Long noncoding RNA *HSCHARME* promotes the maturation of stem cells-derived cardiomyocytes by controlling gene regulatory networks altered in human cardiomyopathies.

Giulia Buonaiuto[1], Fabio Desideri[2], Giulio Storari[1], Alessandro Palma[1], Tiziana Santini[1], Daniela Trisciuoglio[3], Monica Ballarino[1*]

[1]Department of Biology and Biotechnologies "*Charles Darwin*", Sapienza University of Rome, Rome, Italy.
[2]Center for Life Nano- & Neuro-Science of Istituto Italiano di Tecnologia (IIT), 00161, Rome, Italy.
[3]Institute of Molecular Biology and Pathology IBPM-CNR, 00185, Rome, Italy.

[*]*corresponding author:* Monica Ballarino
e-mail address: monica.ballarino@uniroma1.it*; phone*: +39 06 49912201

Running Head
HSCHARME functions in the human heart

Keywords
NcRNA; LncRNA; hiPSCs; Cardiomyocytes; Cardiomyopathies

ABSTRACT

A growing body of evidence suggests that tissue-specific long noncoding RNAs (lncRNA) play pivotal roles in the heart. Here, we exploited the synteny between the mouse and human genomes to identify the novel lncRNA *HSCHARME* (*Human Syntenic CHARME*) and combined single-cell transcriptomics, CAGE-seq data, RNA-FISH imaging and CRISPR-Cas9 genome editing to document its role in cardiomyogenesis. We found that the functional inactivation of *HSCHARME* in hiPSC-derived cardiomyocytes (CM) impairs the expression of calcium handling and contraction genes, with a consequent decline of CM maturation and contractile functions. Consistent with a possible association with disease, large-scale analysis of the lncRNA expression across cardiomyopathy patients revealed increased levels of *HSCHARME* in hypertrophic (HCM) and dilated (DCM) hearts and identified a subset of diseased-associated targets whose expression can be modulated through *HSCHARME* dosage. Overall, our data unlock the potential of *HSCHARME* as a novel non-coding regulator of CM physiology and open new possibilities for the therapeutic application of the lncRNA in disease.

INTRODUCTION

Heart diseases represent the primary cause of death globally. As the cardiac muscle has a limited ability to replace cardiomyocytes, regenerative medicine therapies are thought of as relevant to repairing damaged hearts and preventing transplantation. Nonetheless, the restoration of fully matured and functional cardiomyocytes is still challenging due to the partial knowledge of the factors regulating their development and maturation[1,2]. In the last few decades, significant attention has been given to coding genes[3], with the identification of several protein factors (i.e. *MEF2C*, *GATA4* and *TBX5*) which overexpression can drive transdifferentiation of fibroblast to CM-like cells[4,5]. The recent discovery of long-noncoding RNAs (lncRNA) is expanding the current view by uncovering new, albeit mostly unexplored, modalities for acquiring CM identities[6,7].

LncRNAs represent a heterogeneous class of transcripts longer than 500 nucleotides with limited protein-coding capability[8]. Since their discovery, they have been implicated in regulating every step of the gene life cycle[9], as well as many processes, from physiological to pathological[10]. In the heart, the extraordinary temporal and tissue-specific expression of lncRNAs highlight these molecules as ideal modulators of physiological organ development, making their dysregulation relevant for the insurgence of disease[11–15].

In the nucleus, lncRNAs fine-tune gene expression by interacting with other DNA/RNA molecules, RNA binding proteins, or by shaping the three-dimensional organization of the

genome[16]. Among the existing examples, in mice we have previously identified Charme[17], an architectural lncRNA gene whose expression is specifically restricted to skeletal myotubes and cardiomyocytes (CM). Mechanistically, we found that the nuclear-enriched pCharme lncRNA isoform functions in the formation of chromosome territories that are necessary to coordinate the activation of pro-myogenic genes and the differentiation of muscle cells[18,19]. In mouse embryos, the functional ablation of pCharme leads to morphological alterations of the heart due to the persistent expression of fetal-like gene programs which delays CM maturation[20]. This underscores a crucial and previously overlooked role for Charme in heart physiology and a potential link to cardiac diseases. Nevertheless, no molecular analysis has been performed yet to dissect the possible contribution of this lncRNA in humans.

Here, we applied comparative genomics to explore the evolutionary conservation of Charme and defined a region of synteny between mouse chromosome 7 and human chromosome 19. We found that, in humans, this region transcribes *HSCHARME*, a lncRNA whose expression is restricted to fetal and adult cardiomyocytes, as demonstrated by Cap Analysis of Gene Expression (CAGE), single-cell RNA-sequencing (scRNA-seq) and advanced RNA-FISH imaging approaches. Through functional inactivation of *HSCHARME* in hiPSC-derived CM, we demonstrated a role for the lncRNA in the regulation of CM maturation and contraction capabilities. Moreover, we found the expression of *HSCHARME* being significantly increased in dilated (DCM) and hypertrophic (HCM) human cardiomyopathies. Lastly, we applied a proof-of-principle strategy to demonstrate that *HSCHARME* dosing rescues the expression of genes which are commonly dysregulated in DCM and HCM. Overall, these results identify *HSCHARME* as a novel disease-associated gene with potential implications for cardiomyocyte-specific therapeutic approaches to human cardiomyopathies.

RESULTS

Identification and analysis of *HSCHARME* locus in *Homo sapiens*

To explore the phylogenetic conservation of the murine Charme[17,18] and precisely map the chromosomal location of the human gene in the hg38 genome assembly, we performed a cross-species analysis through the UCSC genome browser “Liftover” tool (**Fig. 1A**). Conversion from the murine (mm10; chr7:44,473,538-44,486,138) into the human coordinates revealed the existence of a syntenic and still uncharacterized locus, named as *HSCHARME*, which maps on chromosome 19 (hg38; chr19:50,486,482-50,500,386) overlapping with a recently annotated gene, *MYREM*, which update our precedent assembly built on skeletal muscle transcriptomics[18].

Vertebrate conservation tracks evidenced an overall high-level sequence conservation of the (~26 kb) region, especially in Rhesus, Mouse, X. Tropicalis and Zebrafish, as well as the synteny of the *EMC10* and *JOSD2* nearby genes. Sequence conservation analysis between the mouse and human genes (intronic and exonic regions) highlighted a degree of sequence identity which is unusually high (45.1 %) for a lncRNA (**Supplementary Fig. 1A**). Consistent with the transcriptional activity of the locus, Transcriptional Start Site (TSS) mapping by CAGE (Cap Analysis of Gene Expression) confirmed the presence of a sharp peak on the negative strand of the *HSCHARME* gene, attributable to the existence of a previously unannotated transcript produced in antisense direction (**Fig. 1A**). We then mined the ENCODE collection from whole-heart tissue samples for epigenetic and gene expression signatures across the human gene. Deposition of histone H3 acetyl-lysine 9 (H3K9ac; GSM706849[21]), trimethylation of histone H3 lysine 4 (H3K4me3; GSM772735[21]) and the absence of the repressive trimethylation of histone H3 lysine 27 (H3K27me3; GSM621450[21]) marks correlate with an overall transcriptional activation of the *HSCHARME* locus in fetal hearts (90-120 days), as also confirmed by the RNA-seq output (GSM1059494; GSM1059495) (**Fig. 1A**). CAGE-sequencing data from the FANTOM5 human promoterome catalogue[22] https://fantom.gsc.riken.jp/zenbu), further confirmed the specific expression of *HSCHARME* in both fetal hearts (84-217 days) and skeletal muscle cells (**Fig. 1B**, left panel). In line with our previous finding in mice[20], scRNA-seq data[23], revealed the highest expression of *HSCHARME* in fetal (72–129 days) cardiomyocytes (**Fig. 1B**, right panel). Restriction of lncRNA expression to cardiomyocytes persists in adult hearts[24], with no evident difference between atria and ventricles (**Fig. 1C**, left and right panels).

We then applied ReMap ChIP-seq atlas to identify possible regulators of *HSCHARME* expression and found enriched peaks for several well-known transcription factors in the 2.5 Kb region from the lncRNA TSS (**Fig. 1D**). Among them, only the T-box transcription factor *TBX5* was identified by the single-cell Human Protein Atlas as having a clear, specific expression in cardiomyocytes (**Fig. 1E** and **Supplementary Fig. 1B**). In addition, the *HSCHARME* regulatory regions contain several TBX5 *consensus* sites as identified by Jaspar database (https://jaspar.genereg.net[25]), which is consistent with its direct binding (**Fig. 1D**). Furthermore, computational analyses of available RNA sequencing datasets from WT, TBX5-KO and TBX5-OE hiPSC-derived cardiomyocytes[26], show a significant decrease in *HSCHARME* expression upon TBX5 loss-of-function (**Fig. 1F**). Along with our previous findings in mice[20], these results underscore a conserved role for TBX5 in positively regulating the expression of *HSCHARME* in the human heart. As the overexpression of TBX5 alone does

not seem to influence the levels of the lncRNA, other factors may be involved in the transcriptional regulation of the locus in cardiac cells.

Overall, these findings provide a high-resolution analysis of *HSCHARME* expression in the human heart and detail the cell-type specific restriction of the lncRNA to cardiomyocytes. The similarities of *HSCHARME* with its murine counterpart, which has already been shown to play an important role in cardiac remodeling[18–20], encouraged further investigations into the relevance of the human transcript in disease.

***HSCHARME* expression and characterization in human iPSC-derived cardiomyocytes**

Considering the specific expression of *HSCHARME* in cardiomyocytes, we used induced pluripotent stem cells (hiPSCs) as a model to study cardiomyogenic differentiation. The applied protocol[27] exploits the induction of the WNT/β-catenin pathway (Day 0, CHIR addition), to guide cells towards mesoderm followed by its subsequent inhibition (Day 3, IWR-1 addition) to promote a cardiomyocyte fate (**Fig. 2A**). RT-qPCR analyses performed at different time points of hiPSCs differentiation show a dynamic wave of cardiac gene expression, which parallels the functional maturation of CM[28]. More specifically, hiPSCs exiting the pluripotency stage display decreased levels of *OCT4* and *NANOG* pluripotency markers while simultaneously increasing the expression of the cardiac mesoderm transcription factor *MESP1* (**Fig. 2B**, left panel). At later maturation stages, genes encoding for cardiomyocyte-specific transcription factors (e.g. *NKX2-5*, *TBX5* and *GATA4*) (**Fig. 2B**, middle panel) and sarcomeric (e.g. *TNNT2* and *MYH6*) (**Fig. 2B**, right panel) proteins were also induced. After 8-10 days, the spontaneous beating of cellular foci was also visible in line with the acquisition of CM maturity (**Supplementary Video 1**). Upon methodological assessment, we then proceeded with the analysis of *HSCHARME* expression and found its gradual increase up to 20 days of differentiation (D20) (**Fig. 2B**, right panel). *TBX5* shows a similar trend of *HSCHARME* activation, further strengthening its functional role as a positive regulator of the lncRNA expression. Overall, these findings suggest a potential association between *HSCHARME* functions and the maturation of human CM. To directly test this hypothesis, we used a CRISPR/Cas9-based gene editing approach to generate *HSCHARME* knock-out (KO) hiPSCs. Specifically, sgRNAs were designed to produce a genomic deletion of ~4.8 Kb, including the newly identified TBX5 binding sites (delta promoter=ΔP) (**Fig. 2C**). Upon confirming the genomic deletion and the successful ablation of *HSCHARME* expression in the KO hiPSCs-derived CM (**Fig. 2D**), we sequenced the genomic regions of 4 putative off-targets for each

guide, which confirmed the absence of any mutations (**Supplementary Fig. 2A**; **Supplementary File 1** and **Supplementary Table 1**).

We then proceeded with transcriptome analysis of isogenic WT (*HSCHARME*^{WT1}) and mutant (*HSCHARME*$^{\Delta P}$) cardiomyocytes from D10 and D20 time-points. *De novo* assembly of reads from the WT samples revealed that two full-length RNA isoforms are produced in cardiomyocytes, with the fully spliced *mCHARME* isoform formed by the exons 1-2-3 and the partially spliced *pCHARME* isoform retaining the first intron (**Fig. 2E**). Both the isoforms were found induced throughout CM differentiation, with *pCHARME* being enriched to nuclei and consistently more abundant than *mCHARME* (**Fig. 2F-G**). Instead, we found a cytoplasmic enrichment for *mCHARME*, although the Coding Potential Calculator tool (CPC2, http://cpc2.gao-lab.org/index.php) excluded the presence of open reading frames possibly translated into micro-peptides[29–32]. Next, we applied high-resolution RNA-fluorescence *in situ* hybridization (RNA-FISH) to specifically visualize *pCHARME* in CM. In agreement with the subcellular fractionation, the imaging experiments confirmed the nuclear localisation of *pCHARME* (**Fig. 2H**). Importantly, no fluorescence signal was detected in the KO conditions, as also assessed by RT-qPCR (**Fig. 3A**, middle panel) and TPM quantification (**Fig. 3A**, right panel). Differently from the murine transcript[20], the *pCHARME* signal in human cardiomyocytes still appears discrete although encompassing multiple nuclear foci. This characteristic adds a step forward into the characterization of *pCHARME* and may reflect an *in trans* regulatory role of the human lncRNA on its physiological targets.

