Long noncoding RNA *HSCHARME* is altered in human cardiomyopathies and promotes stem cell-derived cardiomyocyte differentiation by splicing regulation.

Giulia Buonaiuto[1], Fabio Desideri[2,4], Adriano Setti[1], Alessandro Palma[1], Angelo D'Angelo[1], Giulio Storari[1], Tiziana Santini[1], Pietro Laneve[1,3], 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.

[4]present address: Center for Molecular Biology Severo Ochoa (CBM Severo Ochoa) CSIC/UAM, Madrid 28049, Spain.

[*]*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; Splicing; 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. By investigating the mechanism of action of *HSCHARME* in hiPSC-derived cardiomyocytes, we found that the locus produces the major *pCHARME* isoform that associates with SC35-containing speckles and interacts with the splicing regulator PTBP1. Consistently, the functional inactivation of *pCHARME* influences the splicing of cardiac-specific pre-mRNAs and impacts their expression, which parallels a decline in cardiomyocyte differentiation and physiology. In line with a possible association with disease, large-scale analysis of the lncRNA expression across cardiomyopathy patients revealed increased levels of *pCHARME* in hypertrophic (HCM) and dilated (DCM) hearts and identified a subset of disease-associated targets whose expression can be modulated through *HSCHARME* dosage.

By unlocking mechanistic insights into the role of *pCHARME* in cardiac cells, our data identify a novel non-coding regulator of cardiomyocyte function with potential implications in disease.

INTRODUCTION

Heart diseases represent the primary cause of death globally[1]. As the cardiac muscle fails to renew cardiomyocytes (CM)[2] regenerative medicine therapies are considered crucial for stimulating the repair of damaged hearts, thereby reducing the need for transplants. Nonetheless, the restoration of fully matured and functional cardiomyocytes is still challenging due to the partial knowledge of the factors regulating their endogenous development and maturation[3]. Over the years, considerable attention has been placed on coding genes[4]. More recently, the increasing discovery of long-noncoding RNAs (lncRNAs) has revealed unexplored modalities that are essential for CM to acquire their identity and functionality[5,6].

LncRNAs represent a heterogeneous class of transcripts longer than 500 nucleotides, with limited protein-coding capability[7]. Since their discovery, they have been implicated in regulating every step of the gene life cycle[8], as well as a wide range of processes, from physiological to pathological[9]. The remarkable temporal and tissue-specific expression of these molecules designates them as optimal regulators of physiological organ development. In the heart, several studies conducted in murine models have underscored the critical roles of lncRNAs in sustaining cardiac homeostasis, with their dysregulation frequently associated to

69 the onset of disease[6,10–14]. Despite the presence of many orthologues, their direct investigations
70 in human cells or tissues remain scarce. Nonetheless, several lncRNAs have been implicated
71 in human CM processes, including proliferation and regeneration[5,15,16].
72 In the nucleus, lncRNAs fine-tune gene expression by interacting with DNA/RNA molecules,
73 RNA binding proteins, or by shaping the genome three-dimensional organization[17]. Their role
74 as RNA scaffolds is exemplified by their ability to form RNA-rich condensates and create high-
75 local molecule concentrations at specific loci. This is the case of *NEAT1* and *MALAT1*, which
76 accumulate within the nucleus as RNA components of specific sub-nuclear compartments
77 enriched in splicing factors, thereby regulating pre-mRNA splicing[18,19]. In mice, we have
78 previously identified *pCharme*, an architectural lncRNA coordinating the activation of pro-
79 myogenic genes at specific nuclear condensates[20,21]. In the heart, the genetic ablation of
80 *pCharme* causes persistent expression of fetal-like gene programs delaying CM maturation,
81 which ultimately leads to dilated cardiomyopathy[22]. This highlights an important role for the
82 lncRNA in the pathophysiology of the heart and encourages further analyses of its possible
83 contribution in humans. Built on this rationale, we used comparative genomics to further
84 investigate conserved synteny at the previously identified *pCharme* locus[20]. We found that the
85 human syntenic gene, *HSCHARME*, is expressed in the human heart, both in fetal and adult
86 CM. By using cardiomyocytes derived from human induced pluripotent stem cells (hiPSC-
87 CM), we show that the *pCHARME* transcript localises close to SC35-containing speckles and
88 directly binds to the Polypyrimidine tract-binding protein 1 (PTBP1) to influence the splicing
89 of cardiac-specific pre-mRNAs involved in CM differentiation. Finally, we found that the
90 endogenous *pCHARME* expression is significantly increased in dilated (DCM) and
91 hypertrophic (HCM) human cardiomyopathies, correlating with dysregulation of disease-
92 associated genes. Overall, these results identify *pCHARME* as a new gene linked to cardiac
93 disease, with potential implications for therapeutic approaches in human CM.
94
95 **RESULTS**
96 **Identification of conserved synteny of the *pCharme* locus in *Homo sapiens***
97 To explore the phylogenetic conservation of the murine *pCharme*[20,21] and precisely map the
98 chromosomal location of the gene in the human genome assembly, we performed a cross-
99 species analysis through the UCSC genome browser "LiftOver" tool (**Fig. 1A**). Conversion
100 from the murine (mm10; chr7:44,473,538-44,486,138) into the human coordinates revealed the
101 existence of a syntenic locus (hg38; chr19:50,486,482-50,500,386), herein named

102 *HSCHARME* (gene aliases lncFAM[23] and MYREM[24]), whose mapping located a 94-nucleotide
103 extension at the beginning of the gene, as compared to the previous annotation[20].
104 As shown in **Fig. 1A**, we found an overall high-level sequence conservation of the ~26 kb-long
105 region, especially in Rhesus, Mouse, X. Tropicalis and Zebrafish, as well as the synteny of the
106 *EMC10* and *JOSD2* nearby coding genes. More specifically, sequence conservation analysis
107 between the mouse and human genes (intronic and exonic regions) highlighted a degree of 45.1
108 % of sequence identity (**Supplementary Fig. 1A**), which positions *HSCHARME* between the
109 range of well-known lncRNAs, such as *NEAT1* and *XIST*[25,26] (**Supplementary Fig. 1B**).
110 Consistent with the transcriptional activity of the locus, Transcriptional Start Site (TSS)
111 mapping by CAGE (Cap Analysis of Gene Expression) confirmed the presence of a sharp peak
112 on the negative strand of the *HSCHARME* gene, attributable to the existence of a transcript
113 produced in antisense direction (**Fig. 1A**).
114 We then searched available datasets for epigenetic and gene expression signatures across the
115 human gene focusing on cardiac samples. We found that embryonal whole-heart samples show
116 deposition of histone H3 acetyl-lysine 9 (H3K9ac; GSM706849[27]), trimethylation of histone
117 H3 lysine 4 (H3K4me3; GSM772735[27]) and the absence of the repressive trimethylation of
118 histone H3 lysine 27 (H3K27me3; GSM621450[27]) marks, correlating with the transcriptional
119 activation of the locus in fetal hearts (90-120 days), as further confirmed by the gene expression
120 RNA-seq outputs (GSM1059494; GSM1059495) (**Fig. 1A**). CAGE-sequencing data from the
121 FANTOM5 human promoterome catalogue[28] (https://fantom.gsc.riken.jp/zenbu), further
122 confirmed the specific expression of *HSCHARME* in both fetal hearts (84-217 days) and
123 skeletal muscle cells (**Fig. 1B**, left panel). In line with our previous finding in mice[22], scRNA-
124 seq data from Descartes[29] and Tabula Sapiens[30] atlases revealed the highest expression of
125 *HSCHARME* in fetal (72–129 days) cardiomyocytes (**Fig. 1B**, right panel). Restriction of
126 lncRNA expression to cardiomyocytes persists in adult hearts, with no evident difference
127 between atria and ventricles (**Fig. 1C**, left and right panels). Further examination of scRNA-
128 seq data derived from hiPSC differentiated into cardiac cells[31] definitively confirmed the highly
129 specific expression of *HSCHARME* within the CM cluster (**Fig. 1D**, **Supplementary Fig. 1C-**
130 **E** and **Supplementary Table 1,** see **Materials and Methods** for details).
131 Searching for possible regulators of *HSCHARME* expression we mined ReMap ChIP-seq atlas
132 and found the binding of well-known transcription factors in the 2.5 Kb region from the TSS
133 (**Fig. 1E**) with only *TBX5* (T-box transcription factor) showing clear and specific expression
134 in CM (**Supplementary Fig. 1F**). Consistent with the functional implication of TBX5 in
135 *HSCHARME* expression, computational analyses of available transcriptomic datasets from

