CC	GO:0099512	Supramolecular fiber	68	1.488×10^{-12}
GO:CC	GO:0005788	Endoplasmic reticulum lumen	35	5.127×10^{-12}
GO:CC	GO:0031674	I band	23	1.046×10^{-10}
GO:CC	GO:0099080	Supramolecular complex	75	7.206×10^{-10}
KEGG	KEGG:04512	ECM-receptor interaction	15	2.194×10^{-6}
KEGG	KEGG:04510	Focal Adhesion	20	9.739×10^{-5}
KEGG	KEGG:04350	TGF-beta signaling pathway	13	2.084×10^{-4}
KEGG	KEGG:05414	Dilated cardiomyopathy	13	2.664×10^{-4}
KEGG	KEGG:04550	Signaling pathways regulating pluripotency of stem cells	16	2.869×10^{-4}
KEGG	KEGG:04260	Cardiac muscle contraction	12	5.994×10^{-4}
KEGG	KEGG:04974	Protein digestion and absorption	13	6.676×10^{-4}
KEGG	KEGG:05410	Hypertrophic cardiomyopathy	12	8.583×10^{-4}
KEGG	KEGG:04151	PI3K-Akt signaling pathway	25	2.335×10^{-3}
KEGG	KEGG:04261	Adrenergic signaling in cardiomyocytes	13	3.488×10^{-2}
b. Mouse top 10 ranked GO and KEGG terms.				
GO:MF	GO:0004308	Exo-alpha-sialidase activity	4	1.652×10^{-3}
GO:MF	GO:0016997	Alpha-sialidase activity	4	1.652×10^{-3}
GO:MF	GO:0052794	Exo-alpha-(2- > 3)-sialidase activity	4	1.652×10^{-3}
GO:MF	GO:0052795	Exo-alpha-(2- > 6)-sialidase activity	4	1.652×10^{-3}
GO:MF	GO:0052796	Exo-alpha-(2- > 8)-sialidase activity	4	1.652×10^{-3}
GO:MF	GO:0008138	Protein tyrosine/serine/threonine phosphatase activity	10	5.683×10^{-3}
GO:BP	GO:0006689	Ganglioside catabolic process	5	4.131×10^{-3}

Table 1 (continued)

Category	GO ID	Description	Gene Count	P-value
GO:BP	GO:0009313	Oligosaccharide catabolic process	5	2.322×10^{-2}
GO:BP	GO:0050877	Nervous system process	128	3.665×10^{-2}
GO:CC	GO:0005882	Intermediate filament	26	1.448×10^{-4}
GO:CC	GO:0045111	Intermediate filament cytoskeleton	26	3.344×10^{-3}
KEGG	GO:04915	Estrogen signaling pathway	19	1.095×10^{-3}

Table 2 The shared DEGs between the humans and mouse data sets

Gene Symbol	Human		Mouse	
	Undifferentiated iPSC Average Count Value	iPSC-CM Average Count Value	Undifferentiated iPSC Average Count Value	iPSC-CM Average Count Value
TMOD1	-4.21854	4.16611	3.72087	-3.75938
FRZB	-3.10053	3.00953	9.05710	-9.05943
OLR1	-2.91656	2.80220	-2.31776	2.298521
PRPH	-2.83795	2.67263	-2.49088	2.479128
SPP1	-2.79341	2.60690	-2.03960	2.022672
VEGFC	-2.45415	2.42974	-2.16861	2.16018
DUSP1	-2.31521	2.24014	-2.66451	2.51438
SH3RF2	-2.00987	1.99280	2.76080	-2.79137
NANOG	2.35317	-2.36439	-2.48918	2.44925
VASH2	2.47692	-2.51421	2.36900	-2.38212

Analysing the GO and KEGG terms and up- or down-regulation of genes

Tables were produced to show the top 5 up- and down-regulated genes for each GO and KEGG term in the networks, as well as each of the subclusters analysed, as shown in Table 6, Additional file 2 Table S4, and Additional file 3 Table S5 in the Supplementary Information. Table 6 summarises the full data from the Additional file 2 Table S4 and Additional file 3 Table S5 by showing most commonly up- and down-regulated genes from the GO and KEGG terms for the human and mouse data sets. All the software and websites used in this paper can be found in the Additional file 4 Table S6. Figure 6 also shows a summary of the top 10 human and mouse hub genes (Table 3), as well as the shared DEGs (Table 2), and their count values to show if they are up- or down-regulated.

Discussion

Several key pathways and genes related to iPSC-CM development and various physiological processes, such as vascular regeneration and prevention of calcification, were identified. These provide guidance and hypotheses for the development of iPSC therapies, such as for aortic stenosis, where cells could not only regenerate

endothelium and muscle cells, but also coded for proteins which helped to prevent calcification. Angiogenesis is one of the key challenges in tissue engineering [17], and there are many novel methods which are being used to try and increase the vascularisation in angiogenesis, including using stem cells. Studies by Geiger et al. showed that transfection of mesenchymal stem cells with VEGF-plasmid DNA increased vascularisation and resorption of bone substitute scaffolds, as well as leading to a more homogenous vascularisation, demonstrating the potential of angiogenic factors [18].

From the human data, there are several pathways relating to angiogenesis. Firstly, the PI3K-Akt pathway is involved in the development of blood vessels during both normal and tumour development, where it mediates various angiogenic factors, especially VEGF [19]. Looking at Table 6, we can also see that COL3A1 is a commonly up-regulated gene, and it is involved in the production of type III collagen, a major component of larger blood vessels [20]. Moreover, the results show upregulation of platelet-derived growth factor (PDGF), which is involved in angiogenesis, especially PDGF receptor beta [21, 22]. Moreover, VASH2, a potent angiogenic growth factor, is downregulated in the human model for the GO term *actin binding*. This may be attributed to its potential