Differential expression analysis performed on the RNA-seq datasets revealed that 175 (D10) and 3637 (D20) genes (DEGs) were dysregulated upon *HSCHARME* ablation in CM, as compared to WT (-1<FC>1; FDR<0.05) (**Fig. 3B; Supplementary Fig. 3A** and **Supplementary Table 2**). Of them, 1465 genes were up-regulated (n=68 at D10; n=1397 at D20) whereas 2347 were down-regulated (n=107 at D10; n=2240 at D20) in KO cardiomyocytes. Upon validation by RT-qPCR of the topmost ranked DEGs in WT (WT1) and *HSCHARME*$^{\Delta P}$ CM (**Fig. 3C** and **Supplementary Fig. 3B**), we noticed the absence of *HSCHARME* targets within the neighbouring chromatin loci (around ±260 Kb *HSCHARME* locus), which is in line with the RNA-FISH analyses and consistently excludes a possible *cis*-acting activity of the lncRNA (**Supplementary Fig. 3C**). Moreover, no sgRNAs off-target gene was found among DEGs (**Supplementary Fig. 3D**) that, together with the absence of genomic mutation in their loci, conclusively excludes their involvement in the transcriptomic alterations. To consider possible genotype-specific effects, which are independent by

HSCHARME ablation, DEGs were tested in parallel in a different KO edited line, henceforth defined as *HSCHARME*^{PA}. Specifically, the *HSCHARME*^{PA} hiPSCs were generated on the WTSIi004-A background through the insertion of a strong termination cassette inside the *HSCHARME* locus, a strategy which was previously shown to be successful in mice[18] (**Supplementary Fig. 3E**). In these cells, DNA sequencing of the off-target regions confirmed the absence of any mutations (**Supplementary Fig. 3F** and **Supplementary File 1**). Moreover, no off-target gene was found in common between the two genotypes, which excludes possible effects due to unwanted Cas9 background activities (**Supplementary Fig. 3G**).

Our previous evidence classified the murine lncRNA Charme as a positive regulator of cardiac gene programs[20]. In accordance, RT-qPCR analyses performed on the selected genes confirmed the observed RNA-seq trends of down-regulation for many genes in both the ΔP and PA knockout genotypes (**Fig. 3C** and **Supplementary Fig. 3B**). Gene Ontology (GO) term enrichment analysis performed across the RNA-seq list shows that in early cardiomyocytes, down-regulated DEGs enrich GO categories tied to developmental pathways, such as limb development (adjusted p-value 0.0095) (**Supplementary Fig. 3H**, left). At later differentiation stages, the GO categories appear more strictly related to cardiomyocyte maturation and functionality, such as those related to the regulation of heart contraction (adjusted p-value $6.30E^{-12}$), cardiac muscle cell development (adjusted p-value $7.81E^{-11}$) and sarcomere organization (adjusted p-value $5.83E^{-10}$) (**Fig. 3D**, left). A more deepened gene set enrichment analysis (GSEA) performed on the cardiac muscle contraction (GO:0060048), cardiac cell development (GO:0055006) and sarcomere organization (GO:0045214) categories confirmed that the majority of the genes belonging to these GO categories were down-regulated upon *HSCHARME* ablation (**Fig. 3E**). Examples refer to *NKX2-5* and *NPPB*, which are involved in pivotal steps of cardiomyocyte development[33] and homeostasis[34]. These ontologies also include functional components, like *MYH6*[35], *TNNT2*[36], Myosin-18b (*MYO18B*), known to regulate cardiac sarcomere organisation[37], and the voltage-dependent calcium channel subunit *CACNA1C*[38].

In contrast to down-regulated DEGs, genes which were found up-regulated upon *HSCHARME* ablation behave differently in the two genotypes and enrich GO categories linked to neuronal processes, such as axonogenesis (adjusted p-value $1.33E^{-6}$) and nervous system development (adjusted p-value $1.39E^{-7}$) (**Fig. 3D**, right). These results may reflect the predisposition of the embryonic stem cells to differentiate toward default neuronal fate in the absence of additional stimuli[39]. In this case, *HSCHARME* depletion would act by turning off activating signals of

cardiomyocyte maturation, thus favouring the upregulation of secondary and genotype-specific pathways towards the default state.

Together with the transcriptomic results, the functional interpretation of the datasets supports the role of *HSCHARME* as a positive regulator of genes whose activity is physiologically relevant for the maturation of human CM.

Functional implication of *HSCHARME* in human cardiomyogenesis

To give functional validation to the gene-expression results, we tested the capacity of *HSCHARME*$^{\Delta P}$ hiPSCs to differentiate into mature CM. We found that the lack of *HSCHARME* results in a decreased percentage of TNNT2 positive (+) differentiating cardiomyocytes, as compared to the isogenic WT control (**Fig. 4A** and **Supplementary Fig. 4A**). Additionally, we observed a significant reduction of the TNNT2 intensity in the fraction of TNNT2+ *HSCHARME*$^{\Delta P}$ cells (**Supplementary Fig. 4B**), which parallels the decreased levels of *TNNT2* mRNA. As TNNT2 represent the cardiac-specific isoform of troponin T, essential in anchoring the troponin complex for the effective contraction of muscle cells[36], these results are in line with the transcriptomic output and add value to the physiological role of *HSCHARME* in promoting the maturation of CM.

Interestingly, we found that other phenotypical traits underlying CM maturation, such as cellular multinucleation and morphology[40,41] were also influenced by *HSCHARME* loss-of-function. Specifically, we observed a significant reduction in the percentage of multinucleated *HSCHARME*$^{\Delta P}$ cardiomyocytes (**Fig. 4B**), which also appear smaller and rounder than the isogenic control cells (**Fig. 4C**). To address whether the observed alteration in cell shape leads to functional impairment, we analyzed the capacity of these cells to contract and found a significant decrease in the beat rate of CM in response to *HSCHARME* knock-out (**Fig. 4D**). Further tracking of the contraction dynamics by video recordings revealed a significant increase in the peak time, with a parallel increase in the peak amplitude in *HSCHARME*$^{\Delta P}$ CM (**Fig. 4E** and **Supplementary Fig. 4C**). These changes are coherent with an overall slowing of the beating rate, as each contraction of *HSCHARME*$^{\Delta P}$ CM tends to be longer and more pronounced than WT cardiomyocytes. As the ability to contract underlies the physiological maturation of cardiomyocytes, these results are consistent with the above and bolster the evidence of *HSCHARME* involvement in CM physiology.

***HSCHARME* dosage impacts the expression of disease-linked genes.**

The influence of the human transcript in cardiomyocyte maturation, along with our early evidence in mice[20] encouraged further investigation into the role of *HSCHARME* in human diseased contexts. In this direction, GSEA analyses performed on *HSCHARME*$^{\Delta P}$ *vs* *HSCHARME*^{WT1} DEGs (D20) revealed the significant enrichment of *HSCHARME* targets among genes dysregulated in two principal human cardiomyopathies, both ventricular hypertrophy (HCM) and dilated (DCM) types (**Fig. 5A**). Further confirmation was obtained by KEGG pathways enrichment analyses performed on the same subsets, showing the significant clustering of *HSCHARME* targets into HCM (adjusted p-value $1.51E^{-10}$) and DCM (adjusted p-value $3.86E^{-9}$) categories (**Supplementary Fig. 5A**). These findings strongly imply a possible implication of the human gene in the development or the homeostasis of these pathological states. More direct evidence was obtained through DE analyses of available RNA-seq datasets of HCM, DCM and Healthy human cohorts (GSE130036[42]; GSE116250[43]). This revealed a significant up-regulation of the *pCHARME* isoform in both pathological contexts (**Fig. 5B**). To more directly identify the genes whose expression in pathology may be influenced by *pCHARME* dysregulation, we then compared the lists of HCM and DCM DEGs (HCM *vs* healthy and DCM *vs* healthy, **Supplementary Table 3**) to *HSCHARME* targets (**Supplementary Table 2**). Out of the ~17.000 commonly expressed genes, we observed more than 1500 significant DEGs for each condition (n=1598 for HCM; n=1603 for DCM) (**Fig. 5C**). In particular, 29.5% of the disease-associated DEGs (category "A", **Fig. 5C**, left panel) were following the *pCHARME* expression trends, being up-regulated in HCM and down-regulated in the *pCHARME* knockout cells. Intriguingly, these DEGs clustered into categories of interest for heart functionality, such as actomyosin structure organization, myofibril assembly, sarcomere organization and regulation of heart contraction. Similarly, the comparison with DCM DEGs revealed that 30.3% of them (category "a", **Fig. 5C**, right panel) were up-regulated in the DCM cohorts and significantly down-regulated upon *HSCHARME* ablation. Again, these genes clustered into categories functionally linked to heart development, such as sarcomere organization, myofibril assembly and cardiac muscle tissue development. On the other hand, the intersection between genes up-regulated upon *HSCHARME* ablation and down-regulated in pathologies was limited to a few cases, which enriches only one GO category (Transmembrane receptor protein tyrosine kinase signaling pathway) with little interest in CM development (categories "B" and "b", **Fig. 5C**). The remaining DEGs display expression trends which are not correlated with *pCHARME* (categories "C" and "c", **Fig. 5C**). As such, their alteration could be ascribed to either *HSCHARME* or context/pathology-specific effects.

Interestingly, by comparing the categories “A” and “a” (**Fig. 5D**), we discovered a subset of 184 genes whose expression in CM was *HSCHARME*-dependent and coherently up-regulated in both cardiomyopathies. Among them, we found structural components for muscular contraction (i.e. myofibril assembly, sarcomere organization and actomyosin structure organization), calcium handling (i.e. *ATP2A3*, *CAMK2B* and *CAPN3*) and regulators of morphogenesis (i.e. *LEFTY2*, *SEMA4A*). We also found *MYO18B*, which was proposed as a causative gene for cardiomyopathy[44]. On the other hand, we could not find any significant enrichment of cardiac-related GO categories upon comparing the “B” and “b” lists (**Supplementary Fig. 5B**). Collectively, these data confirm *HSCHARME* role in the activation of cardiac gene expression programs and identify a specific subset of disease-linked genes for testing the lncRNA dosage.

With the purpose of fine-tuning (either silence or induce) the expression of the functional *pCHARME* isoform within the same cellular genetic background, we then established an inducible hiPSCs CRISPR-based system (*HSCHARME*^{RHE}) acting as a dynamic rheostat. CRISPR-cas9 mediated homologous recombination was achieved by designing a single sgRNA to add the doxycycline (DOXY)-responsive element (TRE promoter) upstream of the *HSCHARME* locus (**Supplementary Fig. 5C; Supplementary Table 1**). RT-qPCR analysis confirmed the success of this strategy as in the untreated conditions (-DOXY) the levels of *pCHARME* were comparable to knockout CM whereas they increased 4-5 fold upon the addition of doxycycline (+DOXY) (**Supplementary Fig. 5D-E)**. Within this context, we then performed RT-qPCR analyses focusing on specific *HSCHARME* targets with a clear role in cardiac sarcomere organisation and voltage-dependent calcium release, such as *MYH7*, *TNNT2*, *CACNA1C* and *CACNB1* (**Supplementary Table 2**). We found that the expression of these genes was correlated to *pCHARME* modulation as it was reduced in the -DOXY condition while triggered upon +DOXY treatment. In line with its role as an upstream regulator of *HSCHARME* expression, the *TBX5* levels were not affected by the lncRNA dosage (**Supplementary Fig. 5F**).

Next, we extended the analysis to *HSCHARME* targets whose expression was also dysregulated in HCM and DCM (**Supplementary Table 2** and **3**). When tested by RT-qPCR, we found that the expression of *MYO18B*, *ATP2A3*, *SEMA4A* and *PPFIA4* was responsive to *pCHARME* dosage as it was decreased in -DOXY CM and rescued upon doxycycline addition. In contrast, *CAMK2B* and *NPPB* did not show any evident response in CM, possibly due to their functional implication in other biological contexts and actions[45,46] (**Fig. 5E**). The responsiveness of cardiac gene expression to *HSCHARME* levels along with its significant influence on the

expression of disease-linked genes represent an important advancement for the relevance of *HSCHARME* in diagnostic and therapeutic applications for diseases.

DISCUSSION

Understanding the molecular causes of cardiac pathologies represents a priority for human health. However, the inability of the heart to regenerate and its mixed cellularity add complexity to medical management, making cardiomyopathies the leading cause of death in many developed countries[47,48]. Current clinical practices and pharmacological treatments aim to restore cardiac functionality, but they are unable to repair the damaged tissue. Consequently, a more effective approach is envisioned through regenerative medicine, which aims to develop feasible strategies to replenish the pool of functional CM in the injured heart. Along this direction, lncRNAs may represent promising candidates for fine-tuning CM-specific gene expression programs due to their precise expression pattern, which is strictly regulated in terms of time and location[49]. New biopharmaceutical companies are starting to show the potential of lncRNA-based platforms in enhancing precision medicine strategies for a wide range of illnesses influenced by lncRNA levels, including cardiovascular diseases. This is the case of the HAYA's therapeutic candidate HTX-001, an antisense oligonucleotide targeting the heart-specific lncRNA Wisper[50].