136 WT, TBX5-KO and TBX5-OE hiPSC-derived CM[31], demonstrate a significant decrease of the
137 lncRNA upon the loss-of-function of TBX5 (**Fig. 1F**). Along with our previous findings in
138 mice[22], these results underscore a conserved role for TBX5 in positively regulating the
139 expression of *HSCHARME* in the human heart. Since the overexpression of TBX5 alone does
140 not influence the levels of the lncRNA, we argue that it may establish its basal transcriptional
141 expression, while other factors could be involved in the transcriptional regulation of the locus
142 in cardiac cells.
143 Overall, these new data provide a high-resolution map of *HSCHARME* expression in the human
144 heart and detail the cell-type specific restriction of the lncRNA to CM. The similarities of
145 *HSCHARME* with its murine counterpart, which has already been shown to play an important
146 role in cardiac remodeling[20–22], encouraged further investigations into the relevance of the
147 human transcript in disease.
148
149 ***HSCHARME* characterization in hiPSC-derived human cardiomyocytes.**
150 Considering the specific expression of *HSCHARME* in CM, we used hiPSCs as a model to
151 study cardiomyogenic differentiation. The applied protocol[32] exploits the induction of the
152 WNT/β-catenin pathway (Day 0, CHIR addition), to guide cells towards mesoderm followed
153 by its subsequent inhibition (Day 3, IWR-1 addition) to promote a CM fate (**Fig. 2A**). In line
154 with the acquisition of the CM identity, RT-qPCR analyses performed at specific time points
155 of hiPSC differentiation show the expected dynamic wave of cardiac gene expression[32] (**Fig.
156 2B**), alongside the onset of spontaneous beating of cellular foci, starting at days 8-10
157 (**Supplementary Video 1**). The expression of *HSCHARME* at day 10 and its progressive
158 increase (**Fig. 2B**) suggest its potential association with human CM commitment and
159 differentiation. To directly test this hypothesis, we used a CRISPR/Cas9-based gene editing
160 technology to generate *HSCHARME* knock-out (KO) hiPSCs. Specifically, sgRNAs were
161 designed to produce a genomic deletion (~4.8 Kb sized) overlapping *HSCHARME* TSS, which
162 also includes the TBX5 binding sites (delta promoter=ΔP) (**Fig. 2C**). Mutant hiPSCs-CM were
163 checked to confirm the correct genomic editing, the abrogation of *HSCHARME* expression
164 (**Supplementary Fig. 2A**) and the absence of possible off-targets (**Supplementary Fig. 2B**,
165 **Supplementary File 1** and **Supplementary Table 2)**. Following these verifications, we
166 proceeded with the transcriptome analysis of WT *versus* isogenic ΔP-CM.
167 *De novo* assembly of reads from the WT condition suggested that two *HSCHARME* full-length
168 isoforms are produced in CM, the partially spliced transcript (precursor) *pCHARME* stably
169 retaining the first intron and the fully spliced transcript (mature) *mCHARME* (**Fig. 2D** and

170 **Supplementary Fig. 2C**). Both isoforms were induced throughout CM differentiation, with
171 *pCHARME* being consistently more abundant than *mCHARME* and enriched to nuclei (**Fig.**
172 **2E-F**). Subcellular fractionation assays followed by RT-qPCR analyses revealed a cytoplasmic
173 enrichment for *mCHARME.* The presence of open reading frames within *mCHARME*, possibly
174 translated into micro-peptides[33–36], was excluded by the Coding Potential Calculator tool
175 (CPC2, http://cpc2.gao-lab.org/index.php). In line with the biochemical fractionation, high-
176 resolution RNA-fluorescence *in situ* hybridization (RNA-FISH) experiments using probes
177 against *pCHARME* confirmed its nuclear localisation (**Fig. 2G**). No fluorescence was detected
178 in ΔP-CM, which proves the specificity of *pCHARME* signals and the efficiency of the lncRNA
179 KO. The latter was further confirmed by i) RT-qPCR analysis of WT and ΔP-CM with
180 *pCHARME* and *mCHARME* specific primers (**Fig. 3A**, middle panel), ii) TPM quantification
181 of the transcriptomic reads and ii) the RNA-seq genomic plots (**Fig. 3A**, right panel and
182 **Supplementary Fig. 2D**).
183 As previously observed for the murine orthologue[22], the staining of *pCHARME* in human CM
184 exhibits a discrete nuclear pattern. Nonetheless, here we noticed that the fluorescent signals
185 encompass multiple foci, which suggests *trans*-regulatory roles for the lncRNA. In this
186 direction, differential expression analysis performed to compare the WT and ΔP transcriptomic
187 datasets, revealed an impact of *HSCHARME* on the expression of 175 and 3637 genes at 10
188 (D10) and 20 (D20) differentiation days, respectively ($-1<\log_2 FC>1$; FDR<0.05) (**Fig. 3B**,
189 **Supplementary Fig. 3A** and **Supplementary Table 3**). Of them, a total of 1465 differential
190 expressed genes (DEGs) were up-regulated (n=68 at D10; n=1397 at D20), whereas 2347 were
191 down-regulated (n=107 at D10; n=2240 at D20) in ΔP-CM. Importantly, none of the putative
192 off-target genes was found among DEGs (**Supplementary Fig. 3B**), which conclusively
193 excludes their involvement in the transcriptomic alterations. Moreover, we did not find
194 alteration in the expression of *pCHARME* neighbouring loci (±260 Kb), which conclusively
195 excludes a role for the lncRNA on nearby gene expression in cardiac cells (**Supplementary**
196 **Fig. 3C**). To strength these results, we produced an independent knock-out cell line (PA-
197 hiPSC), with a different *HSCHARME* mutation and genetic background (WTSIi004-A;
198 referred to as WT2). Specifically, the PA-hiPSC were generated through the insertion of a
199 strong termination cassette inside the *HSCHARME* locus using a strategy previously setup in
200 mice[20] (**Supplementary Fig. 3D**). Upon checking the correct abrogation of *pCHARME* and
201 *mCHARME* (**Supplementary Fig. 3E**), the absence of off-target mutations was confirmed by
202 DNA sequencing (**Supplementary Fig. 3F**, **Supplementary File 1** and **Supplementary Table**
203 **2**). To note, no off-target gene was found in common between the two genotypes, which

204 excludes possible effects due to unwanted Cas9 background activities. Validation by RT-qPCR
205 analyses performed on the top-most down-regulated DEGs from ΔP-CM showed their
206 significant decrease in both the mutant cell lines (**Fig. 3C** and **Supplementary Fig. 3G**). As
207 both CRISPR-Cas9 strategies demonstrate high efficiency in the *pCHARME* knockout (~99%),
208 the minor alterations in gene expression observed within the PA context can be attributed to
209 the intrinsic features of the iPSC-WTSIi004-A background.
210 Gene Ontology (GO) term enrichment analysis performed on transcriptomic datasets revealed
211 that at early stages of CM differentiation (D10), down-regulated DEGs are associated with
212 developmental pathways, such as limb development (**Supplementary Fig. 3H**, left and
213 **Supplementary Table 3**). At later stages (D20), enriched GO categories relate to CM structure
214 and function, including myofibril assembly, heart contraction and cardiac muscle cell
215 development gene classes (**Fig. 3D**, left and **Supplementary Table 3**). Gene set enrichment
216 analysis (GSEA) confirmed the significant downregulation of genes belonging to these
217 categories (**Fig. 3E**), with key examples that include *NKX2-5*, crucial for CM development[37],
218 along with structural and functional components such as *MYH6*[38], *MYO18B*[39], and
219 *CACNA1C*[40]. Conversely, upregulated genes enriched GO categories relates to axonogenesis
220 and nervous system development (**Fig. 3D**, right and **Supplementary Table 3**), may reflect
221 the predisposition of the embryonic stem cells to differentiate toward default neuronal fate in
222 the absence of additional stimuli[41]. We argued that *HSCHARME* depletion, by inhibiting the
223 signals activating CM specification, might promote the upregulation of alternative pathways
224 leading to a default neuronal state.
225 To assess the evolutionarily conservation of *pCHARME* function between human and mouse,
226 significant DEGs from WT and *Charme* KO post-natal hearts (GSE200878[22]) were analyzed,
227 identifying 847 human homologs (**Supplementary Fig. 3I**). Of them, 208 genes (24.55% of
228 the murine DEGs) were commonly deregulated upon *pCHARME* ablation in cardiac muscle
229 (i.e. *MYO18B*, *CACNA1C*, and *NPPB)*, enriched in pathways related to muscle homeostasis
230 (e.g., glycolytic process, heart development), muscle function (e.g., contraction, fatty acid
231 metabolism), and linked to cardiac disorders such as familial atrial fibrillation and hypertrophic
232 cardiomyopathy. Collectively, the transcriptomic data, evolutionary conservation, and the
233 functional interpretation of our datasets support the role of *HSCHARME* as a positive regulator
234 of genes whose expression is physiologically relevant for the differentiation of CM.
235
236 **Functional implication of *HSCHARME* in cardiomyogenesis.**