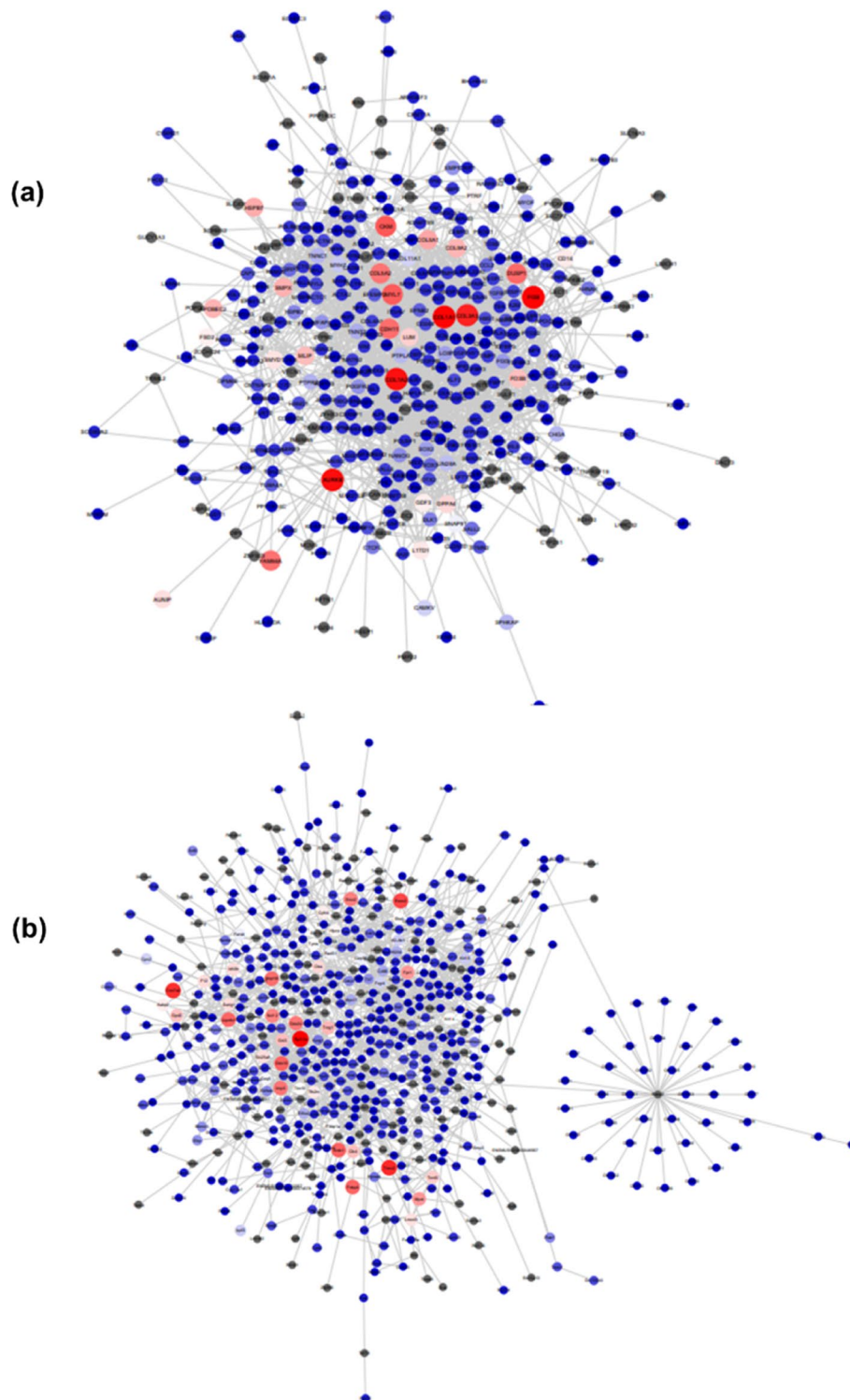


Fig. 4 The PPI networks of all the DEGs. They were edited in Cytoscape with continuous mapping so the colour and size of the node reflected the connection degree. This meant the redder, larger nodes represent a higher connection degree, while the bluer, smaller nodes represent a lower connection degree. For those DEGs, the maximum threshold for coexpression was set as 0.999 and the minimum threshold as 0.0, with values around 0.5019 coloured whiter. The top 10 nodes with the highest connection degrees are known as the hub genes in the PPI networks. **a** the human data set, GSE17579. **b** mouse data set, GSE18514

Table 3 Top 10 most significantly up-regulated genes (Hub Genes) from the PPI Networks of a) human or b) mouse data sets

Rank	Gene Symbol	Coexpression
a. Top 10 human hub genes		
1	FGB	0.992
2	COL1A1	0.991
3	AURKB	0.972
4	COL1A2	0.965
5	COL3A1	0.933
6	CKM	0.825
7	MYL7	0.820
8	DUSP1	0.793
9	CDH11	0.782
10	FAM64A	0.776
b. Top 10 mouse hub genes		
1	RPL10A	0.999
2	TNNC2	0.917
3	COX7A2	0.904
4	ESCO2	0.858
5	SREK1	0.810
6	DDX10	0.804
7	FNBP4	0.801
8	UQCRCF1	0.801
9	MRPL18	0.782
10	IMP3	0.771

inhibition of angiogenesis along with its family member VASH1 reported in [23]. In unsurprising contrast, most of the upregulation of angiogenesis from VASH2 has been shown in tumours rather than normal blood vessel formation [24–27]. VEGFC is also upregulated in Table 2, which encodes for vascular endothelial growth factor C, a protein which promotes angiogenesis [28].

In terms of striated muscle, the human data shows upregulation of significant GO and KEGG terms in relation to the development of the circulatory system, heart, and general anatomy, and the formation of myofibrils, sarcomeres, and contractile fibers. Moreover, key genes are upregulated such as MYOM1, MYH6 and MYH7, as shown in Table 6, with MYH7 being a key marker for ventricular cardiomyocytes (Table S2). We can identify the cells as cardiomyocytes due to the marker TNNT2 which is upregulated in the human iPSCs-CMs, and in contrast, has a low expression in smooth muscle cells [29]. Upregulation of genes related to the formation of smooth muscle or endothelial development are also important because these are key parts of vascular structure. By comparing key genes involved in smooth muscle development with the human data, one of the most upregulated genes, LUM, is also expressed in aortic smooth muscle cells [30]. Here it has a key role in regulating collagen matrix

assembly, as well as aiding in cell migration and proliferation. Other related genes found in this study include ACTA2, which encodes α -smooth muscle actin, involved in smooth muscle contractility and blood pressure homeostasis [31, 32]. This is a known marker of smooth muscle cells (Table S2), indicating the possibility of the presence of smooth muscle cells. MYH11 which codes for myosin in smooth muscle cells [33], MYOM1 which codes for the protein titin, and the aforementioned protein COL3A1 [29], are all shown to be upregulated in the data, and are all also present in smooth muscle cells.