Here, we provide evidence of the importance of *HSCHARME* in the pathophysiology of the heart and set the stage for innovative approaches of lncRNA-targeting precision therapies for the cell-specific treatment of human cardiomyopathies (**Fig. 6**). Indeed, the conservation across species and the high specificity of *HSCHARME* expression for cardiomyocytes make this lncRNA an excellent candidate for pioneering therapeutical treatments that could minimise off-target effects on other cardiac cell types. Currently, several examples of lncRNA playing a role in human cardiomyocyte homeostasis have been identified[6,7,51,52]. However, these studies rely on *in vitro* systems that induce differentiation exclusively towards cardiomyocytes. Therefore, it is unclear whether the role of these lncRNAs in cardiac processes is truly cardiomyocyte-specific or if they could also have roles in other cardiac cell subtypes, such as cardiac fibroblast or endothelial cells. In our study, the use of single-cell omics allowed us to uncover the specific expression of *HSCHARME* in human cardiomyocytes, providing a solid basis to investigate its role more accurately and specifically within the cardiac context. Specifically in pathology, our new findings uncover increased *HSCHARME* expression in DCM and HCM patients. As a proof-of-concept, we provide compelling evidence that adjusting the levels of *HSCHARME* can restore the expression of several disease-linked genes involved in CM physiological

processes, such as calcium handling, heart contraction and cardiomyocyte maturation. Moreover, given its role as a positive regulator of CM maturation, the modulation of *HSCHARME* expression levels might be beneficial for guiding CM de-differentiation into a progenitor-like state, which is crucial for their replenishment after injury and, ultimately, for the effective regeneration of the cardiac muscle *in vivo*[53–55]. Lastly, the proven association we have established between *HSCHARME* deregulation and common cardiac pathologies indicates the necessity of including this lncRNA in genetic testing, particularly for cardiomyopathies with no clear genetic evidence. Alternatively, assessment of *HSCHARME* expression levels could assist in identifying differences in outcome prediction when stratifying cohorts of patients who have been tested positive for genetic markers, based on phenotype manifestation.

MATERIALS AND METHODS

Cell culture conditions and transfection

Human Episomal iPSCc (WT1 from [56] and WT2 are the WTSIi004-A cell line from EBiSC; https://ebisc.org/WTSIi004-A) were maintained in Essential 8 media (Life Technologies) in plates coated with Geltrex LDEV-free reduced growth factor basement membrane matrix (Life Technologies). Differentiation into cardiomyocytes was performed as described in [27]. Specifically, cells were grown as monolayers in RPMI 1640 Medium (Life Technologies) supplemented with B27 Supplement without Insulin (Life Technologies) for 10 days and then switched to RPMI 1640 Medium (Life Technologies) supplemented with B27 Supplement with Insulin (Life Technologies). The induction towards the cardiac lineage was performed by upregulating canonical Wnt signalling (adding CHIR-99021; Selleck) in the early stages of differentiation and by its subsequent inhibition after 3 days (adding IWR-1; Sigma). Genome editing analyses on hiPSCs were performed by co-transfecting the PX333 plasmid[57], harbouring the cloned sgRNAs (designed with https://chopchop.cbu.uib.no/; sequences available upon request) and the CAS9 sequence, together with the HR 110PA-1 donor vector (System Biosciences) in which two homology arm sequences were cloned. Cloning was performed employing In-fusion® kit (Takara), according to the manufacturer's protocol. Cloning with this approach requires adding 15 nt-long tails to the oligonucleotides to create a homology region among insert and plasmid, allowing recombination into bacteria cells (E.Coli STBL3®). The transfection was performed using the Neon Transfection System (Life Technologies). After 48 hours, cells were selected using 1 μg/ml puromycin to identify heterozygous and homozygous *HSCHARME* KO or *HSCHARME* RHE clones.

RNA extraction and RT-qPCR analysis

Total RNA from cultured cells and tissues was isolated using TRI Reagent (Zymo Research), extracted with Direct-zol™ RNA MiniPrep (Zymo Research), treated with DNase (Zymo Research), retrotranscribed using PrimeScript Reagent Kit (Takara) and amplified by RT-qPCR using PowerUp SYBR-Green MasterMix (Life Technologies), as described in [19] (see **Supplementary Table 4**). Nucleus/Cytoplasm fractionation of hiPSC-derived cardiomyocytes was performed using the PARIS kit (Life Technologies), following the manufacturer's specifications.

RNA-Sequencing Analysis

Total RNA was collected from hiPSCs-derived cardiomyocytes (day 10 and day 20) obtained from three independent biological replicates. Illumina Stranded mRNA Prep was used to prepare cDNA libraries. RNA-Seq was performed on an Illumina Novaseq 6000 Sequencing system at the Genomic Facility of the Center for Human Technologies of IIT.

Q30-filtered raw reads were obtained from Illumina BaseSpace Reads and aligned to GRCh38 assembly using STAR aligner software[58]. Gene loci fragment quantification was performed on Ensemble (release 87) gene annotation gtf using STAR –quantMode GeneCounts parameter. The gtf file was edited adding the *HSCHARME* genomic coordinates (**Supplementary Table 2**). Read counts of "reverse" configuration files were combined into a count matrix file, that was given as input to DESeq2 R package[59] for normalization and differential expression analysis, after removing genes with less than 10 counts in at least two samples. Adjusted p-value cutoff for selecting significant differentially expressed genes was set to 0.05 unless otherwise specified. Heatmap of differentially expressed genes was generated using pheatmap R package[60] from normalized scaled data.

Volcano plots were generated using *Enhanced Volcano* R package (https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html). Gene Ontology enrichment analyses were performed with the EnrichR online tool (https://maayanlab.cloud/Enrichr/[61–63]) using standard settings. Gene Set Enrichment Analysis was performed with the GSEA tool from Broad Institute[64,65] running a pre-ranked analysis with statistically significant genes ordered by log2 fold change (KO *vs* WT D20). Gene Ontology was run using the complete GO annotation file (c5.all.v2023.2.Hs.symbols.gmt). Human Phenotype Ontology (HPO) was run using the HPO annotation file

(c5.hpo.v2023.2.Hs.symbols.gmt). Fig.s were prepared with R and Prism 7.0 software and edited with Adobe Illustrator 2024.

HSCHARME RNA Isoforms Analysis

Reads derived from the WT RNA-seq datasets were used as input for the Salmon tool[66] to identify and quantify the *HSCHARME* isoforms. A transcriptome file was generated using the human genome FASTA file and the GTF file containing *HSCHARME* coordinates. Transcripts per million (TPM) output values were used for quantification and statistical analysis. Plot for quantifying the *HSCHARME* isoforms was generated with the *ggtranscript* R package (https://github.com/dzhang32/ggtranscript), using the genomic coordinates provided within the GTF annotation file derived from Salmon alignment. Fig. was edited with Adobe Illustrator 2024.

Pathological dataset analysis

Pathological datasets raw data (FASTQ files) were retrieved from Gene Expression Omnibus (GEO) database with the following accession IDs: GSE130036 for Hypertrophic Cardiomyopathy and GSE116250 for Dilated Cardiomyopathy. Data were pre-processed and analyzed as described in the RNA-Sequencing analysis section (see **Supplementary Table 3**). Intersections between *HSCHARME* KO and the pathological datasets were performed using a 0.05 adjusted p-value cutoff for all genes in all the analyzed datasets and reporting the log2 fold change values for each gene. Fig.s were prepared with Prism 7.0 software and edited with Adobe Illustrator 2024.

Flow cytometry assay

After 20 days of differentiation (D20), hiPSC-derived cardiomyocytes were washed twice with DMEM F12 and dissociated as single cells using 0.25% trypsin-EDTA. Isolated CM were fixed in 4% paraformaldehyde (PFA) in PBS for 15 minutes at room temperature (RT) before staining. After fixation, cells were washed twice by being resuspended in Flow Cytometry Buffer (FCB; 5% FBS in PBS), centrifuged for 5 min at 250 g and the supernatant was discarded. A third wash with FCB-0.2% Triton X-100 was performed as described above. Cells were stained with the diluted (1:100) primary antibody (TNNT2, see **Supplementary Table 4**) in FCB-0.2% Triton X-100 for 45 min at RT. After two washes in FCB, cells were incubated for 30 min at RT with a secondary antibody (Alexa Fluor 488, Life Technologies) at 1:400 dilution, washed twice more and resuspended in PBS. Stained cells were analyzed and sorted

on a flow cytometer (BD Bioscience, Milan, Italy). Data was collected from at least 10,000 events. Flow cytometry data were analyzed with Flowing software 2.5.1. After exclusion of debris and doublets based on light scatter properties, single cells were analyzed for the expression of TNNT2 and reported as the percentage of marker-positive cells or as mean fluorescence intensity.

Immunofluorescence and RNA-FISH assay

Glass coverslips were prepared by treatment with poly-L-lysine (50 mg/ml in water) for 30 minutes under UV light and overnight treatment with Laminin (2 mg/ml in water) at 37 °C. After 20 days of differentiation (D20), hiPSC-derived cardiomyocytes were washed twice with DMEM F12 and dissociated as single cells using 0.25% trypsin-EDTA. Isolated CM were re-plated on the treated coverslips in RPMI 1640 Medium (Life Technologies) supplemented with B27 Supplement and Rock inhibitor (1:1000). After 5 hours, cells were fixed in 4% paraformaldehyde (PFA) in PBS for 20 minutes at 4°C prior staining. For immunofluorescence, cells were permeabilized with 0.2% Triton X-100 in PBS for 20 minutes at room temperature (RT) and blocked with 1% Goat Serum (GS) in PBS for 20 minutes at RT. The primary antibody (TNNT2, see **Supplementary Table 4**) was diluted (1:200) in 1% GS-PBS and incubated overnight at 4°C. The cells were rinsed with PBS and incubated for 1 hour with a secondary antibody (Alexa Fluor 488, Life Technologies) at 1:300 dilution, rinsed, counterstained with DAPI for 5 minutes at RT and mounted on cover-glasses.

For RNA-FISH experiment, staining was carried out with the HCR RNA-FISH technology according to manufacturer protocol (https://www.molecularinstruments.com/hcr-rnafish), with a specific set-up thought for lncRNA staining. Images were acquired on Carl Zeiss Axio Vert.A1 Microscope or with an inverted confocal Olympus IX73 microscope equipped with a Crestoptics X-LIGHT V3 spinning disk system and a Prime BSI Express Scientific CMOS camera. The acquisitions were obtained using a 20X air objective (IF-only samples) or UPLANSApo 60X (NA 1.35) oil objective (FISH/IF samples) and collected with the MetaMorph software (Molecular Devices).

Multinucleation and morphological analysis

Multinucleation was calculated as the percentage of nuclei in multinucleated CM (defined by at least 2 nuclei) divided by the total number of nuclei in each microscope field (20x acquisition). For each biological replicate, 8 to 10 fields were analysed with ImageJ Software[67]. For morphological analysis, 10-12 cells for each field were individually selected and analysed

using ImageJ standard tools. 10 fields were considered for each biological replicate. The five largest and smallest cells for each replicate were discarded in the analysis leaving 90 cells/replicate. Cell area and circularity were computed with the ImageJ measure tool with circularity=1 indicating a perfect circle. Elongation represents the ratio of the major axis length to the minor axis length, as described in [41]. For each parameter, the average of each analysed field was represented as a dot in the plots while statistical analyses were performed on the overall average measure of the 3 biological replicates.

Cardiomyocyte contraction analysis

Cardiomyocyte spontaneous contraction was assessed after 20 days of differentiation in 12-well multidishes. For each biological replicate, two independent wells from the same differentiation experiment were analysed. To quantify beat rate, the contraction events in 30 s were counted under Carl Zeiss Axio Vert.A1 Microscope equipped with 20X air objectives. The number of contraction events (N) was converted into beats per minute (BPM) using the formula: (N/30)*60. To dissect amplitude and timing of contraction events, light brightfield videos were recorded with Carl Zeiss Axio Vert.A1 Microscope equipped with 20X air objectives and analysed using ImageJ tools, as described in [68].

Statistical methods and rigour

Statistical tests, p-values, and n for each analysis are reported in the corresponding figure legend.

Data availability

The data presented in this study will be openly available in NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) with reference number GSE273100.