237 To functionally characterize WT and gene-edited hiPSC-CM, we tested their capacity to
238 contract. By tracking the contraction dynamics through MUSCLEMOTION (**Supplementary**
239 **Fig. 4A**) we found a significant decrease in the beat rate of mutant ΔP-CM as well as the
240 alteration of other parameters, such as time to peak, duration and peak to peak time (**Fig. 4A**).
241 These changes are consistent with a general decrease of the spontaneous beating frequency in
242 ΔP-CM compared to WT cells, with individual contractions becoming longer upon *pCHARME*
243 loss. Importantly, the inspection of cells over differentiation evidenced a significant delay in
244 the onset of beating of mutant CM (ΔP and PA) compared to their WT counterparts (11 *versus*
245 8 days on average, **Fig. 4B** and **Supplementary Fig. 4B**). Other phenotypical traits associated
246 with cellular morphology[42] were influenced by *pCHARME* loss-of-function with mutant CM
247 appearing smaller and rounder than the isogenic WT (**Fig. 4C** and **Supplementary Fig. 4C**),
248 indicating a possible delay in CM differentiation. To deepen our understanding of the timing
249 of *pCHARME* regulation across differentiation, we performed flow cytometry analysis of
250 $PDGFRA^+/CD56^+$ cardiac progenitors and $PDGFRA^+/CD82^+$ cardiac-committed progenitors
251 which show that *pCHARME* loss leads to a 43% reduction in cardiac progenitors and 20%
252 reduction in the cardiac committed progenitors (**Fig. 4D**). Given that CD82 is a key marker of
253 CM fate specification[43], these data align with an early role for the lncRNA in CM specification,
254 which is also coherent with the emergence of default alternative fates observed by
255 transcriptomic analyses. Along the same direction, flow cytometry analysis performed at later
256 stages (D20) shows an 80% reduction of differentiated CM, as inferred by the quantification
257 of TNNT2[44] positive (+) cells (**Fig. 4E** and **Supplementary Fig. 4D**). These findings perfectly
258 match with cellular deconvolution analysis used to interpret our iPSC-derived CM in terms of
259 cell-type composition. Indeed, we found that *pCHARME* ablation causes a reduction of
260 approximately 75% of CM (D20) (**Fig. 4F** and **Supplementary Table 1**), with most of the
261 downregulated DEGs enriched in this population (KL-GSEA, **Supplementary Fig. 4E**).
262 However, as the impact of *pCHARME* ablation in CM was higher than in precursor cells, we
263 asked whether the lncRNA may also play an intrinsic role in differentiated cells. To this end,
264 we used antisense LNA-GapmeRs to deplete *pCHARME* directly in CM (D20). Our findings
265 indicate that 80% reduction of *pCHARME* reduces, although to a lesser extent than in the KO
266 cells, the expression of genes induced with differentiation (i.e *TNNT2*, *MYH7* and *CACNA1C*)
267 (**Fig. 4G**). These findings indicate that *pCHARME* plays a dual regulatory role in the
268 acquisition and maintenance of CM identity by regulating the expression of genes which are
269 crucial for the development and for the physiology of cardiac muscle.
270

271 ***pCHARME* regulates the alternative splicing of cardiac-specific pre-mRNAs.**
272 The high-local distribution of *pCHARME* in the nucleus led us to determine whether specific
273 features are associated with these lncRNA-enriched domains. We first explored the potential
274 for *pCHARME* as a splicing regulator, given the similar sarcomere and contraction defects that
275 splicing factor loss often causes in the heart[45]. In this direction, we combined RNA-FISH
276 targeting *pCHARME* with SC35 immunofluorescence (IF) to determine the proximity of the
277 lncRNA to SC35-containing nuclear speckles[46,47]. Intriguingly, in-depth quantification of
278 three-dimensional signal distribution revealed that *pCHARME* stains in close contact with
279 SC35 domains with 40% of *pCHARME* colocalising with SC35 domains and 60% of SC35
280 domains containing *pCHARME* (**Fig. 5A-B**).
281 The observed RNA-protein proximity led us to test a possible involvement of *pCHARME* in
282 splicing regulation. By large-scale analysis of splicing patterns from our RNA-seq datasets, we
283 identified a total of 9615 expressed isoforms in both WT and ΔP cardiomyocytes (see
284 **Materials and Methods**). We found that the abundance of a consistent fraction of these
285 isoforms (16.3%; corresponding to 1877 transcript isoforms) was significantly altered upon
286 *pCHARME* ablation and functionally linked to cardiomyogenic ontologies related to cardiac
287 muscle cell development and heart contraction (**Supplementary Fig. 5A**). On this preliminary
288 evidence, we used rMATS[48] for alternative splicing (AS) analysis of our RNA-seq datasets,
289 including exon-skipping (ES), intron retention (IR), alternative 3' splice site (A3SS),
290 alternative 5' splice site (A5SS), and mutually excluded exons (MXE) and events. Also in this
291 case, we observed a strong impact of *HSCHARME* ablation, with a total of 6.552 aberrant pre-
292 mRNA splicing events (FDR<0.05, Absolute Inclusion level>0.1) (**Fig. 5C**, **Supplementary**
293 **Fig. 5B** and **Supplementary Table 3**). Splicing alterations were found in each of the analysed
294 classes, with the ES events accounting the 70.1% of all the significant alterations (**Fig. 5C**,
295 right panel). Among the most interesting examples of cardiac genes, we found *CACNB1*, a gene
296 implicated in voltage-dependent calcium release[49]. In KO CM (ΔP and PA), we found that the
297 skipping of exon 7 of *CACNB1* causes a shift from the EX6-EX7-EX8 isoform into the shorter
298 EX6-EX8 one (**Supplementary Fig. 5C-D**). Another example is *MYL6*, a gene encoding for a
299 hexameric ATPase cellular motor protein expressed in muscle and non-muscle tissues[50]. *MYL6*
300 gene produces two mRNA variants which differ in exon 6[51] (**Supplementary Fig. 5E**). We
301 found that *HSCHARME* KO (ΔP and PA) leads to the overabundance of the longest EX5-EX6-
302 EX7 isoform (**Supplementary Fig. 5E-F**), which is described in the literature as the major
303 non-muscle transcript. Therefore, the *HSCHARME*-mediated regulation of MYL6 splicing
304 facilitates the production of a structural myosin variant optimized for CM contraction.

305 To evaluate whether *pCHARME* regulation of alternative splicing is also evolutionarily
306 conserved between human and mouse CM, we performed AS analysis of RNA-seq datasets
307 from WT and *Charme* KO post-natal hearts (GSE200878[22]). We found that, *in vivo*, the lack
308 of *Charme* results in 327 exon exclusion and 265 exon inclusion events (**Supplementary Fig.**
309 **5G** and **Supplementary Table 3**). We mapped the coordinates of these differentially spliced
310 murine exons to the human genome and searched for overlaps with *pCHARME*-regulated
311 exons. Focusing on concordant events, we observed conservation for 10.2% of inclusion events
312 and 8.3% of exclusion events (**Fig. 5D**, upper panel). Notably, these proportions were
313 significantly higher than expected by chance (**Fig. 5D**, lower panel), underscoring the
314 conserved regulatory role for the lncRNA. Among the conserved splicing targets, we identified
315 key cardiac genes, including *TNNT2*[44], *MFF*[52], *SYNC*[53], *SORBS2*[54], *CREM*[55] and *RBFOX2*[56],
316 which are known to play crucial roles in heart function (**Supplementary Table 3)**. Overall,
317 these results indicate *pCHARME* as a novel, evolutionarily conserved regulator of CM splicing.
318
319 ***pCHARME* modulates alternative splicing by direct interaction with PTBP1**
320 To gain mechanistic insights into the splicing alterations observed in the absence of
321 *pCHARME*, we leveraged data from 223 eCLIP experiments covering 150 RNA-binding
322 proteins (RBP) across cell lines (K562 and HepG2)[57] that do not encode the lncRNA. By
323 focusing on ES, which represent the most altered and conserved *pCHARME*-regulated splicing
324 events, and comparing them to a control set of similarly expressed but unregulated junctions
325 (**Supplementary Fig. 5H**), we identified the splicing repressor PTBP1 as significantly
326 enriched at the 5′ junctions of the excluded exons (Fisher's exact test, FDR < 0.05, log2[Odds
327 Ratio] > 0) (**Fig. 5E** and **Supplementary Table 4**). Among them we found several known
328 PTBP1 splicing targets, such as *TPM1*, *ACTN1*, *FHOD3*, and *TNNT2*[58], expressed in CM and
329 displaying exon exclusion in *pCHARME* KO (**Supplementary Table 4**).
330 Overall, these data prioritise PTBP1 as a regulator of the *pCHARME*-dependent ES events and
331 suggest a possible interplay between the two in CM. To deepen this evidence with a focus into
332 the physical engagement of *pCHARME*, we analyzed its nucleotide composition (7-mers)
333 compared to the set of the 1987 lncRNAs expressed in CM (**Supplementary Fig. 5I**). We
334 found that *pCHARME* possesses a unique UC-rich sequence signature, primarily driven by an
335 extended simple repeat (75-2955 nt) in the 5' region of intron 1 (**Supplementary Table 4**).
336 RNA secondary structure predictions using the RNAfold software[59] suggested these UC-rich
337 stretches as unstructured and likely accessible to RBP (**Supplementary Fig. 5J**). On this
338 prediction, we employed four independent approaches to identify RBP with high affinity for