There are fewer genes in the human data related to endothelium formation, apart from the only notable one being VEGFC, which is also involved in endothelial cell growth [28]. In comparison, looking at the top 10 hub genes in the mouse data, TNNC2 (which encodes troponin, a key protein in striated muscle contraction), is the only hub gene to show specific expression to cardiomyocytes. The other hub genes show several different pathways expressed in the GO and KEGG terms, such as RNA processing and splicing, as well as chromosome, ribosome, and cytoskeletal formation, showing less gene expression in terms of angiogenesis, smooth muscle, and endothelial formation. However, several terms can still be linked to cardiomyocytes. Firstly, the estrogen pathway is upregulated, which is known to promote cardiac regeneration and have cardioprotective effects [34]. Moreover, the terms *Cellular calcium ion homeostasis* and *Cellular chemical homeostasis* show that there are processes regulating calcium and transmitting calcium ion currents throughout the mouse iPSC-CMs. These results reflect that the mouse iPSC-CMs have more pathways related to the prevention of calcification compared to human iPSC-CMs. From the human data, the KEGG term *TGF-beta signaling pathway* is upregulated, which may relate to calcification and aortic stenosis. The TGF-beta signaling pathway is involved in elongating glycosaminoglycan (GAG) chains, which causes lipoproteins to accumulate. Lipids will then take part in the calcification process by being modified by enzymes into various lipid-derived compounds [2].

Comparing the shared DEGs in Table 2, most of them are expressed with similar trends, except for TMOD1, FRZB and NANOG. TMOD1 is gene coding for the protein tropomodulin 1, involved in capping the end of actin filaments in sarcomeres in the mouse. The CVDs associated with TMOD1 include hypertrophic and dilated cardiomyopathy [35, 36]. Opposite trends of TMOD1 in both iPSCs and iPSC-CMs are detected, which is downregulated in the mouse iPSC-CM but upregulated in human's, and vice versa for undifferentiated iPSCs. Moreover, expression of the KEGG terms *Hypertrophic cardiomyopathy* and *Dilated cardiomyopathy* in subcluster

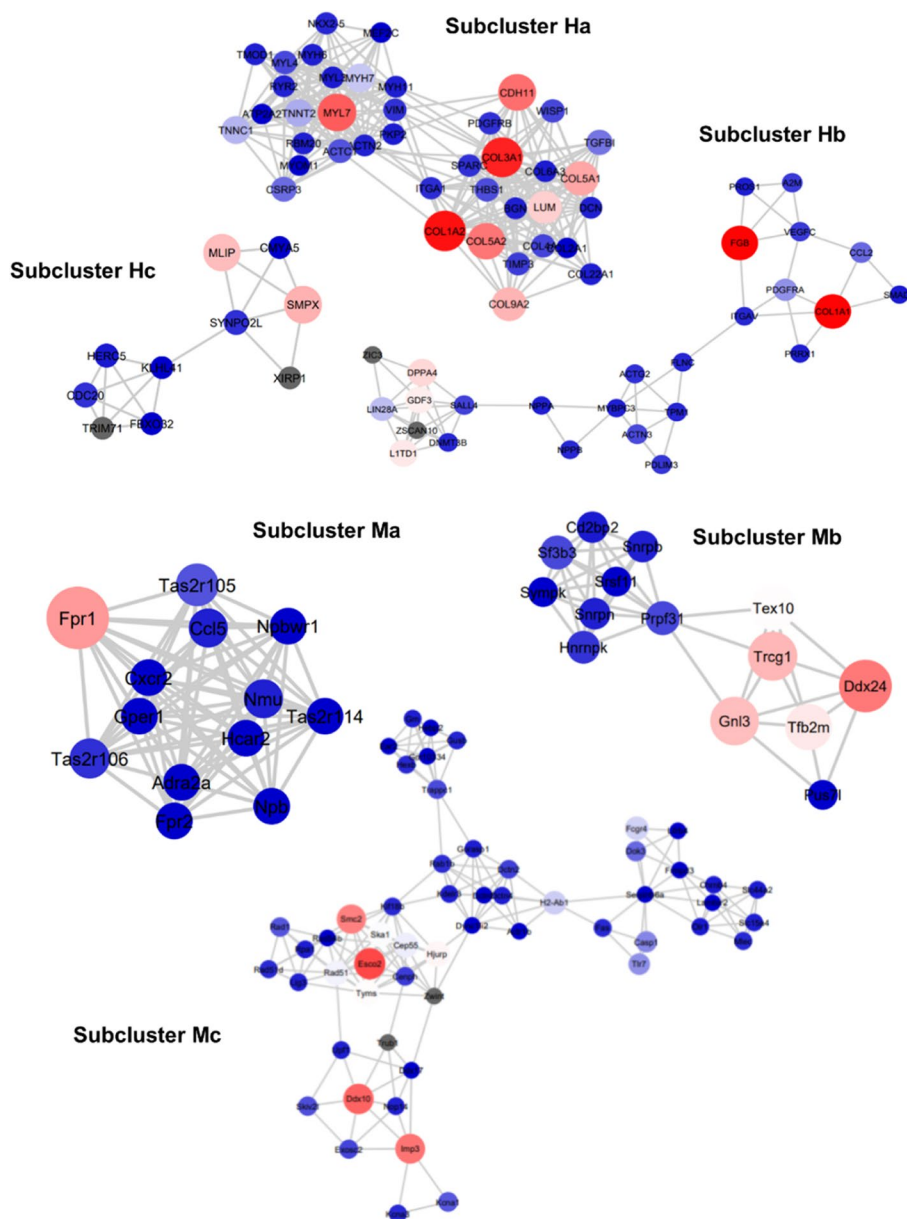


Fig. 5 Subclusters produced from the PPI network using MCODE. These show the top three most significant modules in which hub genes from the original PPI networks are present. The colour and size of the node reflects the connection degree, with the redder, larger nodes having a higher connection degree, while the bluer, smaller nodes having a lower connection degree. **a** Subclusters Ha (Human a), Hb (Human b), and Hc (Human c) from the human data set, GSE17579 **(b)** Subclusters Ma (Mouse a), Mb (Mouse b), and Mc (Mouse c) from the mouse data set, GSE18514. These show which subclusters and genes are related to which processes