REFERENCES

1. Roth, G. A. et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. J Am Coll Cardiol 76, 2982–3021 (2020).
2. Eschenhagen, T. et al. Cardiomyocyte Regeneration: A Consensus Statement. Circulation 136, 680 (2017).
3. Buijtendijk, M. F. J., Barnett, P. & van den Hoff, M. J. B. Development of the human heart. Am J Med Genet C Semin Med Genet 184, 7–22 (2020).
4. Ieda, M. et al. Direct Reprogramming of Fibroblasts into Functional Cardiomyocytes by Defined Factors. Cell 142, 375–386 (2010).
5. Hashimoto, H. et al. Cardiac Reprogramming Factors Synergistically Activate Genome-wide Cardiogenic Stage-Specific Enhancers. Cell Stem Cell 25, 69-86.e5 (2019).
6. Ounzain, S. et al. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. J Mol Cell Cardiol 89, 98–112 (2015).
7. Sato, M. et al. The lncRNA Caren antagonizes heart failure by inactivating DNA damage response and activating mitochondrial biogenesis. Nature Communications 2021 12:1 12, 1–21 (2021).
8. Mattick, J. S. et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nature Reviews Molecular Cell Biology 2023 24:6 24, 430–447 (2023).
9. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22, 96–118 (2021).
10. Oo, J. A., Brandes, R. P. & Leisegang, M. S. Long non-coding RNAs: novel regulators of cellular physiology and function. Pflugers Archiv 474, 191 (2022).
11. Ballarino, M., Morlando, M., Fatica, A. & Bozzoni, I. Non-coding RNAs in muscle differentiation and musculoskeletal disease. J Clin Invest 126, 2021–2030 (2016).
12. McMullen, J. R. & Drew, B. G. Long non-coding RNAs (lncRNAs) in skeletal and cardiac muscle: potential therapeutic and diagnostic targets? Clin Sci (Lond) 130, 2245–2256 (2016).
13. Hobuß, L., Bär, C. & Thum, T. Long non-coding RNAs: At the heart of cardiac dysfunction? Front Physiol 10, 386731 (2019).
14. Mably, J. D. & Wang, D. Z. Long non-coding RNAs in cardiac hypertrophy and heart failure: functions, mechanisms and clinical prospects. Nature Reviews Cardiology 2023 21:5 21, 326–345 (2023).
15. Zeinelabdeen, Y., Abaza, T., Yasser, M. B., Elemam, N. M. & Youness, R. A. MIAT LncRNA: A multifunctional key player in non-oncological pathological conditions. Noncoding RNA Res 9, 447–462 (2024).
16. Ferrer, J. & Dimitrova, N. Transcription regulation by long non-coding RNAs: mechanisms and disease relevance. Nature Reviews Molecular Cell Biology 2024 25:5 25, 396–415 (2024).
17. Ballarino, M. et al. Novel Long Noncoding RNAs (lncRNAs) in Myogenesis: a miR-31 Overlapping lncRNA Transcript Controls Myoblast Differentiation . Mol Cell Biol 35, 728–736 (2015).
18. Ballarino, M. et al. Deficiency in the nuclear long noncoding RNA Charme causes myogenic defects and heart remodeling in mice . EMBO J 37, (2018).
19. Desideri, F. et al. Intronic Determinants Coordinate Charme lncRNA Nuclear Activity through the Interaction with MATR3 and PTBP1. Cell Rep 33, (2020).
20. Taliani, V. et al. The long noncoding RNA Charme supervises cardiomyocyte maturation by controlling cell differentiation programs in the developing heart. Elife 12, (2023).

21. Zhang, J. et al. An integrative ENCODE resource for cancer genomics. Nature Communications 2020 11:1 11, 1–11 (2020).
22. Noguchi, S. et al. FANTOM5 CAGE profiles of human and mouse samples. Sci Data 4, (2017).
23. Cao, J. et al. A human cell atlas of fetal gene expression. Science 370, (2020).
24. Jones, R. C. et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science (1979) 376, (2022).
25. Castro-Mondragon, J. A. et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res 50, D165–D173 (2022).
26. Churko, J. M. et al. Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. Nature Communications 2018 9:1 9, 1–14 (2018).
27. Nakahama, H. & Di Pasquale, E. Generation of cardiomyocytes from pluripotent stem cells. Methods in Molecular Biology 1353, 181–190 (2016).
28. Jeziorowska, D., Korniat, A., Salem, J. E., Fish, K. & Hulot, J. S. Generating patient-specific induced pluripotent stem cells-derived cardiomyocytes for the treatment of cardiac diseases. Expert Opin Biol Ther 15, 1399–1409 (2015).
29. Nelson, B. R. et al. Muscle physiology: A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science (1979) 351, 271–275 (2016).
30. Makarewich, C. A. et al. MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β-Oxidation. Cell Rep 23, 3701–3709 (2018).
31. Stein, C. S. et al. Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. Cell Rep 23, 3710-3720.e8 (2018).
32. Yan, Y. et al. The cardiac translational landscape reveals that micropeptides are new players involved in cardiomyocyte hypertrophy. Molecular Therapy 29, 2253–2267 (2021).
33. Anderson, D. J. et al. NKX2-5 regulates human cardiomyogenesis via a HEY2 dependent transcriptional network. Nature Communications 2018 9:1 9, 1–13 (2018).
34. Matsuo, A., Nagai-Okatani, C., Nishigori, M., Kangawa, K. & Minamino, N. Natriuretic peptides in human heart: Novel insight into their molecular forms, functions, and diagnostic use. Peptides (N.Y.) 111, 3–17 (2019).
35. Anfinson, M. et al. Significance of α-Myosin Heavy Chain (MYH6) Variants in Hypoplastic Left Heart Syndrome and Related Cardiovascular Diseases. J Cardiovasc Dev Dis 9, (2022).
36. Landim-Vieira, M. et al. Cardiac troponin T N-domain variant destabilizes the actin interface resulting in disturbed myofilament function. Proc Natl Acad Sci U S A 120, (2023).
37. Latham, S. L. et al. Myosin-18B Regulates Higher-Order Organization of the Cardiac Sarcomere through Thin Filament Cross-Linking and Thick Filament Dynamics. Cell Rep 32, (2020).
38. Wang, K. et al. Arrhythmia mutations in calmodulin cause conformational changes that affect interactions with the cardiac voltage-gated calcium channel. Proc Natl Acad Sci U S A 115, E10556–E10565 (2018).
39. Muñoz-Sanjuán, I. & Brivanlou, A. H. Neural induction, the default model and embryonic stem cells. Nature Reviews Neuroscience 2002 3:4 3, 271–280 (2002).
40. Funakoshi, S. et al. Generation of mature compact ventricular cardiomyocytes from human pluripotent stem cells. Nature Communications 2021 12:1 12, 1–23 (2021).

41. Sutcliffe, M. D. et al. High content analysis identifies unique morphological features of reprogrammed cardiomyocytes. Sci Rep 8, (2018).
42. Liu, X. et al. Long non-coding and coding RNA profiling using strand-specific RNA-seq in human hypertrophic cardiomyopathy. Scientific Data 2019 6:1 6, 1–7 (2019).
43. Sweet, M. E. et al. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. BMC Genomics 19, (2018).
44. Malfatti, E. et al. A Premature Stop Codon in MYO18B is Associated with Severe Nemaline Myopathy with Cardiomyopathy. J Neuromuscul Dis 2, 219 (2015).
45. Falcetta, D. et al. CaMKIIβ deregulation contributes to neuromuscular junction destabilization in Myotonic Dystrophy type I. Skelet Muscle 14, (2024).
46. Giovou, A. E., Gladka, M. M. & Christoffels, V. M. The Impact of Natriuretic Peptides on Heart Development, Homeostasis, and Disease. Cells 13, (2024).
47. Dattani, S., Samborska, V., Ritchie, H., Roser, M. Cardiovascular Diseases. Preprint at (2023).
48. Jimenez-Tellez, N. & Greenway, S. C. Cellular models for human cardiomyopathy: What is the best option? http://www.wjgnet.com/ 11, 221–235 (2019).
49. Buonaiuto, G., Desideri, F., Taliani, V. & Ballarino, M. Muscle Regeneration and RNA: New Perspectives for Ancient Molecules. Cells 2021, Vol. 10, Page 2512 10, 2512 (2021).
50. Micheletti, R. et al. The long noncoding RNA Wisper controls cardiac fibrosis and remodeling. Sci Transl Med 9, (2017).
51. Cai, B. et al. The Long Noncoding RNA CAREL Controls Cardiac Regeneration. J Am Coll Cardiol 72, 534–550 (2018).
52. Ponnusamy, M. et al. Long Noncoding RNA CPR (Cardiomyocyte Proliferation Regulator) Regulates Cardiomyocyte Proliferation and Cardiac Repair. Circulation 139, 2668–2684 (2019).
53. Porrello, E. R. et al. Transient regenerative potential of the neonatal mouse heart. Science (1979) 331, 1078–1080 (2011).
54. Porrello, E. R. et al. Regulation of neonatal and adult mammalian heart regeneration by the miR-15 family. Proceedings of the National Academy of Sciences 110, 187–192 (2013).
55. Lam, N. T. & Sadek, H. A. Neonatal Heart Regeneration: A Comprehensive Literature Review. Circulation 138, 412 (2018).
56. Lenzi, J. et al. ALS mutant FUS proteins are recruited into stress granules in induced pluripotent stem cell-derived motoneurons. DMM Disease Models and Mechanisms 8, 755–766 (2015).
57. Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature 516, 423–428 (2014).
58. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 1–21 (2014).
60. Kolde, R. pheatmap: Pretty Heatmaps. Preprint at (2019).
61. Xie, Z. et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc 1, (2021).
62. Chen, E. Y. et al. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 1–14 (2013).
63. Kuleshov, M. V et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44, (2016).

64. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550 (2005).
65. Mootha, V. K. et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics 2003 34:3 34, 267–273 (2003).
66. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat Methods 14, 417 (2017).
67. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 2012 9:7 9, 671–675 (2012).
68. Astro, V., Ramirez-Calderon, G. & Adamo, A. Protocol to measure calcium spikes in cardiomyocytes obtained from human pluripotent stem cells using a ready-to-use media. STAR Protoc 4, (2023).

Acknowledgments

The authors acknowledge Prof. Alessandro Rosa for kindly providing the WT1 hiPSCs, Dr. Davide Mariani and the Genomic Facility of the Center for Human Technologies of IIT for support in RNA-sequencing experiments and Dr. Adriano Setti and Dr. Alessandra Luppino for helpful discussion. This work was supported by grants from Sapienza University [AR223188B40CB2D0] to GB and from Sapienza University [RM12117A5DE7A45B, RM123188F6B80CE4], Project ‘National Center for Gene Therapy and Drugbased on RNA Technology’ [CN00000041] financed by NextGenerationEU PNRR MUR – M4C2 – Action 1.4- Call ‘Potenziamento strutture di ricerca e di campioni nazionali di R&S’ [CUP: B83C22002870006], PRIN 2022 – Progetti di Rilevante Interesse Nazionale [2022BYB33L, CUP: B53D23016100006], PRIN 2022 PNRR – Progetti di Rilevante Interesse Nazionale [P2022FFEWN, CUP: B53D23026140001] to MB.

Author contributions

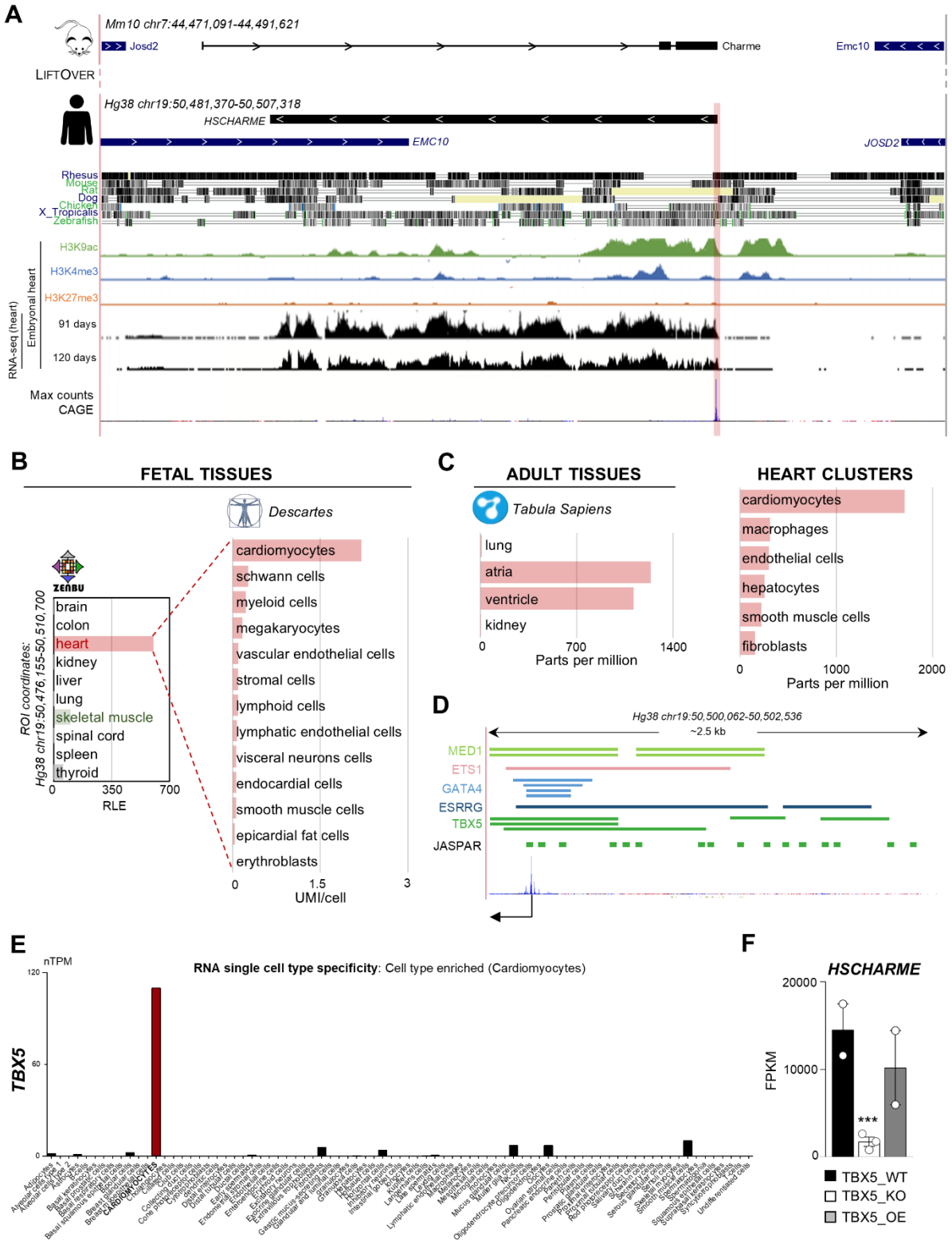
M.B. designed, conceived and acquired funding for the study; G.B developed the *HSCHARME*^{PA} and *HSCHARME*^{RHE} CRISPR-Cas9 edited cell lines and performed all the gene expression, morphological and functional analyses in hiPSC-derived cardiomyocytes; F.D. developed the *HSCHARME*$^{\Delta P}$ cell line and provided support for the gene expression analyses; G.S. provided support with the hiPSC handling and gene expression analyses; A.P. performed the computational analyses on the *HSCHARME*^{WT}, *HSCHARME*$^{\Delta P}$, GSE130036 and GSE116250 datasets. T.S. performed the RNA-FISH experiments; D.T. performed the flow cytometry assay. The project was supervised and written by M.B. with major contributions from G.B. and suggestions from all the other authors.