339 the UC-rich sequence, and specifically (i) RBP affinity scores derived from crosslinking-
340 induced truncations from eCLIP datasets, (ii) *in vitro* RNAcompete experiments assessing RBP
341 binding preferences for short nucleotide sequences[60], (iii) in silico predictions, and (iv) motif
342 analysis using CatRAPID software[61]. The results showed only two RBP (PTBP1 and PCBP2;
343 **Supplementary Fig. 5K** and **Supplementary Table 4)** as the most likely interactors of the
344 *pCHARME* UC-rich signature. Notably, PTBP1, previously found as a *pCharme* interactor in
345 murine myocytes[21], showed binding propensity scores closely mirroring the abundance of the
346 5' UC-rich sequences (**Fig. 5F**).
347 In support of the PTBP1/*pCHARME* interaction, we performed PTBP1-IF assay combined with
348 *pCHARME* RNA/FISH in CM, which revealed that approximately 40% of the lncRNA signal
349 overlapped with PTBP1 (**Fig. 5G**). Additionally, we tested for a direct interaction between
350 PTBP1 and the lncRNA by using UV cross-linking immunoprecipitation (CLIP) experiments
351 in CM. Western blot analysis confirmed successful recovery of PTBP1 after
352 immunoprecipitation (**Fig. 5H**, left panel). Subsequent RT-qPCR analysis of the retrieved
353 RNA using isoform-specific primers confirmed a robust and specific interaction between
354 PTBP1 and the 5′ region of *pCHARME* intron 1 (**Fig. 5H**, right panel), which aligns with the
355 presence of the UC-rich motifs and the highest PTBP1 binding propensity.
356 Together with the observation that PTBP1 binds *pCHARME* splicing-regulated targets in the
357 absence of the lncRNA, this PTBP1/*pCHARME* interaction in CM suggests a role for the
358 lncRNA as a decoy for PTBP1. To test this hypothesis, we examined the potential binding of
359 PTBP1 to the *pCHARME* splicing target *CACNB1*, whose exon 7 was excluded upon the
360 lncRNA ablation (**Supplementary Fig. 5C-D)**. Supporting the decoy mechanism, RT-qPCR
361 analysis from PTBP1-CLIP experiments performed in WT and KO conditions shows the
362 specific binding of PTBP1 at the level of *CACNB1* exon 7 in KO CM, while no enrichment
363 was found in WT cells (**Fig. 5I**). As PTBP1 was recently shown to negatively regulate CM
364 specification[62], the suggested role of *pCHARME* as a PTBP1 decoy could support a mechanism
365 ensuring the acquisition of splicing patterns that are specific for CM.
366
367 ***HSCHARME* dosage impacts the expression of disease-linked genes.**
368 The influence of *pCHARME* in CM differentiation encouraged further investigation into the
369 role of the human transcript in pathology. GSEA analyses performed on *HSCHARME* DEGs
370 (ΔP *vs* WT) evidenced a substantial enrichment of *HSCHARME* targets among genes
371 dysregulated in HCM (**Fig. 6A**, left panel) and DCM (**Fig. 6A**, right panel) cardiomyopathies.
372 Further confirmation was obtained by KEGG pathways enrichment analysis performed on the

373 same subsets, showing a significant clustering of *HSCHARME* targets into the HCM (adjusted
374 p-value $1.51E^{-10}$) and the DCM (adjusted p-value $3.86E^{-9}$) categories (**Fig. 6B**). These findings
375 strongly suggest that the human gene may play a role in the development or the homeostasis
376 of these cardiac pathological states. To strengthen this hypothesis, we performed DE analyses
377 of available RNA-seq datasets of HCM and DCM cardiomyopathies and Healthy human
378 cohorts (GSE130036[63]; GSE116250[64]). Interestingly, we found aberrant expression of the
379 *HSCHARME* locus with a significant up-regulation of the *pCHARME* isoform in both HCM
380 and DCM diseased hearts (**Fig. 6C**).
381 To more directly pinpoint the genes whose aberrant expression in pathology is influenced by
382 *pCHARME* alteration, we then used a guilt-by-association approach and compared the lists of
383 *HSCHARME* targets (**Supplementary Table 3**) to HCM and DCM DEGs (**Supplementary**
384 **Table 5**). We observed more than 1500 significant DEGs for each comparison, n=1598 for
385 HCM and n=1603 for DCM (**Fig. 6D**). Among them, the expression of 29.5% disease-
386 associated DEGs (category "A", **Fig. 6D**, left panel) was correlated with *pCHARME*, being up-
387 regulated in HCM and down-regulated in the lncRNA knockout cells. Intriguingly, these DEGs
388 clustered into categories of interest for heart functionality, such as actomyosin structure
389 organization, myofibril assembly, sarcomere organization and regulation of heart contraction.
390 Notably, the "A" category also contains the *CACNB1* transcript whose splicing pattern was
391 significantly altered upon *HSCHARME* ablation (**Supplementary Fig. 5C-D**). Similarly, the
392 comparison with DCM revealed that 30.3% of DEGs (category "a", **Fig. 6D**, right panel) were
393 up-regulated in the DCM cohorts and significantly down-regulated upon *HSCHARME*
394 ablation. Again, these genes clustered into categories functionally related to heart development,
395 such as sarcomere organization, myofibril assembly and cardiac muscle tissue development.
396 These results are in line with recent evidence demonstrating the central role of CM and the
397 contractile apparatus in DCM[65] and highlight the potential involvement of *pCHARME* in
398 disease.
399 On the other hand, the intersection between genes up-regulated upon *HSCHARME* ablation and
400 down-regulated in pathologies was limited to a few cases with little interest in CM development
401 (categories "B" and "b", **Fig. 6D**). The remaining DEGs (categories "C" and "c", **Fig. 6D**)
402 display expression trends that are opposite to those of *pCHARME*. Consequently, their changes
403 in pathology cannot be attributed to *pCHARME* upregulation but are likely influenced by other
404 context- or pathology-specific effects. The use of the same compound in the treatment of
405 cardiomyopathy can constitute a novel opportunity for therapeutic applications. In this
406 direction, the comparison of DEGs included in the categories "A" and "a" led to the discovery

407 of a common subset of cardiac homeostasis and contraction genes (n=184; **Fig. 6E**). Notably,
408 these transcripts, whose expression in CM was activated by *HSCHARME* and, coherently with
409 its pathological upregulation, aberrantly overexpressed in both HCM and DCM hearts. This
410 subset represents a group of candidates for developing therapeutic strategies. With the aim to
411 test their responsiveness to the modulation of *pCHARME* levels, we established an inducible
412 hiPSCs CRISPR-based system (RHE) acting as a dynamic rheostat to either silence or induce
413 *HSCHARME* within the same genetic background. CRISPR-cas9 mediated homologous
414 recombination was achieved by designing a single sgRNA (**Supplementary Table 2**) to add
415 the doxycycline (DOXY)-responsive element (TRE promoter) upstream of the *HSCHARME*
416 locus (**Supplementary Fig. 6A**). Assessment of *pCHARME* expression by RT-qPCR and
417 RNA-FISH analyses confirmed the success of the editing strategy, as the lncRNA levels were
418 comparable to knockout cells in the untreated CM and increased significantly upon
419 doxycycline addition (-/+ DOXY; **Fig. 7A**). To gain functional validation on the impact of
420 *pCHARME* dosage in RHE-CM, we then quantified the expression of several lncRNA targets
421 identified by our RNA-seq analysis (**Fig. 3C** and **Supplementary Fig. 3G**). In line with our
422 loss-of-function analyses, we focused on genes critical for CM structure (i.e. *MYH7*, *TNNT2*)
423 and functionality (*CACNA1C*, *CACNB1*). RT-qPCR analyses revealed that their expression
424 was consistently modulated by *pCHARME* levels, decreasing in -DOXY conditions while
425 triggered upon +DOXY treatments (**Supplementary Fig. 6B**). This effect was highly specific
426 as demonstrated by *TBX5* which levels were found unaltered, in line with its role as an upstream
427 regulator of the lncRNA expression (**Supplementary Fig. 6B**). Notably, these *pCHARME*-
428 responsive genes (i.e. *MYH7*, *TNNT2* and *CACNA1C*) are recognized as known
429 cardiomyopathy-causative loci[66,67] with essential roles in maintaining CM functionality and
430 homeostasis. Consequently, the ability to regulate their expression *via* modulation of
431 *pCHARME* levels offers a promising avenue to ameliorate the pathological state of patients
432 across various genetic backgrounds. Finally, we extended the analysis to our priority list of
433 targets common to "A" and "a", therefore consistently de-regulated in both DCM and HCM
434 (**Fig. 6E**). Intriguingly, we found that genes such as *PPFIA4*, *NPPB, CAMK2B* and the
435 cardiomyopathy-linked *MYO18B*[68], displayed significant and coherent responses to
436 *pCHARME* modulation (**Fig. 7B**). Finally, we also analysed the responsiveness of these
437 disease-linked genes to the transient knockdown of *pCHARME* in differentiated CM and found
438 the consistent decrease of all of them (**Fig. 7C**). Overall, the ability of *pCHARME* to influence
439 gene expression programs, including the regulation of cardiac genes linked to disease, marks