Ha shows that upregulation of *TMOD1* might be associated with a higher rate of CVDs in the human iPSC-CMs. This may give a warning that potential conflicting results could have been obtained using *TMOD1* as a marker in mice and humans.

CVDs such as aortic stenosis are also associated with the “osteoblastic transition of cardiac valve interstitial cells” [3], so it is also important to look at genes related

to bone formation. *FRZB* is a gene whose overexpression is linked to osteogenesis. Looking at Table 2, it is upregulated in the human data set but downregulated in the mouse data, showing that perhaps the human cells are more prone to osteogenesis in the aortic stenosis process [37]. However, for both data sets, *SPP1* is upregulated, which is related to bone remodelling [38], with the human data set having a higher up-regulation. Overall,

Table 4 Top 5 Ranked genes in each subcluster from a) human or b) mouse data sets

Rank	1	2	3	4	5
a. Top 5 ranked genes in each of the subclusters from the human dataset					
Subcluster Ha (Human a)					
Gene	COL1A2	COL3A1	MYL7	CDH11	COL5A2
Symbol					
Coexpression	0.965	0.933	0.820	0.782	0.768
Subcluster Hb (Human b)					
Gene	FGB	COL1A1	DPPA4	LITD1	GDF3
Symbol					
Coexpression	0.992	0.991	0.578	0.546	0.535
Subcluster Hc (Human c)					
Gene	SMPX	MLIP	CDC20	SYNPO2L	HERC5
Symbol					
Coexpression	0.972	0.968	0.793	0.776	0.631
b. Top 5 ranked genes in each of the subclusters from the mouse dataset					
Subcluster Ma (Mouse a)					
Gene	FPR1	TAS2R105	TAS2R106	NMU	CD5
Symbol					
Coexpression	0.701	0.164	0.094	0.062	0.062
Subcluster Mb (Mouse b)					
Gene	DDX24	TRCG1	GNL3	TFB2M	TEX10
Symbol					
Coexpression	0.763	0.642	0.628	0.547	0.505
Subcluster Mc (Mouse c)					
Gene	ESCO2	DDX10	IMP3	SMC2	HJURP
Symbol					
Coexpression	0.858	0.804	0.771	0.748	0.519

both genes show the human iPSC-CM's networks are associated with increased bone remodeling compared to the mouse iPSC-CM's.

While the human data downregulates NANOG for stem cell proliferation and renewal, the mouse data has upregulation, which could be a reflection on their relative times in culture, which will be discussed in the limitations section.

In comparison, the genes upregulated in angiogenesis and endothelial formation show consistent trends in both mouse and human. Both iPSC-CMs stimulate these processes, as shown by upregulation in both data sets for VEGFC and downregulation in both for VASH2, showing similarities between the data sets in Table 3, in agreement with previous studies in literature [23–28].

For both data sets there is upregulation of OLR1. OLR1 is a risk factor for coronary artery disease (CAD) because

it encodes lectin-like oxidised low-density lipoprotein receptor-1 (LOX-1), which increases the absorption and degradation of oxidised low-density lipoproteins (ox-LDLs) in cells [39, 40]. Therefore, both the human and mouse data sets can be linked to also being involved in calcification and aortic stenosis since many studies show a link between higher levels of ox-LDL and fibro-calcific modelling in aortic valves [41, 42].

There were a couple of limitations with this study, the first being that each data set downloaded had limited biological replicates, with only 3 each for both the undifferentiated iPSCs and differentiated iPSC-CMs, so there is limited in vitro data. Moreover, there was no in vivo data or data from disease states, which means that we would need further data to verify conclusions on these topics.

Moreover, since the experiments were conducted by two separate groups, there were differences in the methodology between the groups. For example, the human iPSC-CMs were produced via co-culture with the murine visceral endoderm-like cell line END2 [9, 10], while the mouse iPSC-CMs were produced via introducing a transgene into the murine iPSCs. Furthermore, slight differences such as the fact that the genetic expression in mouse iPSC-CMs was tested on day 16, while the human iPSC-CMs were tested on day 18, could be the reason why the shared DEG NANOG was more highly expressed in the mouse iPSC-CMs, since it could reflect that they had a more immature phenotype compared to the human iPSC-CMs.

Another issue in our study includes the varying efficacy of differentiation of the iPSCs. Other methods could be investigated so that a higher proportion of the iPSCs can be successfully differentiated into the desired cells. Several biophysical factors and the concentration, timing and the mixture of factors/molecules will affect the efficiency of the differentiation pathways. Our data provides a starting point in helping to identify the genetic makeup of the iPSC-CMs, and by looking at how this can affect their behavior and how they differentiate into the desired cells, this could be utilised in this further research. This variation in differentiation is shown by the fact that the murine iPSC-CMs display fewer myosin related genes, and genes related to things such as sarcomere assembly or muscle function. Typically, iPSC-CMs are cultured for 35–40 days post differentiation to allow for maturation into CMs before any studies are done, so the biological and physiological differences shown in the data could have been due to the mouse and human iPSC-CMs being at different points of maturation than normal. Other possible causes of this could have been due to the differences in the species themselves, any possible epigenetic pattern retention, or the different methods of their differentiation which could have led to the cells going down different developmental routes.