Competing interest

The authors declare no competing interests.

723 **MAIN FIGURES**

Fig. 1



724

725 **Fig. 1. Identification and analysis of the *HSCHARME* lncRNA locus in *Homo sapiens*.**
726 **A)** UCSC visualization showing the chromosome position and the genomic coordinates of
727 Charme gene in mouse and human. Default tracks from "Vertebrate Multiz Alignment &
728 Conservation" are shown together with histone modifications (H3K9ac, green; H3K4me3,
729 blue; H3K27me3, orange) and RNA-seq reads from ENCODE. CAGE-seq track from

FANTOM5 project is also shown. **B)** Left panel: Transcriptional start site (TSS) usage analysis from FANTOM5 CAGE (Phase1 and 2 datasets) taken from the last Zenbu genome browser update (FANTOM5 Human hg19 promoterome) update. *HSCHARME* locus expression in fetal body districts is shown. Bars represent the Relative Logaritmic Expression (RLE) of the Tag Per Million values of *HSCHARME* TSS usage in a specific sample. Right panel: scRNA-seq analysis from fetal tissues as collected from Descartes atlas[23]. Bars represent the UMI/cell. **C)** scRNA-seq analysis showing *HSCHARME* expression in adult tissues, as obtained from Tabula Sapiens[24] organized by full details (left panel) and by tissue and cell type (right panel). Bars represent the Parts Per Million values of *HSCHARME* TSS usage in each of the specific samples. **D)** UCSC visualization of ReMap ChIP-seq track in cardiomyocytes across *HSCHARME* promoter. The analysis was performed by restricting the biotype selection to cardiomyocytes and by focusing on the 2.5 Kb genomic region upstream *HSCHARME* TSS (black arrow) which is characterized by histone markers of open chromatin. TBX5 consensus motifs predicted by JASPAR 2022[25] (relative profile score threshold=80%) are also shown. **E)** Single-cell resolution of *TBX5* expression as reported by the Human Protein Atlas portal. RNA expression is quantified in nTPM (nTPM = normalized Transcripts Per Million). **F)** Quantification of *HSCHARME* expression from RNA-seq analysis performed in wild type (WT), TBX5 knockout and TBX5 overexpression (OE) from hiPSC-derived cardiomyocytes (GSE81585[26]). Plotted values represent the Fragments Per Kilobase of transcript per Million mapped reads. Data information: ***$p < 0.001$, adjusted p-value.

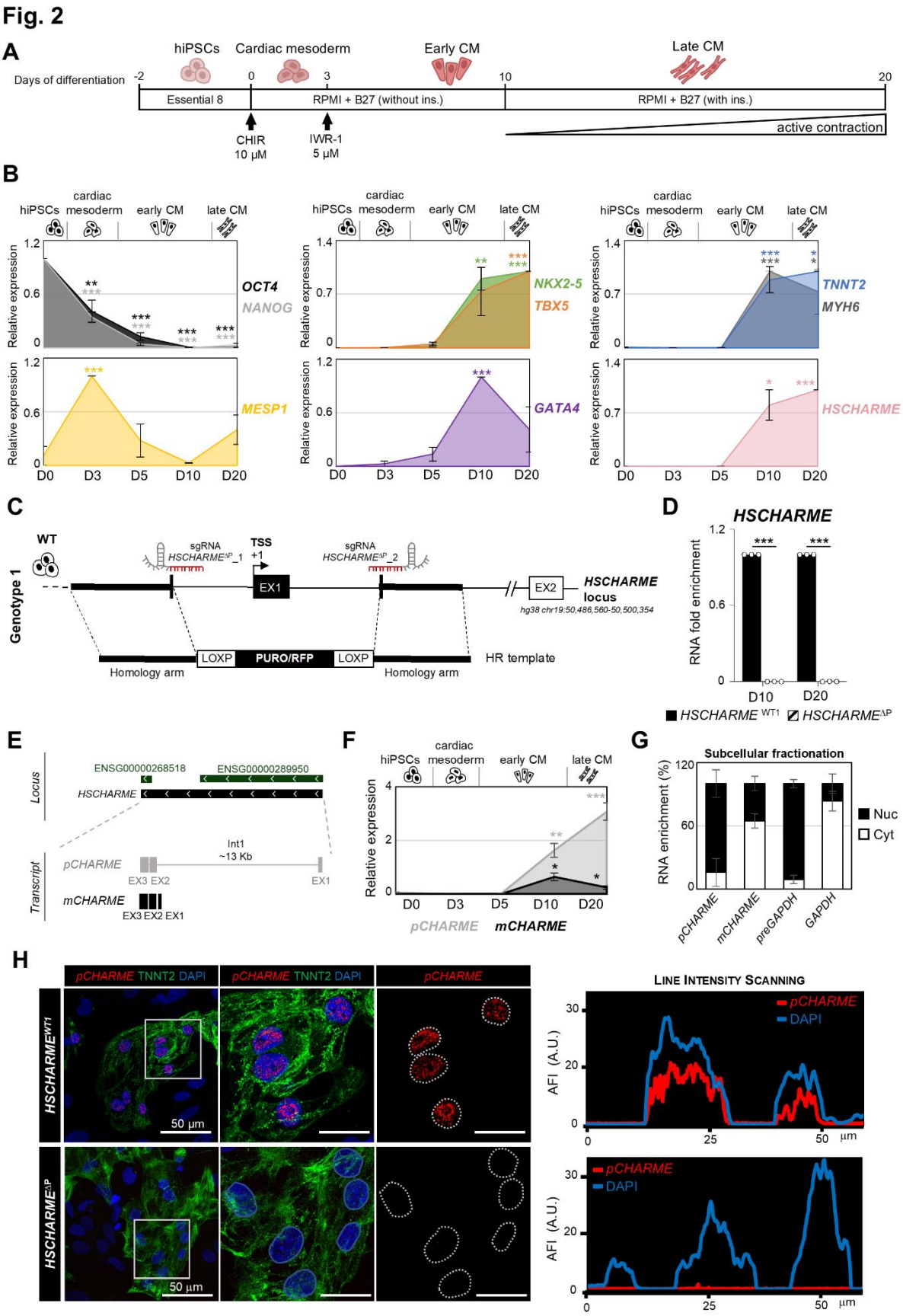


Fig. 2. *HSCHARME* expression and characterization in human iPSC-derived cardiomyocytes. A) Schematic representation of cardiomyocyte differentiation from hiPSCs. **B)** Quantification by RT-qPCR of target gene expression (*OCT4*, *NANOG*, *MESP1*, *NKX2-5*,

TBX5, *GATA4*, *TNNT2* and *MYH6*) as well as *HSCHARME* from total RNA extracted from hiPSCs-derived cardiomyocytes at specific time points (D0-D20). Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological experiments. Statistical tests were performed relative to D0. **C)** Schematic representation of the editing strategy designed to obtain the *HSCHARME*$^{\Delta P}$ hiPSCs genotype using the CRISPR/Cas9 technology. **D)** Quantification by RT-qPCR of *HSCHARME* from *HSCHARME*^{WT1} *vs* *HSCHARME*$^{\Delta P}$ (D10 and D20) cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. **E)** Schematic representation of *HSCHARME* isoforms as reconstructed by genome assembly. **F)** Quantification by RT-qPCR of the *pCHARME* and *mCHARME* isoforms from total RNA extracted from hiPSC-derived cardiomyocytes at specific time points (D0-D20). Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological experiments. Statistical tests were performed relative to D0. **G)** Quantification of the subcellular distribution of *pCHARME* and *mCHARME* isoforms from biochemical fractionation of D10 cardiomyocytes. The histogram shows the RT-qPCR quantification of the relative % of RNA abundance in cytoplasmic *vs* nuclear compartments. Quantification of *GAPDH* mRNA (cytoplasmic) and *preGAPDH* RNA (nuclear) was used to assess the quality of biochemical fractionation. **H)** Left panel: Representative RNA-FISH staining for *pCHARME* (red) combined with TNNT2 immunofluorescence (green) in *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. Panels represent the maximun 2D projection of full-size image of confocal caption, digital magnification of region highlighted by white squares and *pCHARME* signal distribution inside the nuclei highlighted by dotted line. Right panel: plot displaying Average Fluorescence Intensity (AFI) of *pCHARME* RNA-FISH and DAPI signals of a single focal plane. Overlapping of the lines indicates colocalization between the signals scanned along the horizontal distance of the selection (white rectangle). Data information: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, unpaired Student's t-test.

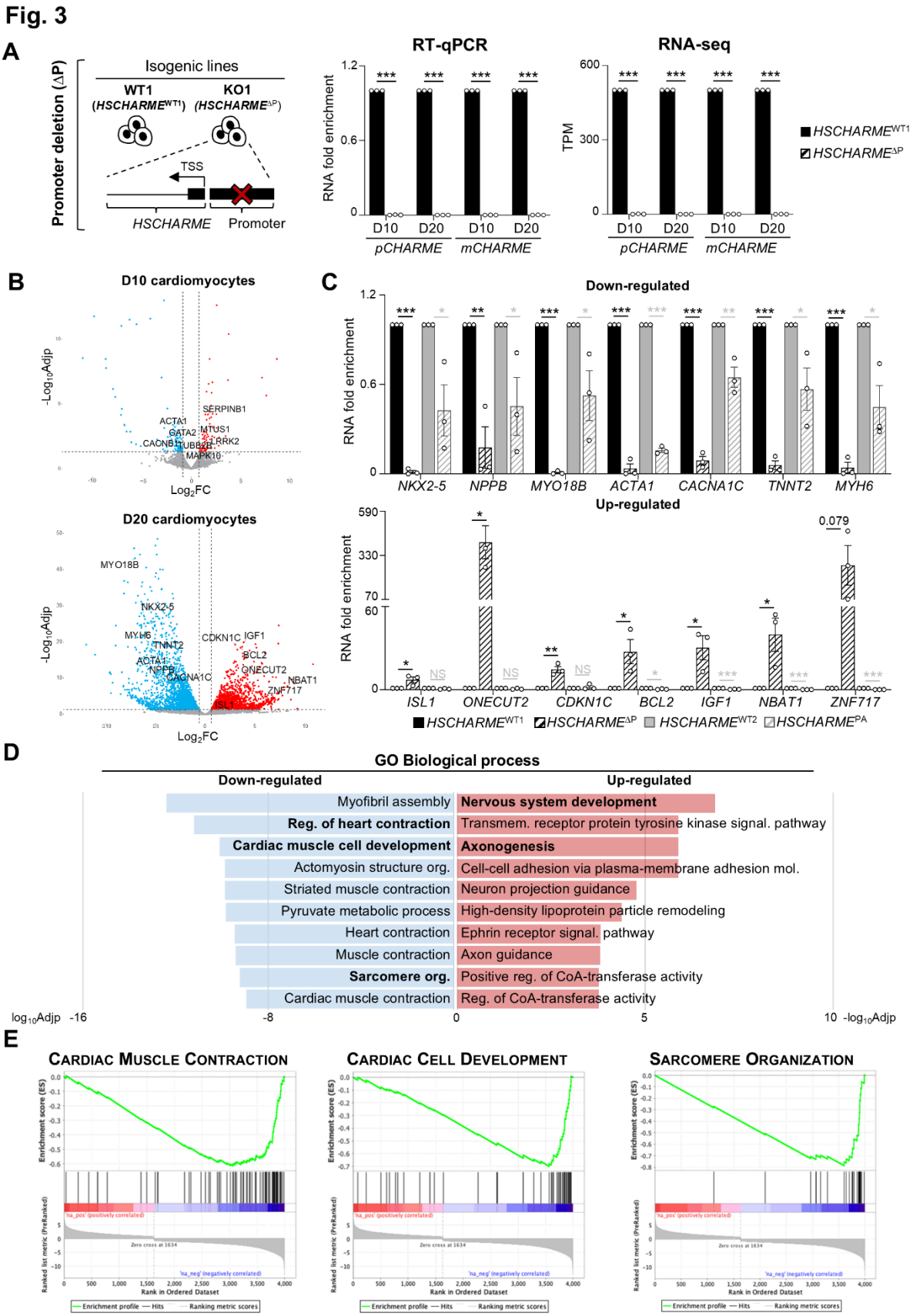


Fig. 3. Genome-wide profiling of *HSCHARME*^{WT} and *HSCHARME*$^{\Delta P}$ hiPSC-derived cardiomyocytes. A) Left panel: Schematic representation of the genome editing strategy designed to obtain the *HSCHARME*$^{\Delta P}$ hiPSCs genotype using the CRISPR/Cas9 technology.