440 an advancement in the lncRNA field and suggests its potential relevance for future studies
441 exploring diagnostic or therapeutic strategies.
442
443 **DISCUSSION**
444 Understanding the molecular causes of cardiac pathologies represents a priority for human
445 health. However, the inability of the heart to regenerate and its mixed cellularity add
446 complexity to medical management, making cardiomyopathies the leading cause of death in
447 developed countries[69,70]. In this context, lncRNAs are emerging as promising candidates for
448 fine-tuning CM-specific gene expression programs due to their precise spatiotemporal
449 regulation[71]. New biopharmaceutical companies are starting to show the potential of lncRNA-
450 based platforms in enhancing precision medicine strategies for a wide range of illnesses
451 influenced by lncRNA levels, including cardiovascular diseases. This is the case of the
452 HAYA's therapeutic candidate HTX-001, an antisense oligonucleotide targeting the heart-
453 specific lncRNA Wisper[72].
454 Here, we provide evidence of the importance of *HSCHARME* in the pathophysiology of the
455 heart and set the stage for innovative approaches of lncRNA-targeting precision therapies for
456 the cell-specific treatment of human cardiomyopathies (**Fig. 8**). Indeed, the conservation across
457 species and the high specificity of *HSCHARME* expression for CM make this lncRNA an
458 excellent candidate for pioneering therapeutical treatments that could minimise off-target
459 effects on other cardiac cell types. Currently, several examples of lncRNA playing a role in
460 human CM homeostasis have been identified[5,6,10,73]. However, these studies rely on *in vitro*
461 systems that induce differentiation exclusively towards cardiomyocytes. Therefore, it is unclear
462 whether the role of these lncRNAs in cardiac processes is truly cardiomyocyte-specific or if
463 they could also have roles in other cardiac cell subtypes, such as cardiac fibroblast or
464 endothelial cells.
465 In our study, publicly available single-cell datasets allowed us to uncover the specific
466 expression of *HSCHARME* in human CM, providing a solid basis to investigate its role more
467 accurately and specifically within the cardiac context. We found that the human syntenic
468 *pCHARME* transcript shares a high level of conservation with its murine counterpart and
469 maintains peculiar features that make the lncRNA functional, such as the retention of intron 1.
470 Mechanistically, we provide evidence of the direct interaction between *pCHARME* and the
471 splicing suppressor PTBP1 and show their functional interplay in regulating a subset of the
472 lncRNA targets at the splicing level. Given the strong binding affinity between PTBP1 and
473 *pCHARME*, we propose a model in which the lncRNA acts as a decoy for PTBP1, sequestering

474 the splicing repressor from pre-mRNA targets, thereby promoting the inclusion of specific
475 exons and ensuring the acquisition of CM-specific splicing patterns. This model aligns with
476 the recently described role of PTBP1 as a negative regulator of CM specification[62]. As such,
477 the interaction between *pCHARME* and PTBP1 underscores the importance of this axis in
478 maintaining proper cardiac splicing programs and highlights its potential relevance in cardiac
479 pathophysiology. Recently, nuclear speckles have been described as storage sites for transcripts
480 with retained introns[74]. These unspliced RNAs retain certain structural features, such as a high
481 GC content and a high density of protein binding motifs[75] that perfectly fit with *pCHARME*
482 and strengthen our proposed model of speckle-associated lncRNA.
483 As the role of nuclear speckles in regulating cellular specification is still only partially
484 understood, the muscle-specific restriction of *pCHARME* can add specificity to their
485 implication in cardiac-specific gene expression programs. Remarkably, as for *HSCHARME*
486 (this paper), alterations in SC35 expression were linked to dilated cardiomyopathy[76], thereby
487 narrowing a possible line of intervention for cardiac disorders. To note, the role of *pCHARME*
488 in regulating structural genes, such as different types of myosins and troponins, is also in line
489 with a broader role for the human lncRNA in the functionality of muscle cells, either cardiac
490 or striated, as previously characterized in mice[20,21]. A newly identified shorter isoform of
491 *pCHARME* was identified in human myocytes as a regulator of *MYBPC2* expression
492 (lncFAM[23]). However, the fact that *MYBPC2* is neither expressed nor influenced by
493 *HSCHARME* in CM (data not shown) would suggest a different role for the two transcripts in
494 the heart.
495 We also provide evidence that modulating *pCHARME* expression in differentiated CM can
496 influence the levels of several disease-linked genes involved in CM physiological processes,
497 such as calcium handling, heart contraction and CM development. Given its role as a positive
498 regulator of CM differentiation, the modulation of *pCHARME* expression levels might be
499 beneficial for guiding CM de-differentiation into a progenitor-like state, which is crucial for
500 their replenishment after injury and, ultimately, for the effective regeneration of the cardiac
501 muscle *in vivo*[77–79]. Lastly, the proven association between *pCHARME* dysregulation and
502 cardiac pathologies, indicates the necessity of including the lncRNA in genetic testing,
503 particularly for cardiomyopathies with no clear genetic evidence. Alternatively, assessment of
504 *HSCHARME* expression levels could assist in identifying differences in outcome prediction
505 when stratifying cohorts of cardiomyopathy patients who have tested positive for genetic
506 markers, based on phenotype manifestation.
507

508 **MATERIALS AND METHODS**

509 **Cell culture conditions and transfection**

510 Human Episomal iPSCc (WT from [80] and WT2 are the WTSIi004-A cell line from EBiSC;
511 https://ebisc.org/WTSIi004-A) were maintained in Essential 8 media (Life Technologies) in
512 plates coated with Geltrex LDEV-free reduced growth factor basement membrane matrix (Life
513 Technologies). Differentiation into cardiomyocytes was performed as described in [32].
514 Specifically, cells were grown as monolayers in RPMI 1640 Medium (Life Technologies)
515 supplemented with B27 Supplement without Insulin (Life Technologies) for 10 days and then
516 switched to RPMI 1640 Medium (Life Technologies) supplemented with B27 Supplement with
517 Insulin (Life Technologies). The induction towards the cardiac lineage was performed by
518 upregulating canonical Wnt signalling (adding CHIR-99021; Selleck) in the early stages of
519 differentiation and by its subsequent inhibition after 3 days (adding IWR-1; Sigma). Genome
520 editing analyses on hiPSCs were performed by co-transfecting the PX333 plasmid, harbouring
521 the cloned sgRNAs (designed with https://chopchop.cbu.uib.no/; sequences available upon
522 request) and the CAS9 sequence, together with the HR 110PA-1 donor vector (System
523 Biosciences) in which two homology arm sequences were cloned. Cloning was performed
524 employing In-fusion® kit (Takara), according to the manufacturer's protocol. Cloning with
525 this approach requires adding 15 nt-long tails to the oligonucleotides to create a homology
526 region among insert and plasmid, allowing recombination into bacteria cells (E.Coli STBL3®).
527 The transfection was performed using the Neon Transfection System (Life Technologies).
528 After 48 hours, cells were selected using 1 µg/ml puromycin to identify heterozygous and
529 homozygous *HSCHARME* KO or *HSCHARME* RHE clones. For RNAi, WT hiPSC-derived
530 CM were dissociated into single cells and transfected with either control (GAP-SCR) or
531 *HSCHARME* (GAP-*pCHARME*) antisense LNA GapmeRs (sequences available upon request)
532 using TransIT-X2® Dynamic Delivery System (Mirus). Total RNA was collected 48 hr after
533 transfection.

534

535 **RNA extraction and RT-qPCR analysis**

536 Total RNA from cultured cells and tissues was isolated using TRI Reagent (Zymo Research),
537 extracted with Direct-zol™ RNA MiniPrep (Zymo Research), treated with DNase (Zymo
538 Research), retrotranscribed using PrimeScript Reagent Kit (Takara) and amplified by RT-
539 qPCR using PowerUp SYBR-Green MasterMix (Life Technologies), as described in [21] (see
540 **Supplementary Table 6**). Data are represented as relative expression ($2^{\wedge -DCt}$).

541 Nucleus/Cytoplasm fractionation of hiPSC-derived CM was performed using the PARIS kit
542 (Life Technologies), following the manufacturer's specifications.
543
544 **RNA-Sequencing Analysis**
545 Total RNA was collected from hiPSCs-derived CM (day 10 and day 20) obtained from three
546 independent biological replicates. Illumina Stranded mRNA Prep was used to prepare cDNA
547 libraries. RNA-Seq was performed on an Illumina Novaseq 6000 Sequencing system at the
548 Genomic Facility of the Center for Human Technologies of IIT.
549 Q30-filtered raw reads were obtained from Illumina BaseSpace Reads and aligned to GRCh38
550 assembly using STAR aligner software[81]. Gene loci fragment quantification was performed on
551 Ensemble (release 87) gene annotation gtf using STAR –quantMode GeneCounts parameter.
552 The gtf file was edited adding the *HSCHARME* genomic coordinates (**Supplementary Table**
553 **3**). Read counts of "reverse" configuration files were combined into a count matrix file, that
554 was given as input to DESeq2 R package[82] for normalization and differential expression
555 analysis, after removing genes with less than 10 counts in at least two samples. Adjusted p-
556 value cutoff for selecting significant differentially expressed genes was set to 0.05 unless
557 otherwise specified. Heatmap of differentially expressed genes was generated using pheatmap
558 R package[83] from normalized scaled data.
559 Volcano plots were generated using *Enhanced Volcano* R package
560 (https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/Enhanced
561 Volcano.html). Gene Ontology enrichment analyses were performed with the EnrichR online
562 tool (https://maayanlab.cloud/Enrichr/[84–86]) using standard settings. Gene Set Enrichment
563 Analysis was performed with the GSEA tool from Broad Institute[87,88] running a pre-ranked
564 analysis with statistically significant genes ordered by log2 fold change (KO *vs* WT D20). Gene
565 Ontology was run using the complete GO annotation file (c5.all.v2023.2.Hs.symbols.gmt).
566 Human Phenotype Ontology (HPO) was run using the HPO annotation file
567 (c5.hpo.v2023.2.Hs.symbols.gmt). Figures were prepared with R and Prism 7.0 software and
568 edited with Adobe Illustrator 2024.
569
570 **RNA Isoforms Analysis**
571 Reads derived from the WT RNA-seq datasets were used as input for the Salmon tool[89] to
572 identify and quantify the *HSCHARME* isoforms. A transcriptome file was generated using the
573 human genome FASTA file and the GTF file containing *HSCHARME* coordinates. Transcripts
574 per million (TPM) output values were used for quantification and statistical analysis. Plot for