Table 5 Pathway enrichment analysis to show the significantly enriched GO and KEGG terms for each subcluster from a) human or b) mouse data sets

Category	GO ID	Description	Gene Count	P-value
a. Pathway enrichment analysis to show the top 3 GO and KEGG terms of Subcluster Ha				
GO:MF	GO:0005201	Extracellular matrix structural constituent	15	1.032×10^{-18}
GO:MF	GO:0005198	Structural molecule activity	21	1.084×10^{-17}
GO:MF	GO:0030020	Extracellular matrix structural constituent conferring tensile strength	8	6.423×10^{-12}
GO:BP	GO:0006936	Muscle contraction	19	3.864×10^{-19}
GO:BP	GO:0003012	Muscle system process	20	1.377×10^{-18}
GO:BP	GO:0070252	Actin-mediated cell contraction	13	5.983×10^{-18}
GO:CC	GO:0099081	Supramolecular polymer	25	2.044×10^{-20}
GO:CC	GO:0099512	Supramolecular fiber	24	5.143×10^{-19}
GO:CC	GO:0099080	Supramolecular complex	25	1.732×10^{-17}
KEGG	KEGG:04260	Cardiac muscle contraction	9	1.136×10^{-9}
KEGG	KEGG:05410	Hypertrophic cardiomyopathy	9	1.554×10^{-9}
KEGG	KEGG:05414	Dilated cardiomyopathy	9	2.554×10^{-9}
b. Pathway enrichment analysis to show the top 3 GO and KEGG terms of Subcluster Hb				
GO:MF	GO:0008307	Structural constituent of muscle	4	1.214×10^{-4}
GO:MF	GO:0005102	Signaling receptor binding	11	2.258×10^{-3}
GO:MF	GO:0019838	Growth factor binding	4	1.264×10^{-3}
GO:BP	GO:0072359	Circulatory system development	14	2.763×10^{-7}
GO:BP	GO:0032502	Developmental process	24	1.135×10^{-5}
GO:BP	GO:0035019	Somatic stem cell population maintenance	5	1.171×10^{-4}
GO:CC	GO:0031093	Platelet alpha granule lumen	4	4.123×10^{-4}
GO:CC	GO:0031091	Platelet alpha granule	4	1.430×10^{-3}
GO:CC	GO:0030017	Sarcomere	5	1.844×10^{-3}
KEGG	KEGG:04933	AGE-RAGE signaling pathway in diabetic complications	4	4.815×10^{-3}
KEGG	KEGG:04510	Focal adhesion	5	5.074×10^{-3}
c. Pathway enrichment analysis to show the top 3 GO and KEGG terms of Subcluster Hc				
GO:BP	GO:0016567	Protein ubiquitination	5	4.006×10^{-2}
GO:CC	GO:0030017	Sarcomere	5	3.883×10^{-6}
GO:CC	GO:0030016	Myofibril	5	6.403×10^{-6}
GO:CC	GO:0043292	Contractile fiber	5	8.076×10^{-6}
d. Pathway enrichment analysis to show the top 3 GO and KEGG terms of Subcluster Ma				
GO:MF	GO:0004930	G protein-coupled receptor activity Transmembrane signalling receptor activity	10	2.363×10^{-6}
GO:MF	GO:0004888	Transmembrane signalling receptor activity	10	1.200×10^{-5}
GO:MF	GO:0038023	Signaling receptor activity	10	2.208×10^{-5}
GO:BP	GO:0007186	G protein-coupled receptor signalling pathway	13	5.287×10^{-10}
GO:BP	GO:0006874	Cellular calcium ion homeostasis	6	2.908×10^{-4}
GO:BP	GO:0055082	Cellular chemical homeostasis	7	3.336×10^{-4}
GO:CC	GO:0043679	Axon terminus	3	1.745×10^{-2}
GO:CC	GO:0044306	Neuron projection terminus	3	2.350×10^{-2}
KEGG	KEGG:04080	Neuroactive ligand-receptor interaction	6	3.988×10^{-4}
KEGG	KEGG:04742	Taste transduction	3	1.856×10^{-2}
e. Pathway enrichment analysis to show the top 3 GO and KEGG terms of Subcluster Mb				
GO:MF	GO:0003723	RNA binding	9	9.759×10^{-8}
GO:MF	GO:0003676	Nucleic acid binding	10	4.231×10^{-5}
GO:MF	GO:0043021	Ribonucleoprotein complex binding	4	1.366×10^{-4}
GO:BP	GO:0006396	RNA processing	9	2.465×10^{-8}
GO:BP	GO:0008380	RNA splicing	7	3.751×10^{-7}
GO:BP	GO:0006397	mRNA processing	7	1.214×10^{-6}
GO:CC	GO:0005687	U4 snRNP	3	2.365×10^{-6}

Table 5 (continued)

Category	GO ID	Description	Gene Count	P-value
GO:CC	GO:0097525	Spliceosomal snRNP complex	4	2.365×10^{-6}
GO:CC	GO:0030532	Small nuclear ribonucleoprotein complex	4	4.758×10^{-6}
KEGG	KEGG:03040	Spliceosome	4	3.249×10^{-5}
f. Pathway enrichment analysis to show the top 3 GO and KEGG terms of Subcluster Mc				
GO:MF	GO:0004386	Helicase activity	6	1.539×10^{-3}
GO:MF	GO:0017111	Nucleoside-triphosphatase activity	11	3.713×10^{-3}
GO:MF	GO:0016462	Pyrophosphatase activity	11	6.465×10^{-3}
GO:BP	GO:0006281	DNA repair	9	5.244×10^{-3}
GO:BP	GO:0022402	Cell cycle process	13	5.321×10^{-3}
GO:BP	GO:0000280	Nuclear division	8	1.472×10^{-2}
GO:CC	GO:0000776	Kinetochores	7	1.582×10^{-5}
GO:CC	GO:0005694	Chromosome	14	5.991×10^{-5}
GO:CC	GO:0098687	Chromosomal region	8	9.539×10^{-5}
KEGG	KEGG:03440	Homologous recombination	4	7.741×10^{-4}
KEGG	KEGG:04962	Vasopressin-regulated water reabsorption	3	3.362×10^{-2}