Middle panel: Quantification by RT-qPCR of *pCHARME* and *mCHARME* isoforms from *HSCHARME*^{WT1} *vs HSCHARME*$^{\Delta P}$ cardiomyocytes (D10 and D20). Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. Right panel: Quantification by RNA-seq (TPM) of *pCHARME* and *mCHARME* isoforms expression in *HSCHARME*^{WT1} *vs HSCHARME*$^{\Delta P}$ hiPSC-derived D10 and D20 cardiomyocytes. **B)** Volcano plots showing differentially expressed genes from the RNA-seq analysis of *HSCHARME*^{WT1} *vs HSCHARME*$^{\Delta P}$ hiPSC-derived cardiomyocytes D10 *vs* CTRL10 (upper panel) and D20 *vs* CTRL20 (lower panel). Significantly up-regulated (FDR<0.05; FC>1) and down-regulated (FDR<0.05; FC<-1) genes are represented in red and blue, respectively. **C)** Quantification by RT-qPCR of down-regulated (upper panel) and up-regulated (lower panel) DEGs in *HSCHARME*^{WT1} *vs HSCHARME*$^{\Delta P}$ (black) and *HSCHARME*^{WT2} *vs HSCHARME*^{PA} (grey) hiPSC-derived D20 cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. **D)** Gene ontology (GO) enrichment analysis performed with EnrichR on down-regulated (left) and up-regulated (right) DEGs in *HSCHARME*^{WT1} *vs HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. Bars indicate +/–$\log_{10}$ adjusted p-value ($\log_{10}$Adjp and –$\log_{10}$Adjp) of the top enriched biological processes. All the represented categories show an Adjp<0.05. **E)** GSEA plot showing the enrichment of "cardiac muscle contraction", "cardiac cell development" and "sarcomere organization" processes down-regulated in *HSCHARME*^{WT1} *vs HSCHARME*$^{\Delta P}$ D20 cardiomyocytes. Data information: *p < 0.05, **p < 0.01, ***p < 0.001; unpaired Student's t-test.

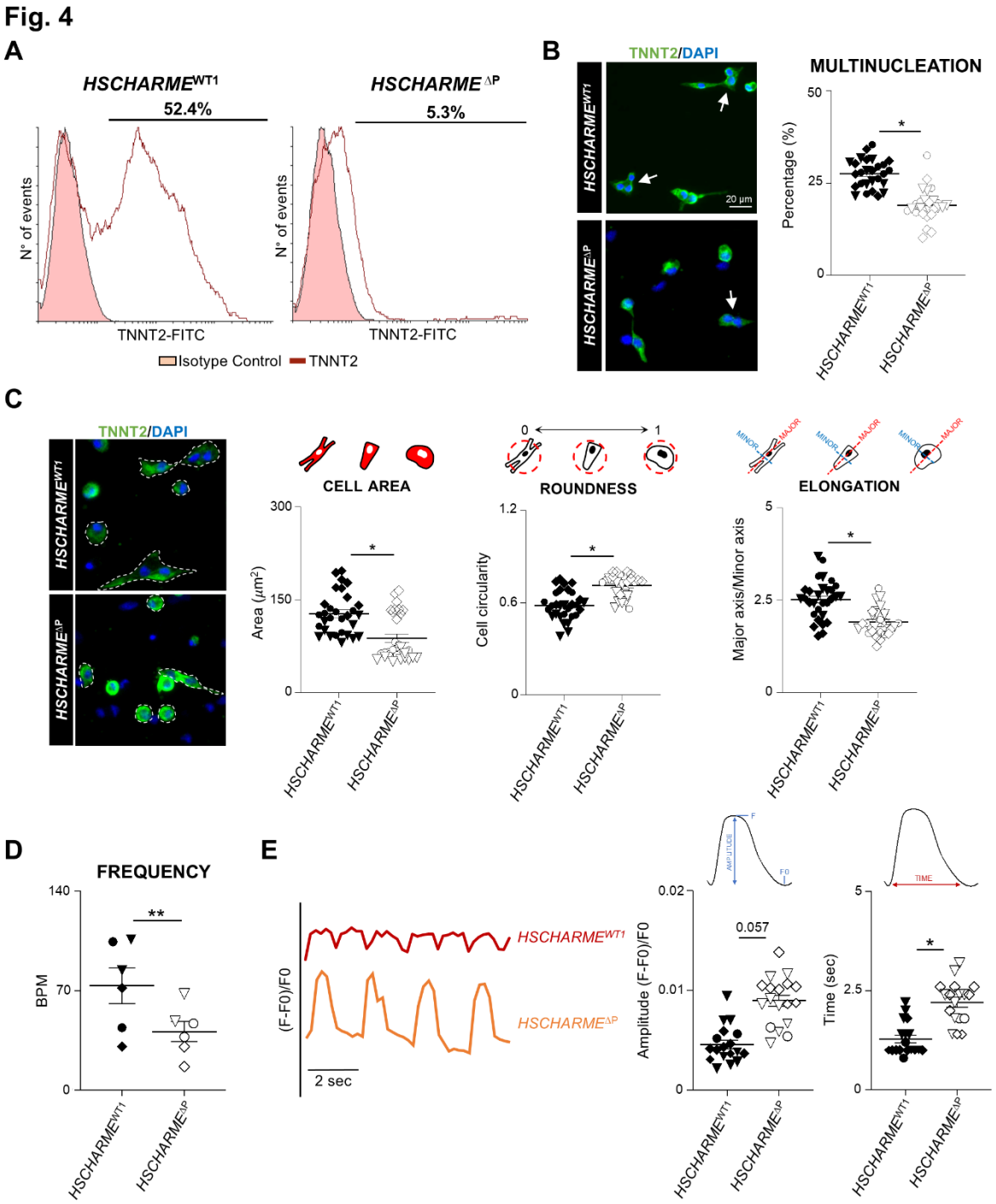


Fig. 4. Functional implication of *HSCHARME* in human cardiomyogenesis and contraction capabilities. A) Representative expression profiles of cardiac troponin-T (TNNT2) of *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. The pink shadow indicates isotype control while the bordeaux line indicates TNNT2 staining. Numbers indicate the percentage of TNNT2-positive cells relative to the isotype control. **B)** Left panel: Representative images of *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes stained for TNNT2 (green) and DAPI (blue) showing cells with two/multiple nuclei (white arrows). Right panel: Percentage (%) of multinucleated/total nuclei ratio from *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. Data are expressed as mean ± SEM of 3 biological replicates. **C)** Left panel: Representative images of *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes stained for TNNT2

(green) and DAPI (blue) and used for morphological analyses (white dashed outlines). Right panel: morphological measurements of CM cell area, roundness and elongation. Black bars represent means ± SEM of 3 biological replicates. **D)** Quantification of the beat rate of *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. Data are expressed as beats per minute (BPM) and represent mean ± SEM of 3 biological replicates. **E)** Left panel: Representative traces of spontaneous contraction activity of *HSCHARME*^{WT1} (red) and *HSCHARME*$^{\Delta P}$ (orange) hiPSC-derived D20 cardiomyocytes. Traces show the changes in light conditions of brightfield 2D videos that are indicative of cell movements (F) as compared to resting cardiomyocytes (F0). Right panel: Quantification of the amplitude (F-F0/F0) and the duration (time in sec) of single contraction events of *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. Data are represented as means +SEM of 3 biological replicates and n=18 peaks. Data information: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, paired Student's t-test.

Fig. 5

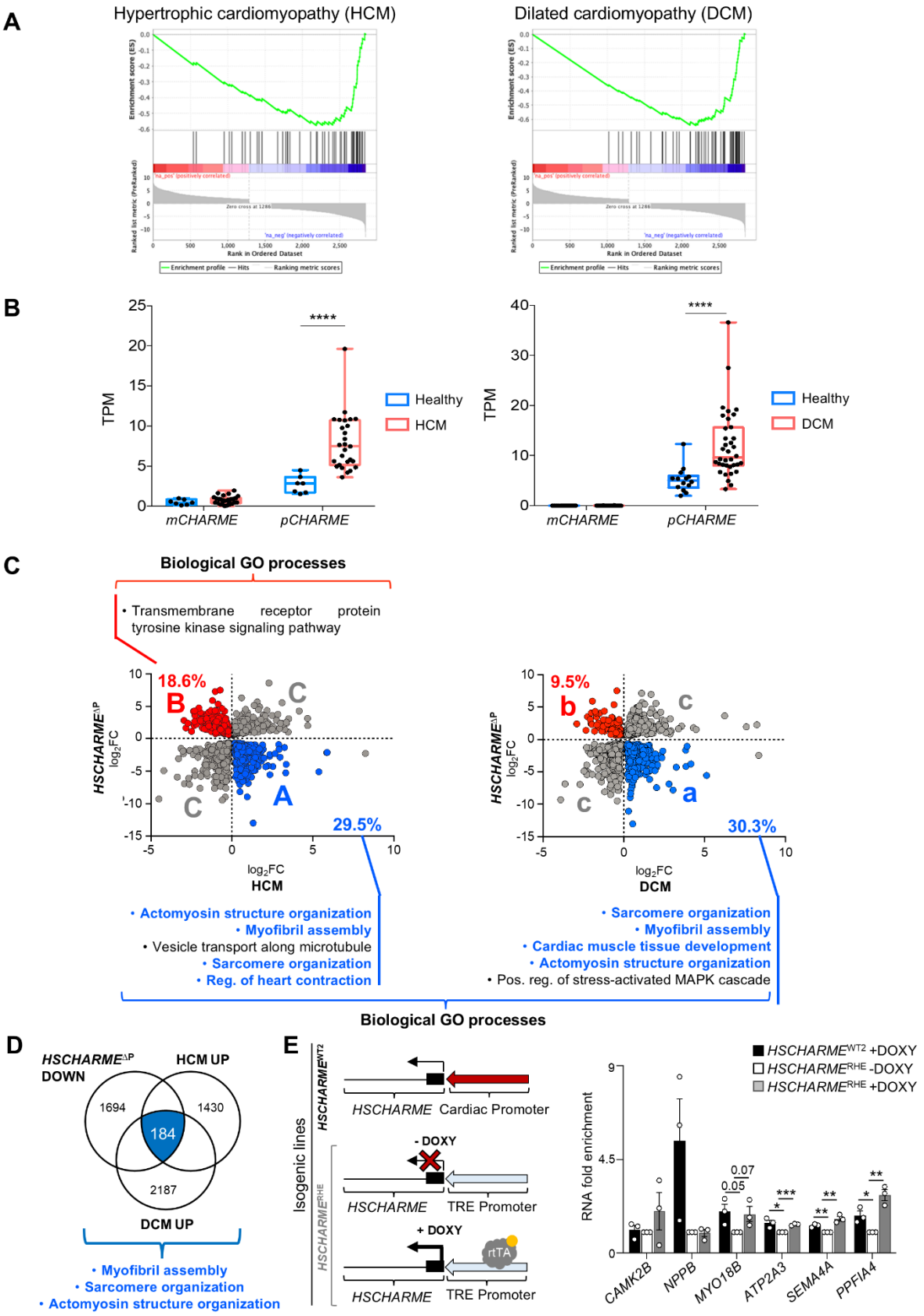


Fig. 5. *HSCHARME* dosage impacts the expression of disease-linked genes. A) GSEA plot showing the enrichment of the Human Phenotype Ontology terms "Hypertrophic Cardiomyopathy" and "Dilated Cardiomyopathy" resulting strongly down-regulated in

HSCHARME$^{\Delta P}$ *vs HSCHARME*^{WT1} D20 cardiomyocytes. **B)** Quantification of *mCHARME* and *pCHARME* isoforms expression from RNA-seq analysis performed in healthy (CTRL) *vs* hypertrophic (left panel) and dilated (right panel) cardiomyopathy patients. Data information: ****p<0.0001, adjusted p-value **C)** Scatterplots representing the log_2FC comparison between DEGs in *HSCHARME*$^{\Delta P}$ and HCM (left panel) and DCM (right panel), showing up- and down-regulated genes with opposite trends in the conditions. Percentage of de-regulated genes are reported as well as Biological Process GO categories in which they cluster. All reported categories show an Adjp <0.05. **D)** Venn diagrams depicting the intersection between *HSCHARME*$^{\Delta P}$ *vs HSCHARME*^{WT1} down-regulated DEGs (*HSCHARME*$^{\Delta P}$ DOWN) with Healthy *vs* HCM and Healthy *vs* DCM up-regulated DEGs (HCM UP and DCM UP, respectively). **E)** Left panel: Schematic representation of the editing strategy designed to obtain the *HSCHARME*^{RHE} hiPSCs genotype using the CRISPR/Cas9 technology. Right panel: RT-qPCR quantification of the expression of disease-linked genes in *HSCHARME*^{WT2}, *HSCHARME*^{RHE} (-DOXY) and *HSCHARME*^{RHE} (+DOXY) hiPSC-derived cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. Data information: *p < 0.05; *p < 0.01; ***p < 0.001; unpaired Student's t-test.