575 quantifying the *HSCHARME* isoforms was generated with the *ggtranscript* R package
576 (https://github.com/dzhang32/ggtranscript), using the genomic coordinates provided within the
577 GTF annotation file derived from Salmon alignment. Figures were edited with Adobe
578 Illustrator 2024. For large-scale isoform analysis from WT and ΔP datasets, transcript
579 quantification was performed with the Salmon tool using standard parameters and FASTQ files
580 as input. Transcripts per million (TPM) counts were used to quantify the expression of each
581 transcript isoform. Differential gene expression analysis was performed with edgeR[90], using a
582 quasi-likelihood negative binomial generalized log-linear model (glmQLFit function) and
583 applying a false discovery rate (FDR) cutoff of 0.05 to identify statistically significant genes.
584
585 **Single Cell RNA-seq deconvolution analysis**
586 Single-cell RNA sequencing (scRNA-seq) data of hiPSC-derived CM (14 and 45 days of
587 differentiation) were obtained from the Synapse database (ID: syn7818379). FASTQ files were
588 extracted from the provided BAM files using CellRanger BamtoFastq (v1.4) and aligned to the
589 GRCh38 reference genome using Ensembl 99 gene annotation, supplemented with a custom
590 annotation for the long non-coding RNA *HSCHARME*. Cells were filtered to retain high-
591 quality single-cell transcriptomes, defined by having more than 200 and fewer than 5,000
592 detected genes and less than 20% mitochondrial gene content. The upper threshold for
593 mitochondrial gene content was extended beyond standard filters to retain metabolically active
594 cell types such as CM. Reanalysis was performed with Seurat v5.2.1 in R[91] the data have been
595 first normalized using NormalizeData function with "LogNormalize" method and then
596 integrated via Seurat's Canonical Correlation Analysis (CCA) standard workflow.
597 After integration, initial unsupervised clustering was performed using the Seurat pipeline;
598 Principal component analysis (PCA) was applied on the 3000 most variable genes identified
599 using the VST method, and the top 15 principal components were used to construct a shared
600 nearest neighbour (SNN) graph. Clusters were identified using the Louvain algorithm with a
601 resolution parameter of 0.5. Cell population identities were assigned using a knowledge-based
602 approach by evaluating canonical marker gene expression across clusters. CM were identified
603 based on the expression of known cardiac muscle genes[92]. Neuronal populations were
604 annotated using gene sets associated with the Gene Ontology (GO) terms for postsynaptic
605 density (GO:0014069) and axon part (GO:0033267). Endothelial populations were annotated
606 based on expression of established endothelial markers[62,93,94]. Differential gene expression
607 analysis was performed using the RunPresto function from the Presto package v1.0 in R,
608 applying the Wilcoxon rank-sum test. Each cell population was compared against all remaining

609 cells in a one-vs-all manner. Genes with log2 fold change > 1 and adjusted p-value < 0.05 were
610 considered significantly upregulated. Identified marker genes for each cluster were subjected
611 to functional enrichment analysis using the WebGestaltR R package v0.4.6. The reference gene
612 background used for enrichment consisted of all genes detected in the scRNA-seq experiment
613 after quality control and filtering.
614 Deconvolution analysis was conducted using the MusiC R package v1.0[95]. First, the gene
615 counts table from the bulk RNA-seq experiment was normalized and converted to counts per
616 million (CPM). The gene expression profile matrix for each cell population, as defined by our
617 single-cell analysis, was then used as input to estimate the proportion of each cell population
618 within each bulk RNA-seq sample. To quantify the relative contribution of each differentially
619 expressed genes upon *HSCHARME* KO to individual cell transcriptomes, and to cross-validate
620 deconvolution analysis results, "KL-GSEA" was performed. We calculated per-cell Kullback-
621 Leibler divergence scores comparing transcriptome expression profiles and a target distribution
622 restricted to either genes downregulated (putative cardiac developmental genes) or upregulated
623 in *HSCHARME* KO (bulk RNA-seq). KL scores for each cell were visualized using UMAP.
624
625 **Alternative Splicing Analysis**
626 Cutadapt v3.2[96] with parameters: *-u 1 -U 1 –trim-n* and Trimmomatic v0.39[97] with parameters:
627 *-PE ILLUMINACLIP:adapter_path:2:30:10:8:true LEADING:3 TRAILING:3*
628 *SLIDINGWINDOW:4:20* were used to remove adapter sequences and poor quality bases; for
629 both software the minimum read length after trimming was set to 35. Then, STAR software
630 v2.7.7a[81] was used to align reads to GRCh38 genome using the following parameters: --
631 *outSAMstrandField intronMotif --outSAMattrIHstart 0 --outFilterType BySJout --*
632 *outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --*
633 *outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 --*
634 *alignIntronMax 1000000 --alignMatesGapMax 1000000 --outFilterIntronMotifs*
635 *RemoveNoncanonical --peOverlapNbasesMin 3.* Alternative splicing events were identified
636 using rMATS turbo version 4.1.1[48], with the following parameters: *–variable-read-length, –*
637 *allow-clipping, –readLength 150*, and *–libType fr-firststrand*. This analysis utilized the
638 Ensembl gene annotation file (release 86) as the reference for splicing events. For downstream
639 analyses, JCEC output from rMATS was selected. Splicing events were filtered based on false
640 discovery rate (FDR) and Inclusion Level criteria. Only events with an FDR value of <0.05
641 and an Inclusion Level difference >0.1 or < -0.1 were considered significant and included for
642 further examination (**Supplementary Table 3**). The rmats2sashimiplot Python script

643 (available at https://github.com/Xinglab/rmats2sashimiplot) was used to visualize the selected
644 alternative splicing events. This tool provided a graphical representation of splicing patterns,
645 aiding in the interpretation of differential splicing events. Alternative splicing analysis in heart
646 tissue comparing WT and *Charme* KO mice was performed following the same approach, using
647 RNA-seq data available from the GEO dataset GSE200878[22]. To assess alternative splicing
648 events conservation, the coordinates of exons regulated by *Charme* (either included or skipped)
649 in mouse heart tissue were extracted and converted from mm10 genome to the hg38 human
650 genome using the LiftOver tool[98]. The fraction of successfully lifted-over mouse-regulated
651 exons that overlapped perfectly with regulated exons in human CM was then calculated,
652 ensuring the consistency of the observed regulation between mouse and human datasets. To
653 estimate the expected overlap by chance, 1,000 random samples of human exons were picked
654 up from the ones detected by rMATS software (including not *pCHARME*-regulated). Then the
655 overlap between human random exons and lifted-over mouse-regulated exons, generating a
656 distribution of expected values.
657
658 ***pCHARME* sequence analysis**
659 For the analysis of the k-mer composition, *pCHARME* was analyzed together with 1,987 genes
660 annotated as "lncRNA" and expressed in WT CM (D20), each with an average FPKMs greater
661 than 1. Gene annotations and transcript sequences were retrieved from Ensembl release 99[99]
662 using the biomaRt R package[100]. When multiple isoforms were associated with a gene, the
663 longest isoform was selected to reduce sequence redundancy. The frequency of all possible
664 combinations of seven nucleotides (7-mers) was computed for each long non-coding RNA. The
665 *pCHARME* 7-mer signature was defined as the set of 7-mers exhibiting the highest frequency
666 in *pCHARME* compared to other lncRNAs (top 0.1%). The sequence logo of the *pCHARME*
667 7-mers was generated using the Logomaker Python package v0.8[101].
668 The RNA secondary structure of *pCHARME* was predicted using the RNAfold algorithm[59]
669 with standard parameters. Repeated and low complexity sequences were identified by scanning
670 the transcript sequence with RepeatMasker v4.1.1[102], using default settings and specifying the
671 appropriate organism (Homo sapiens), with rmblastn v2.11.0+ as the internal search engine.
672 The RepeatMasker output was subsequently converted to BED format using the RM2Bed.py
673 utility provided by the software. To integrate the various analyses and visualize the results
674 along the *pCHARME* sequence, the lncRNA was segmented from 5' to 3' into windows of 50
675 nucleotides in length, with sliding step of 10 nucleotides. The resulting bins were compiled
676 into a BED file. BEDtools intersect v2.29.1[103] was used to assign repeated elements to