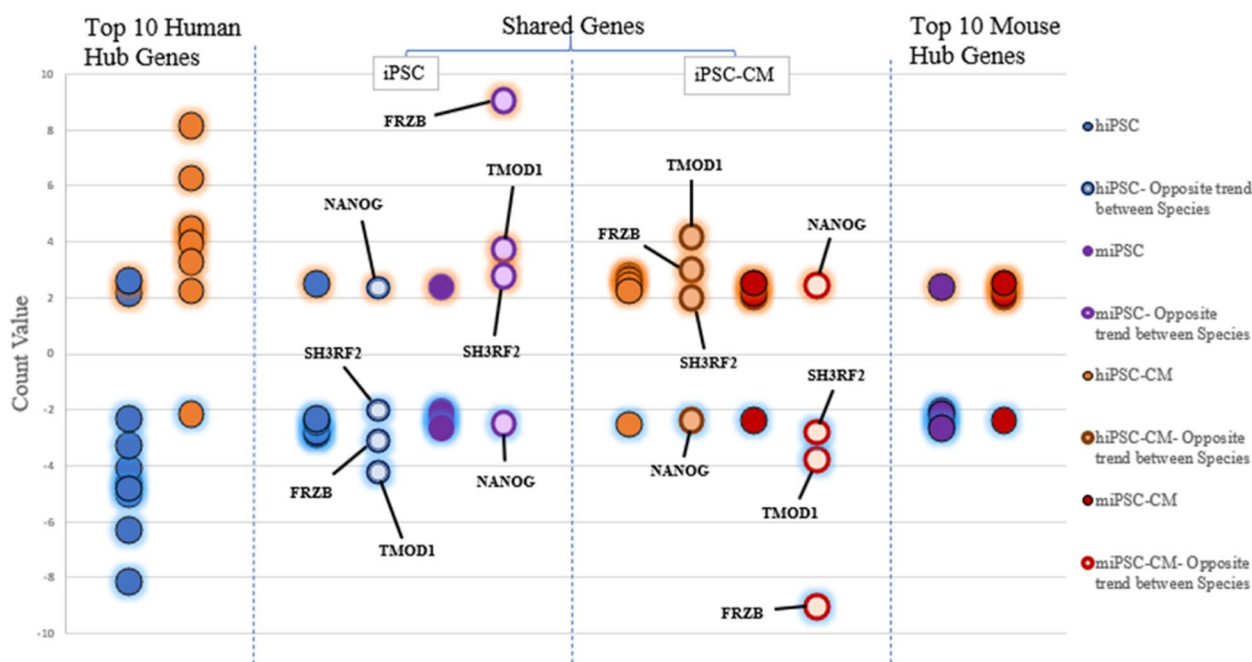


Fig. 6 Bubble plot showing up- and down-regulation of genes- This shows the top 10 human and mouse hub genes, as well as the shared DEGs between the data sets, and their count values for both the undifferentiated iPSCs and the iPSC-CMs. A blue glow behind the dots represents down-regulation of the DEG, while an orange glow represents up-regulation of the DEG. The four shared DEGs which express opposite trends between the human and mouse data sets are labelled (TMOD1, FRZB, SH3RF2, NANOG). Note: For the top 10 mouse hub genes, SREK1 is not plotted because it did not have an equivalent Official Gene Symbol in the conversion table, so its count values were unable to be retrieved

In our future work something that would be pertinent firstly would be to confirm the results of the bioinformatics analysis using qPCR or western blotting to look at the gene expression profiles of the iPSC and iPSC-CMs. Moreover, following on from this we can then start to investigate specific systems discussed above,

such as confirming the mechanisms and genes involved in angiogenesis, calcification, sarcomere assembly and muscle function etc. and be able to see the complex dynamic systems we have only partly touched on so far. For example, the mouse iPSC-CMs have more pathways in the data related to calcification prevention than human

Table 6 Top 5 most commonly up- and down-regulated genes for data sets

Rank	Human Data Set		Mouse Data Set	
	Most commonly up-regulated genes	Most commonly down-regulated genes	Most commonly up-regulated genes	Most commonly down-regulated genes
1	MYH7	ACTN3	ZMAT2	DCTN2
2	MYH6	COL9A2	CD2BP2	PRPF31
3	COL3A1	MAP7	PTGIR	DDX1
4	LUM	ALPL	CALM3	ADRA2A
5	NPPA	PAK1	DYNC1LI2	ZW10

iPSC-CMs, but due to the complex regulatory systems involved in this, things such as the ion channels involved would need to be clarified using experiments such as qPCR. In addition, comparing the iPSC-CMs of mouse and humans to primary cardiomyocytes from the respective organisms is also valuable to see how genetically similar they are.

Further research about the cell delivery to diseased tissues *in vivo* is also important. As shown in Table S3, there are studies investigating the way these cells can actually be delivered/applied to diseased tissues, for instance using heart patches [43]. In future research, we can use iPSCs to produce an iPSC-loaded heart-valve scaffold, and then investigate how this can help to treat CVDs. We can investigate the effect of the scaffold on the delivery and survival of the cells, and the efficacy of treatment *in vivo*. This allows us to further investigate how iPSC-CMs can be more efficiently utilised in cardiovascular diseases, using our current study as a foundation for how iPSC-CMs may differ in their genetic makeup, and thus their behavior in research.

Conclusion

This study analysed the gene expression profiles between human and mouse iPSC-derived cardiomyocytes. DEGs were identified, pathway enrichment analysis was performed to find GO and KEGG terms, and PPI networks and subclusters were constructed and analysed. The results show that iPSC-derived cardiomyocytes could be used as a potential therapeutic for CVD, with the expression of key cardiomyocyte genes. They also show potential in vascular, endothelial, and smooth muscle repair, especially from the human iPSC-CMs, and prevention of calcification in the mouse iPSC-CMs, due to the genes expressed. It also indicates how *in vivo* studies using mouse iPSC-CMs may not reflect those in humans due to the clear differences in gene expression between the species, which affects how

iPSCs will act in research. This is the first study which has shown comprehensive information on the genetic differences between human and mouse iPSC-CMs and highlighting key areas where they may be used for design and validation of iPSC therapeutics.