Fig. 6

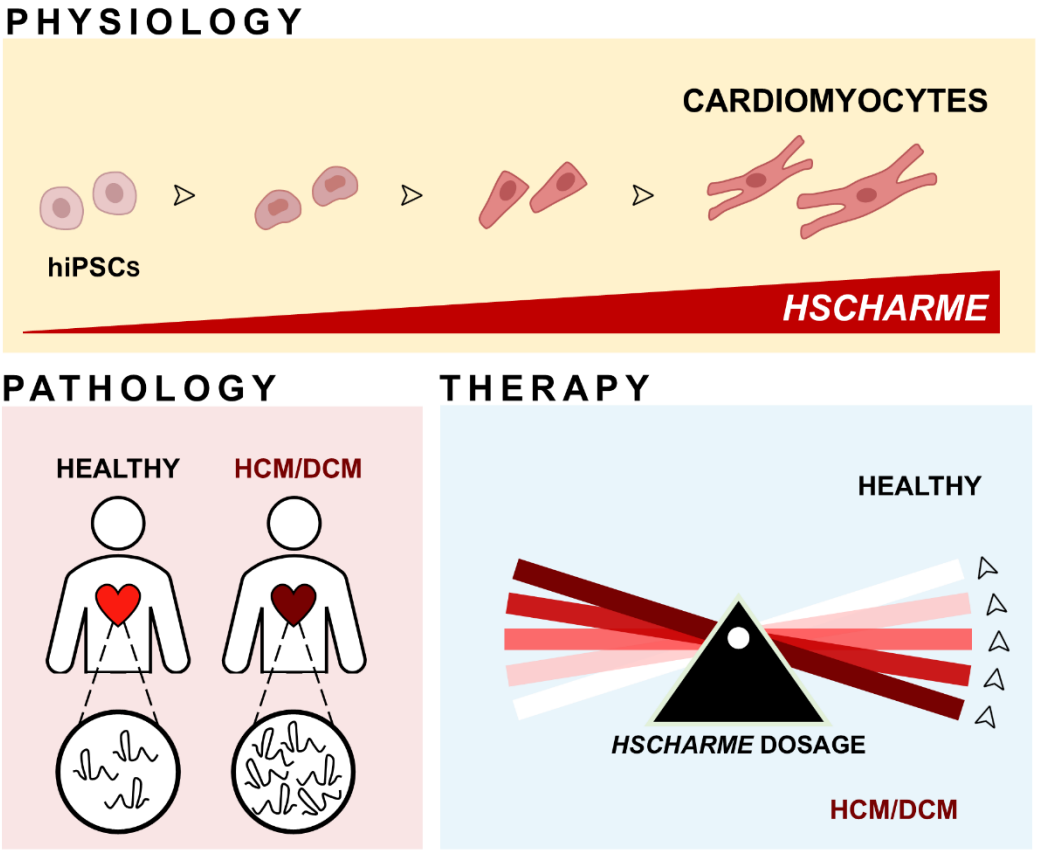
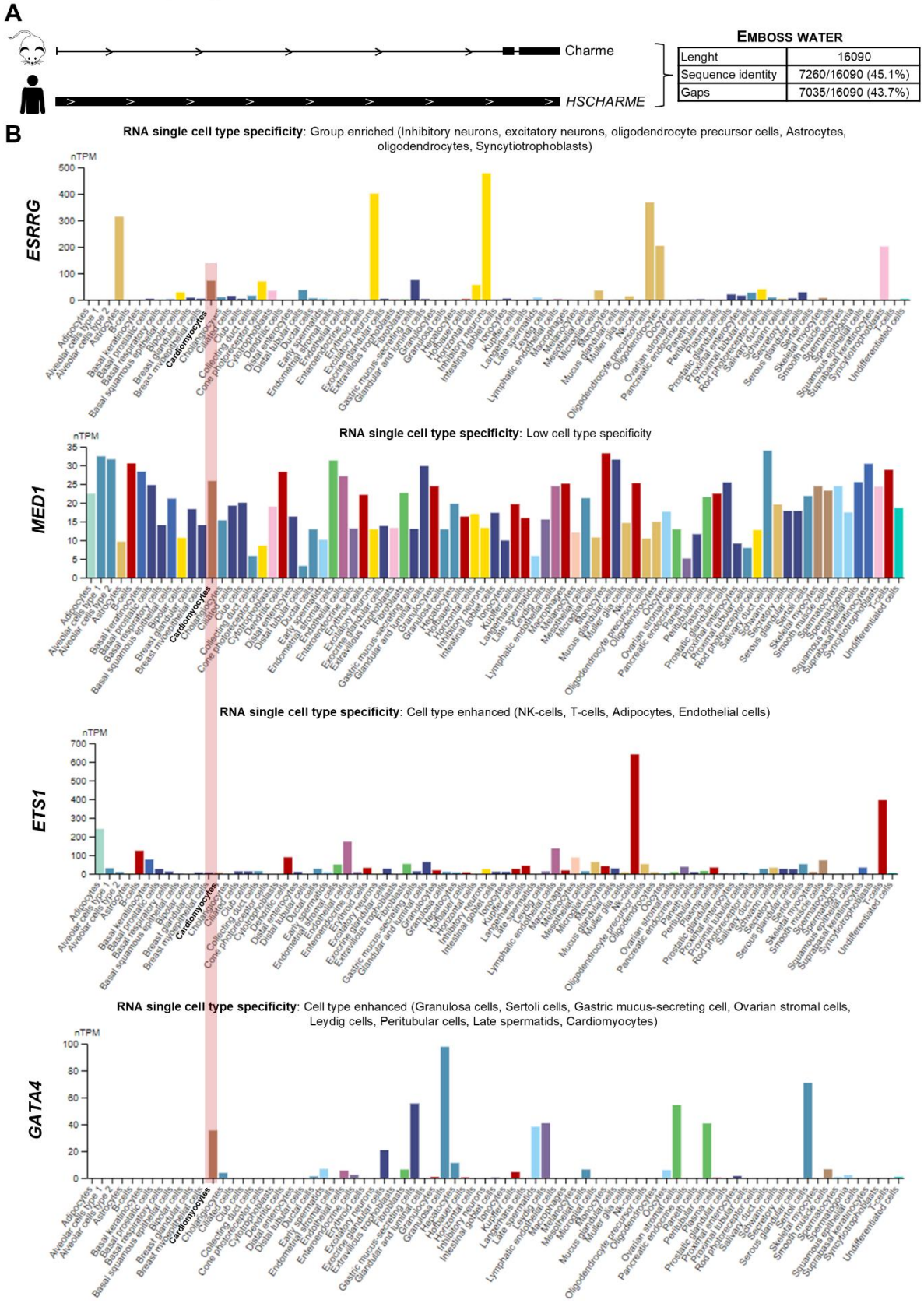


Fig. 6. Schematic overview of *HSCHARME* role in human cardiomyocyte. In physiological conditions (Physiology), *HSCHARME* dosage is crucial to positively regulate the expression of genes involved in cardiomyocyte maturation and contraction. In pathological conditions (Pathology), *HSCHARME* expression results increased in patients affected by DCM and HCM. As such, a novel therapeutic approach (Therapy) can be envisaged, using the modulation of *HSCHARME* dosage to regulate the expression of its target genes.

SUPPLEMENTARY FIGURES

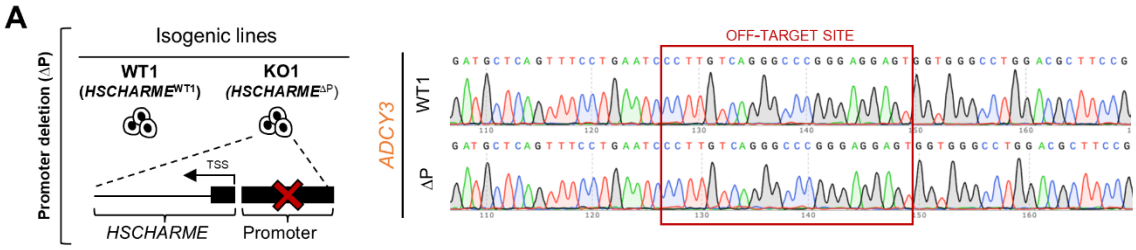
Supplementary Fig. 1



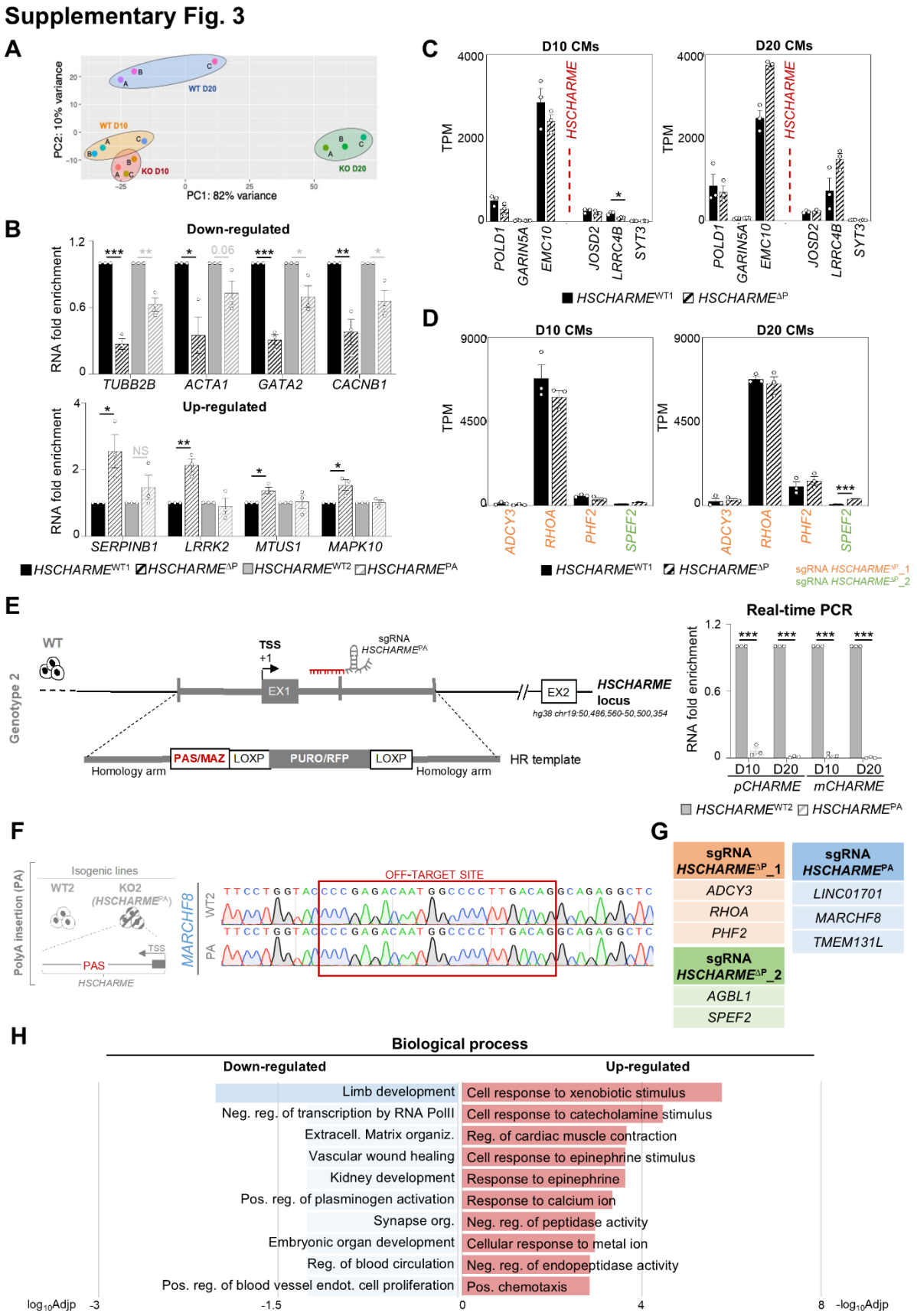
Supplementary Fig. 1. A) Table reports the sequence conservation values between the mouse and human Charme genes. The data were obtained by local sequence using the Smith–

Waterman algorithm available at Emboss water (https://www.ebi.ac.uk/jdispatcher/psa/emboss_water). **B)** *ESRRG*, *MED1*, *ETS1* and *GATA4* gene expression levels at single cell resolution reported in the Human Protein Atlas Portal. RNA expression is quantified in nTPM (nTPM = normalized Transcripts Per Million).