677 overlapping bins, while getfasta was applied to retrieve bin sequences, from which GC content
678 and the occurrence of the *pCHARME* 7-mer signature were calculated. The Accessibility score
679 was computed for each bin using the dot-bracket notation of the previously predicted
680 *pCHARME* MFE structure, by calculating the fraction of unpaired nucleotides within each
681 window. Finally, the ComplexHeatmap R package was used to visualize all results in a unified
682 representation[104].
683
684 **Analysis of Protein-RNA interactions**
685 To identify RNA-binding proteins with high affinity for the *HSCHARME* transcript, we applied
686 four complementary approaches: i. eCLIP-based motif enrichment analysis. We leveraged
687 binding affinity scores of 150 RBPs for all possible 5-mers derived from the analysis of
688 crosslinking induced truncations (CITS)[105] in 223 eCLIP experiments[57]. For each experiment,
689 the RBPs binding motifs were defined as the top 5 5-mers with highest PEKA scores from
690 eCLIP datasets processed with Clippy or narrowPeaks softwares, or mCROSS[105]. To assess
691 RBP affinity for *pCHARME*, we decomposed the *pCHARME*-specific 7-mers into 5-mers and
692 selected RBPs for which the selected 5-mers matched the RBPs binding motifs in at least two
693 analysis methods; ii. *in vitro* RNAcompete-derived binding profiles. We retrieved RBPs
694 affinity scores (Z-scores) for all possible 7-mers from the RNAcompete dataset available in the
695 CisBP-RNA database[60]. RBPs with *pCHARME* 7-mers ranked among their top binders (Z-
696 score > 2) were selected as high-affinity candidates; iii. *in silico* sequence-based binding
697 prediction with catRAPID. We used catRAPID software[61] to predict interactions between full-
698 length *pCHARME* and a reference library of RBPs. Proteins with a predicted interaction Z-
699 score > 2 were considered high-affinity binders; iv. motif enrichment using catRAPID
700 signature. We further identified candidate RBPs by evaluating the enrichment of known
701 binding motifs within the *pCHARME* sequence using the motif analysis module of catRAPID.
702 RBP binding enrichment analysis to pre-mRNA regions proximal to splicing events regulated
703 by *pCHARME* was performed by focusing on exon inclusion and exclusion events upon its
704 KO, as identified using the rMATS software. To map RNA-binding protein (RBP) binding
705 sites, we used IDR peaks from the eCLIP datasets previously described[57]. To ensure
706 compatibility with the eCLIP data, we first filtered the rMATS output from the *HSCHARME*
707 KO versus WT comparison in CM, retaining only those genes also expressed in HepG2 and
708 K562 cell lines (average FPKM > 1), which are the cell lines where eCLIP were performed.
709 Next, for each exon significantly affected by *HSCHARME* loss (differentially spliced), we
710 selected a matched control exon not affected by the lncRNA KO. Control exons were chosen

711 with splicing junctions closely match the regulated exons in terms of splice junction counts
712 (SJC) and intron junction counts (IJC), ensuring comparable expression levels. For each
713 differentially spliced and control exon, we considered the following regions: the 3′ splice site
714 (SS) of the exon upstream to the regulated one (defined as "upstream"); both the 5′ and 3′ SS
715 of the regulated exon (defined as "center"), and the 5′ SS of the exon downstream to the
716 regulated one (defined as "downstream"). For each site, 250 nucleotides of intronic sequence
717 adjacent to the splice junction and 50 nucleotides of exonic sequence were analyzed, generating
718 a total of 8 pre-mRNA region types: exonic-upstream, intronic-upstream, intronic-
719 center(5'SS), exonic-center(5'SS), exonic-center(3'SS), intronic-center(3'SS), intronic-
720 downstream and exonic-downstream.
721 Subsequently, BEDtools intersect was employed with the -s parameter to identify strand-
722 specific overlaps between RBP binding sites, derived from each eCLIP experiment, and the
723 pre-mRNA regions flanking dysregulated exons and their matched controls. For each eCLIP
724 dataset, the number of regions either bound or unbound by the RBP was quantified separately
725 for dysregulated and control exons, allowing the construction of a 2×2 contingency table to
726 evaluate differential binding patterns. Then, Fisher's exact test was applied to assess whether
727 RBP binding was significantly enriched or depleted in the regions associated with dysregulated
728 exons compared to controls. Finally, Benjamini-Hochberg false discovery rate (FDR)
729 correction was used to adjust p-values for multiple hypothesis testing.
730
731 **Pathological dataset analysis**
732 Pathological datasets raw data (FASTQ files) were retrieved from Gene Expression Omnibus
733 (GEO) database with the following accession IDs: GSE130036 for HCM and GSE116250 for
734 DCM. Data were pre-processed and analyzed as described in the RNA-Sequencing analysis
735 section (**Supplementary Table 5**). Intersections between *HSCHARME* KO and the
736 pathological datasets were performed using a 0.05 adjusted p-value cutoff for all genes in all
737 the analyzed datasets and reporting the log2 fold change values for each gene. Figures were
738 prepared with Prism 7.0 software and edited with Adobe Illustrator 2024.
739
740 **Flow cytometry assay**
741 For PDGFRA, CD56 and CD82 cell surface markers, CM were washed twice with DMEM F12
742 and dissociated as single cells using 0.25% trypsin-EDTA. Cells were stained with the diluted
743 (1:10) primary antibody (CD82, see **Supplementary Table 6**) for 45 minutes in Flow
744 Cytometry Buffer (FCB; 5% FBS in PBS). After two washes in FCB, cells were incubated for

745 30 min at RT with a secondary antibody (Alexa Fluor 488, Life Technologies) at 1:300 dilution
746 or with a conjugated antibody (APC-PDGFRA 1:20 and 488-CD56 1:20) in FCB. After two
747 washes, stained cells were analyzed by CytoFLEX SRT Cell Sorter (Beckman Coulter). Data
748 were collected from at least 10.000 events and analyzed using Kaluza software (Beckman
749 Coulter). For TNNT2, were fixed in 4% paraformaldehyde (PFA) in PBS for 15 minutes at
750 room temperature (RT) before staining. After fixation, cells were washed twice by being
751 resuspended in Flow Cytometry Buffer (FCB; 5% FBS in PBS), centrifuged for 5 min at 250
752 g and the supernatant was discarded. A third wash with FCB-0.2% Triton X-100 was performed
753 as described above. Cells were stained with the diluted (1:100) primary antibody (TNNT2, see
754 **Supplementary Table 5**) in FCB-0.2% Triton X-100 for 45 min at RT. After two washes in
755 FCB, cells were incubated for 30 min at RT with a secondary antibody (Alexa Fluor 488, Life
756 Technologies) at 1:400 dilution, washed twice more and resuspended in PBS. Stained cells
757 were analyzed and sorted on a flow cytometer (BD Bioscience, Milan, Italy). Data was
758 collected from at least 10,000 events. Flow cytometry data were analyzed with Flowing
759 software 2.5.1. After exclusion of debris and doublets based on light scatter properties, single
760 cells were analyzed for the expression of TNNT2 and reported as the percentage of marker-
761 positive cells or as mean fluorescence intensity.
762
763 **Immunofluorescence and RNA-FISH assays**
764 Glass coverslips were prepared as for [106] by treatment with poly-L-lysine (50 mg/ml in water)
765 for 30 minutes under UV light and overnight treatment with Laminin (2 mg/ml in water) at 37
766 °C. After 20 days of differentiation, hiPSC-derived CM were washed twice with DMEM F12
767 and dissociated as single cells using 0.25% trypsin-EDTA. Isolated CM were re-plated on the
768 treated coverslips in RPMI 1640 Medium (Life Technologies) supplemented with B27
769 Supplement and Rock inhibitor (1:1000). After 24 hours, cells were fixed in 4%
770 paraformaldehyde (PFA) in PBS for 20 minutes at 4°C prior staining. For immunofluorescence,
771 cells were permeabilized with 0.2% Triton X-100 in PBS for 20 minutes at room temperature
772 (RT) and blocked with 1% Goat Serum (GS) in PBS for 20 minutes at RT. The primary
773 antibody (TNNT2, see **Supplementary Table 6**) was diluted (1:200) in 1% GS-PBS and
774 incubated overnight at 4°C. The cells were rinsed with PBS and incubated for 1 hour with a
775 secondary antibody (Alexa Fluor 488, Life Technologies) at 1:300 dilution, rinsed,
776 counterstained with DAPI for 5 minutes at RT and mounted on cover-glasses.
777 For RNA-FISH experiment, staining was carried out with the HCR RNA-FISH technology
778 according to manufacturer protocol (https://www.molecularinstruments.com/hcr-rnafish), with

779 a specific set-up thought for lncRNA staining. Images were acquired on Carl Zeiss Axio
780 Vert.A1 Microscope or with an inverted confocal Olympus IX73 microscope equipped with a
781 Crestoptics X-LIGHT V3 spinning disk system and a Prime BSI Express Scientific CMOS
782 camera. The acquisitions were obtained using a 20X air objective (IF-only samples) or
783 UPLANSApo 60X (NA 1.35) oil objective (FISH/IF samples) and collected with the
784 MetaMorph software (Molecular Devices). SC35 domains/*pCHARME* and PTBP1/*pCHARME*
785 overlapping quantification was performed by ImageJ tolls in JACoP package in order to obtain
786 a Mander's overlap coefficient (MOC) from single ROIs of original confocal images.
787
788 **Cross-linking immunoprecipitation (CLIP) assay.**
789 hiPSC-CM (D20) were washed twice with PBS and UV-crosslinked (4000 μJ) in a
790 Spectrolinker UV Crosslinker (Spectronics Corporation). Cells were harvested in lysis buffer
791 (20 mM Tris-HCl, 100 mM NaCl, 0.5 mM EDTA, 0.5% [v/v] NP40, 0.1% SDS, 0.5 mM DTT,
792 1× PIC), incubated on ice for 15 min and sonicated at low-intensity four times with Bioruptor
793 Plus sonication device to ensure membrane lysis. Lysate was diluted to a final concentration of
794 1 mg/ml and PTBP1 CLIP was carried out as in [107]
795
796 **Morphological analysis**
797 For morphological analysis, 10 cells for each field were individually selected and analysed
798 using ImageJ standard tools. 5 fields were considered for each biological replicate. Cell area
799 and circularity were computed with the ImageJ measure tool with circularity=1 indicating a
800 perfect circle. Elongation represents the ratio of the major axis length to the minor axis length,
801 as described in [42]. For each parameter, the average of each biological replicate was represented
802 as a dot in the plots. Statistical analyses were performed on the overall average measure of the
803 5 biological replicates.
804
805 **Cardiomyocytes contraction analysis**
806 Spontaneous CM contraction was assessed after 20 days of differentiation. To quantify beat
807 rate and contraction dynamics, light brightfield videos of WT and ΔP beating monolayers were
808 recorded with an OkoLab stage top incubator equipped with 20X air objectives under
809 controlled CO_2, pH and temperature conditions. One video per well was taken across n=5
810 biological replicates. Video recordings were processed using a MuscleMotion macro[108] in
811 FIJI/ImageJ software (v2.9.0). MUSCLEMOTION outputs included contraction duration, time
812 to peak and peak-to-peak time.