Abbreviations

iPSCs	Induced pluripotent stem cells
iPSC-CM	Induced pluripotent stem cell-derived cardiomyocytes
CVD	Cardiovascular Disease
TEHV	Tissue engineered heart valve
ESCs	Embryonic stem cells
CM	Cardiomyocyte
MSC-PA	Mesenchymal stem-cell loaded patches
ELISA	Enzyme-linked immunosorbent assay
LCMS	Liquid chromatography-mass spectroscopy
GEO	Gene Expression Omnibus
RNA	Ribonucleic acid
DEGs	Differentially Expressed Genes
GO	Gene Ontology
MF	Molecular function
BP	Biological processes
CC	Cellular component
KEGG	Kyoto Encyclopedia of Genes and Genomes
PPI	Protein-protein interaction
STRING	Search Tool for the Retrieval of Interacting Genes
ECM	Extracellular matrix
Ha	Human a
Hb	Human b
Hc	Human c
Ma	Mouse a
Mb	Mouse b
Mc	Mouse c
PDGF	Platelet-derived growth factor
GAG	Glycosaminoglycans
CAD	Coronary artery disease
LOX-1	Lectin-like oxidised low-density lipoprotein receptor-1
ox-LDLs	Oxidised low-density lipoproteins

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41231-023-00139-7>.

Additional file 1: Fig. S1. Important cell types in the heart. **Table S1.** Current treatment options and unmet clinical needs for valvular heart disease (S1-S7). **Table S2.** Main cell types present in the heart and their markers (S8-S11). **Table S3.** Summary of studies showing the application of iPSCs for CVDs. (S12-S19).

Additional file 2: Table S4. Top 5 Up- and down- regulated Genes for term from the human data set.

Additional file 3: Table S5. Top 5 Up- and down- regulated Genes for each term from the mouse data set.

Additional file 4: Table S6. Software and websites which were used in this paper.

Acknowledgments

The authors acknowledge financial support by the Engineering and Physical Science Research Council, the United Kingdom (EPSRC grant no. EP/L020904/1, EP/M026884/1, and EP/R02961X/1, WS).

Authors' contributions

RB made substantial contributions to the data acquisition, the analysis and writing the manuscript. JC and WS contributed to the original conception and mentored the work. All authors read, discussed and commented on the manuscript. The author(s) read and approved the final manuscript.

Funding

Professor Wenhui Song is supported by the Engineering and Physical Science Research Council, the United Kingdom (EPSRC grant no. EP/L020904/1, EP/M026884/1, and EP/R02961X/1).

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the GEO repository, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18879>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹UCL Centre for Biomaterials in Surgical Reconstruction and Regeneration, Department of Surgical Biotechnology, Division of Surgery & Interventional Science, University College London, London NW3 2PF, UK. ²Centre of Maxillofacial Surgery and Digital Plastic Surgery, Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100144, People's Republic of China. ³Centre for Nanomedicine and Surgical Theranostics, Department of Surgical Biotechnology, Division of Surgery & Interventional Science, University College London, London NW3 2PF, UK.

Received: 3 January 2023 Accepted: 20 February 2023

Published online: 17 March 2023

References

- Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart disease and stroke Statistics-2017 update: a report from the American Heart Association. *Circulation*. 2017;135(10):e146–603.
- Lindman BR, Clavel M-A, Mathieu P, Iung B, Lancellotti P, Otto CM, et al. Calcific aortic stenosis. *Nat Rev Dis Primers*. 2016;2(1):16006.
- Dweck MR, Boon NA, Newby DE. Calcific aortic stenosis: a disease of the valve and the myocardium. *J Am Coll Cardiol*. 2012;60(19):1854–63.
- Liew LC, Ho BX, Soh BS. Mending a broken heart: current strategies and limitations of cell-based therapy. *Stem Cell Res Ther*. 2020;11(1):138.
- Wang KL, Xue Q, Xu XH, Hu F, Shao H. Recent progress in induced pluripotent stem cell-derived 3D cultures for cardiac regeneration. *Cell Tissue Res*. 2021;384:231–40.
- Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, et al. Epigenetic memory in induced pluripotent stem cells. *Nature*. 2010;467(7313):285–90.
- Bizy A, Klos M. Optimizing the use of iPSC-CMs for cardiac regeneration in animal models. *Animals (Basel)*. 2020;10(9):1561.
- Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207–10.
- Saric T GM, Illich DJ, Gaarz A, Schultze JL, Hescheler J. Comparative global transcriptomic profiling of human ES and iPSCs cells and their derived microdissected cardiac clusters gene expression omnibus- GEO: NCBI; 2010 Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17579>.
- Gupta MK, Illich DJ, Gaarz A, Matzkies M, Nguemo F, Pfannkuche K, et al. Global transcriptional profiles of beating clusters derived from human induced pluripotent stem cells and embryonic stem cells are highly similar. *BMC Dev Biol*. 2010;10:98.
- Saric T FA, Gaarz A, Schultze JL, Hescheler J. Transcriptomic profiling of murine ES and iPSC cells, embryoid bodies, and ES and IPS cell-derived cardiomyocytes gene expression omnibus (GEO): NCBI; 2012[updated Jan 16, 2019. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18514>.
- Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*. 2019;47(W1):W234–w41.
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13(8):1194–202.
- Reimand J, Kull M, Peterson H, Hansen J, Vilo J. G:profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. 2007;35(suppl_2):W193–200.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–d13.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
- Novosel EC, Kleinhans C, Kluger PJ. Vascularization is the key challenge in tissue engineering. *Adv Drug Deliv Rev*. 2011;63(4-5):300–11.
- Geiger F, Lorenz H, Xu W, Szalay K, Kasten P, Claes L, et al. VEGF producing bone marrow stromal cells (BMSC) enhance vascularization and resorption of a natural coral bone substitute. *Bone*. 2007;41(4):516–22.
- Karar J, Maity A. PI3K/AKT/mTOR pathway in angiogenesis. *Front Mol Neurosci*. 2011;4:51.
- Kuivaniemi H, Tromp G. Type III collagen (COL3A1): gene and protein structure, tissue distribution, and associated diseases. *Gene*. 2019;707:151–71.
- Risau W, Drexler H, Mironov V, Smits A, Siegbahn A, Funa K, et al. Platelet-derived growth factor is Angiogenic in vivo. *Growth Factors*. 1992;7(4):261–6.
- Battegay EJ, Rupp J, Iruela-Arispe L, Sage EH, Pech M. PDGF-BB modulates endothelial proliferation and angiogenesis in vitro via PDGF beta-receptors. *J Cell Biol*. 1994;125(4):917–28.
- Shibuya T, Watanabe K, Yamashita H, Shimizu K, Miyashita H, Abe M, et al. Isolation and characterization of Vasohibin-2 as a homologue of VEGF-inducible endothelium-derived angiogenesis inhibitor Vasohibin. *Arterioscler Thromb Vasc Biol*. 2006;26(5):1051–7.
- Koyanagi T, Suzuki Y, Saga Y, Machida S, Takei Y, Fujiwara H, et al. In vivo delivery of siRNA targeting vasohibin-2 decreases tumor angiogenesis and suppresses tumor growth in ovarian cancer. *Cancer Sci*. 2013;104(12):1705–10.
- Xue X, Gao W, Sun B, Xu Y, Han B, Wang F, et al. Vasohibin 2 is transcriptionally activated and promotes angiogenesis in hepatocellular carcinoma. *Oncogene*. 2013;32(13):1724–34.
- Takahashi Y, Koyanagi T, Suzuki Y, Saga Y, Kanomata N, Moriya T, et al. Vasohibin-2 expressed in human serous ovarian adenocarcinoma accelerates tumor growth by promoting angiogenesis. *Mol Cancer Res*. 2012;10(9):1135–46.
- Iida-Norita R, Kawamura M, Suzuki Y, Hamada S, Masamune A, Furukawa T, et al. Vasohibin-2 plays an essential role in metastasis of pancreatic ductal adenocarcinoma. *Cancer Sci*. 2019;110(7):2296–308.