Supplementary Fig. 2



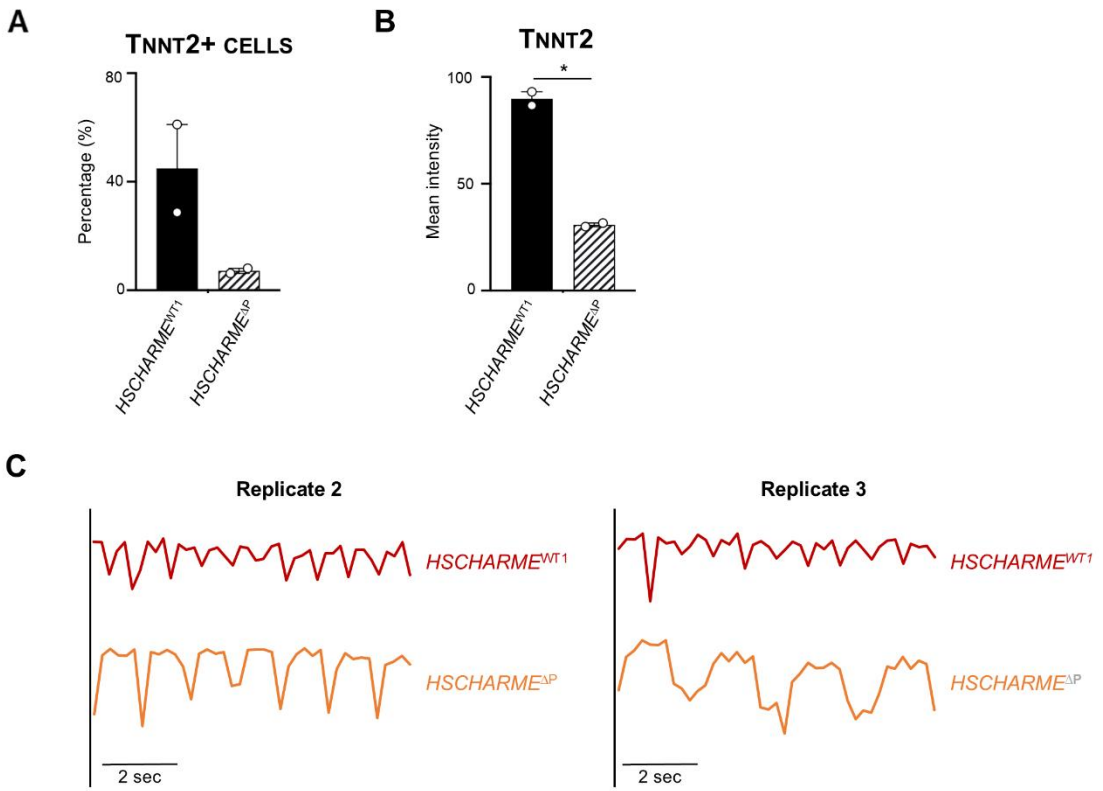
Supplementary Fig. 2. A) Representative example of sequencing results from *HSCHARME*^{WT1} (WT1) and *HSCHARME*$^{\Delta P}$ (ΔP) on the predicted off-target ADCY3 gene. The red square indicates the region interested by the off-targeting. A schematic representation of the genome editing strategy is represented on the left.



Supplementary Fig. 3. A) Principal Component Analysis (PCA) performed on *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ RNA-seq datasets at D10 and D20 hiPSC-derived cardiomyocytes. **B)** RT-qPCR quantification of down-regulated (upper panel) and up-regulated

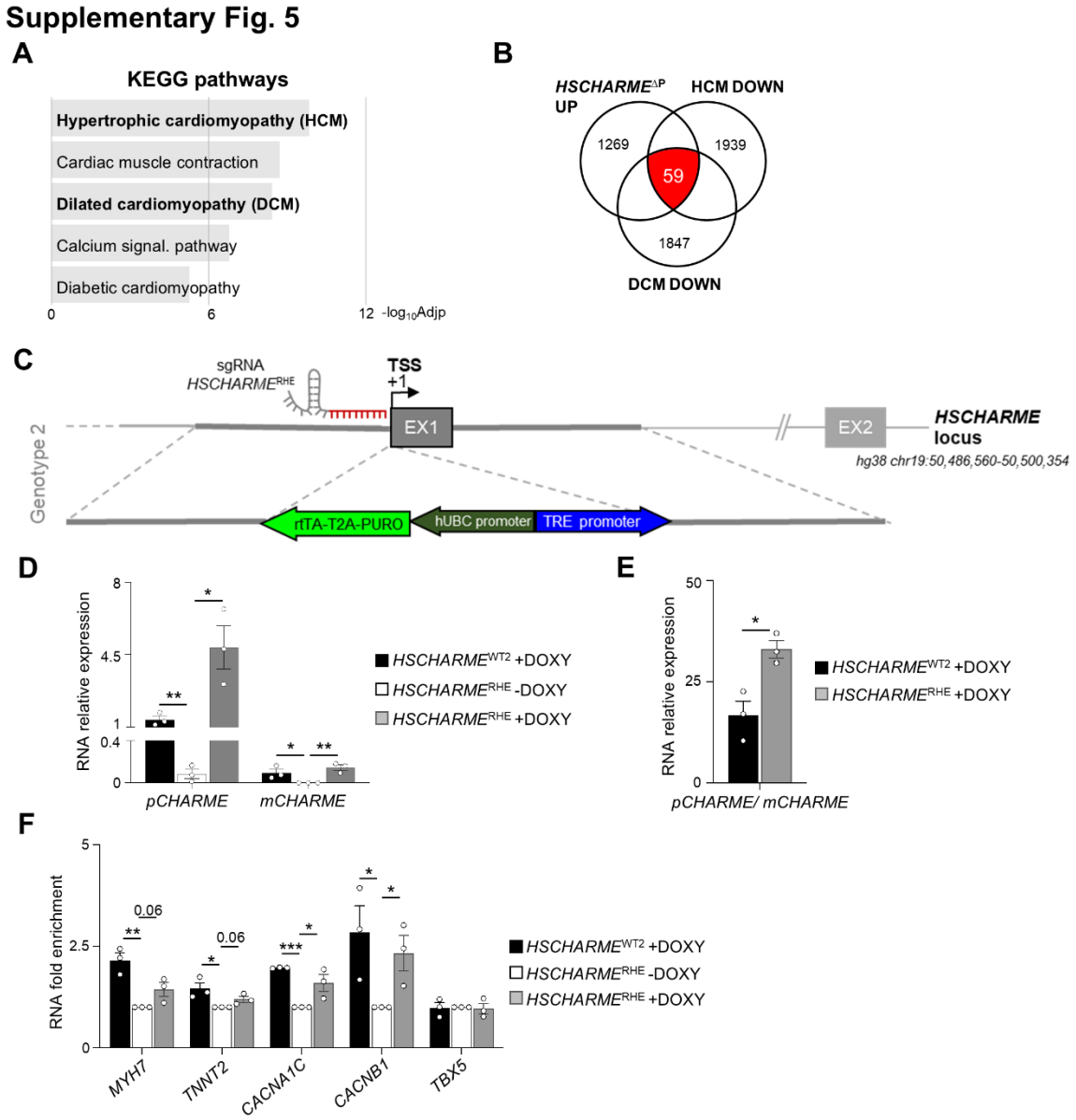
(lower panel) DEGs in *HSCHARME*^{WT1} vs *HSCHARME*$^{\Delta P}$ (black) and *HSCHARME*^{WT2} vs *HSCHARME*^{PA} (grey) hiPSC-derived D10 cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. **C)** Quantification by RNA-seq (TPM) of *HSCHARME*-neighbouring genes expression in *HSCHARME*^{WT1} vs *HSCHARME*$^{\Delta P}$ hiPSC-derived D10 (left panel) and D20 (right panel) cardiomyocytes. **D)** Quantification by RNA-seq (TPM) of sgRNAs off-targets expression in *HSCHARME*^{WT1} vs *HSCHARME*$^{\Delta P}$ hiPSC-derived D10 (left panel) and D20 (right panel) cardiomyocytes. **E)** Left panel: Representation of the genome editing strategy design followed to obtain the *HSCHARME*^{PA} hiPSC cell line by CRISPR/Cas9. Right panel: RT-qPCR amplification of *pCHARME and mCHARME* from *HSCHARME*^{WT2} vs *HSCHARME*^{PA} D10 and D20 cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. **F)** Representative example of sequencing results from *HSCHARME*^{WT2} (WT2) and *HSCHARME*^{PA} (PA) on the predicted off-target *MARCHF8.* The red square indicates the region interested by the off-targeting. A schematic representation of the genome editing strategy is represented on the left. **G)** Table reporting the main off-targets for each guide used for CRISPR/Cas9-based genome editing. H) Gene ontology (GO) enrichment analysis performed with EnrichR on down-regulated (left) and up-regulated (right) DEGs in *HSCHARME*^{WT1} vs *HSCHARME*$^{\Delta P}$ hiPSC-derived D10 cardiomyocytes. Bars indicate the +/– log_{10} adjusted p-value (log_{10}Adjp and –log_{10}Adjp) of the top enriched biological processes. Categories represented by darker bars show an Adjp <0.05. Data information: *p < 0.05, **p < 0.01, ***p < 0.001, unpaired Student's t-test.

Supplementary Fig. 4



Supplementary Fig. 4. A) Flow cytometry quantification of cardiac troponin-T (TNNT2) positive cells percentage in *HSCHARME*^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. Data represent mean ± SEM of 2 biological replicates. **B)** Flow cytometry quantification of cardiac troponin-T (TNNT2) mean intensity in TNNT2+ cells in

HSCHARME^{WT1} and *HSCHARME*$^{\Delta P}$ hiPSC-derived D20 cardiomyocytes. Data represent mean ± SEM of 2 biological replicates. Data information: *p < 0.05; paired Student's t-test. **C)** Representative traces of spontaneous contraction activity of *HSCHARME*^{WT1} (red) and *HSCHARME*$^{\Delta P}$ (orange) hiPSC-derived D20 cardiomyocytes biological replicates. Traces show the changes in light conditions of brightfield 2D videos that are indicative of cell movements (F) as compared to resting cardiomyocytes (F0).



Supplementary Fig. 5. A) KEGG pathways enrichment analysis performed with EnrichR on significant DEGs from *HSCHARME*$^{\Delta P}$ *vs HSCHARME*^{WT1} D20 cardiomyocytes. Bars indicate the top categories of biological processes in decreasing order of $-\log_{10}$ adjusted p-value ($-\log_{10}$Adjp). All the represented categories show an Adjp value <0.05. **B)** Venn diagrams depicting the intersection between *HSCHARME*$^{\Delta P}$ vs *HSCHARME*^{WT1} up-regulated DEGs (*HSCHARME*$^{\Delta P}$ UP) with Healthy *vs* HCM and Healthy *vs* DCM down-regulated DEGs (HCM DOWN and DCM DOWN, respectively). **C)** Schematic representation of the genome editing strategy design followed to obtain *HSCHARME*^{RHE} hiPSC cell line using CRISPR/Cas9 technology. **D)** RT-qPCR quantification of *pCHARME* and *mCHARME* expression in *HSCHARME*^{WT2}, *HSCHARME*^{RHE} (-DOXY) and *HSCHARME*^{RHE} (+DOXY) hiPSC-derived

cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. **E)** RT-qPCR quantification of *pCHARME/mCHARME* ratio in *HSCHARME*^{WT2} and *HSCHARME*^{RHE} (+DOXY) hiPSC-derived cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. **F)** RT-qPCR quantification of *HSCHARME* target RNAs in *HSCHARME*^{WT2}, *HSCHARME*^{RHE} (-DOXY) and *HSCHARME*^{RHE} (+DOXY) hiPSC-derived cardiomyocytes. Data were normalized to *ATP5O* mRNA and represent means ± SEM of 3 biological replicates. Data information: *$p < 0.05$; **$p < 0.01$; ***$p<0.001$; unpaired Student's t-test.