28. Joukov V, Pajusola K, Kaipainen A, Chilov D, Lahtinen I, Kukk E, et al. A novel vascular endothelial growth factor, VEGF-C, is a ligand for the Flt4 (VEGFR-3) and KDR (VEGFR-2) receptor tyrosine kinases. *EMBO J*. 1996;15(2):290–8.
29. Kwong G, Marquez HA, Yang C, Wong JY, Kotton DN. Generation of a purified iPSC-derived smooth muscle-like population for cell sheet engineering. *Stem Cell Reports*. 2019;13(3):499–514.
30. Qin H, Ishiwata T, Asano G. Effects of the extracellular matrix on lumican expression in rat aortic smooth muscle cells in vitro. *J Pathol*. 2001;195(5):604–8.
31. Yuan SM. α -Smooth muscle actin and ACTA2 gene expressions in Vasculopathies. *Braz J Cardiovasc Surg*. 2015;30(6):644–9.
32. Skalli O, Pelte MF, Pecllet MC, Gabbiani G, Gugliotta P, Bussolati G, et al. Alpha-smooth muscle actin, a differentiation marker of smooth muscle cells, is present in microfilamentous bundles of pericytes. *J Histochem Cytochem*. 1989;37(3):315–21.
33. Pipes GCT, Sinha S, Qi X, Zhu C-H, Gallardo TD, Shelton J, et al. Stem cells and their derivatives can bypass the requirement of myocardin for smooth muscle gene expression. *Dev Biol*. 2005;288(2):502–13.
34. Luo T, Kim JK. The role of estrogen and estrogen receptors on cardiomyocytes: an overview. *Can J Cardiol*. 2016;32(8):1017–25.
35. Ly T, Pappas CT, Johnson D, Schlecht W, Colpan M, Galkin VE, et al. Effects of cardiomyopathy-linked mutations K15N and R21H in tropomyosin on thin-filament regulation and pointed-end dynamics. *Mol Biol Cell*. 2019;30(2):268–81.
36. Sussman MA, Welch S, Cambon N, Klevitsky R, Hewett TE, Price R, et al. Myofibril degeneration caused by tropomodulin overexpression leads to dilated cardiomyopathy in juvenile mice. *J Clin Invest*. 1998;101(1):51–61.
37. Thyssen S, Cailotto F, Lories R. Osteogenesis induced by frizzled-related protein (FRZB) is linked to the netrin-like domain. *Lab Invest*. 2016;96(5):570–80.
38. Denhardt DT, Noda M. Osteopontin expression and function: role in bone remodeling. *J Cell Biochem*. 1998;72(S30-31):92–102.
39. Salehipour P, Rezagholizadeh F, Mahdiannasser M, Kazerani R, Modarressi MH. Association of OLR1 gene polymorphisms with the risk of coronary artery disease: a systematic review and meta-analysis. *Heart Lung*. 2021;50(2):334–43.
40. Jin P, Cong S. LOX-1 and atherosclerotic-related diseases. *Clin Chim Acta*. 2019;491:24–9.
41. Mohty D, Pibarot P, Després JP, Côté C, Arsenault B, Cartier A, et al. Association between plasma LDL particle size, valvular accumulation of oxidized LDL, and inflammation in patients with aortic stenosis. *Arterioscler Thromb Vasc Biol*. 2008;28(1):187–93.
42. Côté C, Pibarot P, Després JP, Mohty D, Cartier A, Arsenault BJ, et al. Association between circulating oxidised low-density lipoprotein and fibrocalcific remodelling of the aortic valve in aortic stenosis. *Heart*. 2008;94(9):1175–80.
43. Weinberger F, Breckwoldt K, Pecha S, Kelly A, Geertz B, Starbatty J, et al. Cardiac repair in guinea pigs with human engineered heart tissue from induced pluripotent stem cells. *Sci Transl Med*. 2016;8(363):363ra148.

Publisher's Note

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

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

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

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