Statistical methods and rigour

For RT-qPCR and flow cytometry experiments, statistical analyses were performed on log_2-transformed fold change (LogFC) values to stabilise variance and approximate a normal distribution using a two-tailed one-sample t-test against the null hypothesis of zero (no change). For experiments with n=5, data were assessed for normal distribution with a Shapiro-wilk test, and a parametric/non-parametric statistical test was used accordingly (Student's t-test when data are normally distributed and Wilcoxon-Mann-Whitney when they are not). Details regarding statistical tests, p-values, and sample sizes (n) are provided in the corresponding figure legends.

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. Bergmann, O. *et al.* Evidence for cardiomyocyte renewal in humans. *Science (1979)* **324**, 98–102 (2009).
3. Eschenhagen, T. *et al.* Cardiomyocyte Regeneration: A Consensus Statement. *Circulation* **136**, 680 (2017).
4. 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).
5. 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).
6. 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).
7. 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).
8. 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).
9. 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).
10. Ponnusamy, M. *et al.* Long Noncoding RNA CPR (Cardiomyocyte Proliferation Regulator) Regulates Cardiomyocyte Proliferation and Cardiac Repair. *Circulation* **139**, 2668–2684 (2019).
11. Gao, R. *et al.* Long Noncoding RNA Cardiac Physiological Hypertrophy-Associated Regulator Induces Cardiac Physiological Hypertrophy and Promotes Functional Recovery after Myocardial Ischemia-Reperfusion Injury. *Circulation* **144**, 303–317 (2021).
12. Anderson, D. M. *et al.* A myocardin-adjacent lncRNA balances SRF-dependent gene transcription in the heart. *Genes Dev* **38**, 1–6 (2021).
13. Rogala, S. *et al.* The lncRNA Sweetheart regulates compensatory cardiac hypertrophy after myocardial injury in murine males. *Nature Communications 2023 14:1* **14**, 1–14 (2023).
14. He, X. *et al.* The long noncoding RNA CARDINAL attenuates cardiac hypertrophy by modulating protein translation. *J Clin Invest* **134**, (2024).
15. Li, X. *et al.* Loss of AZIN2 splice variant facilitates endogenous cardiac regeneration. *Cardiovasc Res* **114**, 1642–1655 (2018).
16. Chen, Y. *et al.* Long Non-coding RNA ECRAR Triggers Post-natal Myocardial Regeneration by Activating ERK1/2 Signaling. *Molecular Therapy* **27**, 29–45 (2019).
17. 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).
18. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39**, 925–938 (2010).
19. Clemson, C. M. *et al.* An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol Cell* **33**, 717–726 (2009).
20. Ballarino, M. *et al.* Deficiency in the nuclear long noncoding RNA Charme causes myogenic defects and heart remodeling in mice. *EMBO J* **37**, (2018).

21. Desideri, F. *et al.* Intronic Determinants Coordinate Charme lncRNA Nuclear Activity through the Interaction with MATR3 and PTBP1. *Cell Rep* **33**, (2020).
22. Taliani, V. *et al.* The long noncoding RNA Charme supervises cardiomyocyte maturation by controlling cell differentiation programs in the developing heart. *Elife* **12**, (2023).
23. Chang, M. W. *et al.* Enhanced myogenesis through lncFAM-mediated recruitment of HNRNPL to the MYBPC2 promoter. *Nucleic Acids Res* **50**, 13026–13044 (2022).
24. Mcleod, J. C. *et al.* Network-based modelling reveals cell-type enriched patterns of non-coding RNA regulation during human skeletal muscle remodelling. *NAR Molecular Medicine* **1**, (2024).
25. Chen, J. *et al.* Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol* **17**, 1–17 (2016).
26. Lewandowski, J. P. *et al.* The Tug1 lncRNA locus is essential for male fertility. *Genome Biol* **21**, 1–35 (2020).
27. Zhang, J. *et al.* An integrative ENCODE resource for cancer genomics. *Nature Communications 2020 11:1* **11**, 1–11 (2020).
28. Noguchi, S. *et al.* FANTOM5 CAGE profiles of human and mouse samples. *Sci Data* **4**, (2017).
29. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, (2020).
30. Jones, R. C. *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science (1979)* **376**, (2022).
31. 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).
32. Nakahama, H. & Di Pasquale, E. Generation of cardiomyocytes from pluripotent stem cells. *Methods in Molecular Biology* **1353**, 181–190 (2016).
33. 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).
34. Makarewich, C. A. *et al.* MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β-Oxidation. *Cell Rep* **23**, 3701–3709 (2018).
35. Stein, C. S. *et al.* Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Rep* **23**, 3710-3720.e8 (2018).
36. Yan, Y. *et al.* The cardiac translational landscape reveals that micropeptides are new players involved in cardiomyocyte hypertrophy. *Molecular Therapy* **29**, 2253–2267 (2021).
37. 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).
38. 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).
39. 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).
40. 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).
41. 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).

928 42. Sutcliffe, M. D. *et al.* High content analysis identifies unique morphological features of
929 reprogrammed cardiomyocytes. *Sci Rep* **8**, (2018).
930 43. Takeda, M. *et al.* Identification of Cardiomyocyte-Fated Progenitors from Human-
931 Induced Pluripotent Stem Cells Marked with CD82. *Cell Rep* **22**, 546–556 (2018).
932 44. Landim-Vieira, M. *et al.* Cardiac troponin T N-domain variant destabilizes the actin
933 interface resulting in disturbed myofilament function. *Proc Natl Acad Sci U S A* **120**,
934 (2023).
935 45. Montañés-Agudo, P., Pinto, Y. M. & Creemers, E. E. Splicing factors in the heart:
936 Uncovering shared and unique targets. *J Mol Cell Cardiol* **179**, 72–79 (2023).
937 46. Spector, D. L. & Lamond, A. I. Nuclear Speckles. *Cold Spring Harb Perspect Biol* **3**,
938 a000646 (2011).
939 47. Faber, G. P., Nadav-Eliyahu, S. & Shav-Tal, Y. Nuclear speckles – a driving force in
940 gene expression. *J Cell Sci* **135**, (2022).
941 48. Wang, Y. *et al.* rMATS-turbo: an efficient and flexible computational tool for alternative
942 splicing analysis of large-scale RNA-seq data. *Nat Protoc* **19**, 1083–1104 (2024).
943 49. Dolphin, A. C. Voltage-gated calcium channels and their auxiliary subunits: physiology
944 and pathophysiology and pharmacology. *J Physiol* **594**, 5369–5390 (2016).
945 50. England, J. & Loughna, S. Heavy and light roles: myosin in the morphogenesis of the
946 heart. *Cell Mol Life Sci* **70**, 1221–1239 (2013).
947 51. Olivieri, J. E. *et al.* RNA splicing programs define tissue compartments and cell types
948 at single-cell resolution. *Elife* **10**, e70692 (2021).
949 52. Sun, Y. Le, Li, S. H., Yang, L. & Wang, Y. miR-376b-3p attenuates mitochondrial
950 fission and cardiac hypertrophy by targeting mitochondrial fission factor. *Clin Exp*
951 *Pharmacol Physiol* **45**, 779–787 (2018).
952 53. Fan, L. L. *et al.* Haploinsufficiency of syncoilin leads to hypertrophic cardiomyopathy.
953 *Genes Dis* **9**, 1412–1414 (2022).
954 54. Timmer, L. T. *et al.* Cardiomyocyte SORBS2 expression increases in heart failure and
955 regulates integrin interactions and extracellular matrix composition. *Cardiovasc Res* **00**,
956 1–16 (2025).
957 55. Schulte, J. S. *et al.* Cardiac expression of the CREM repressor isoform CREM-IbΔC-X
958 in mice leads to arrhythmogenic alterations in ventricular cardiomyocytes. *Basic Res*
959 *Cardiol* **111**, 1–12 (2016).
960 56. Shen, J., Shentu, J., Zhong, C., Huang, Q. & Duan, S. RNA splicing factor RBFOX2 is
961 a key factor in the progression of cancer and cardiomyopathy. *Clin Transl Med* **14**,
962 (2024).
963 57. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-
964 binding proteins. *Nature 2020 583:7818* **583**, 711–719 (2020).
965 58. Charlet-B, N., Logan, P., Singh, G. & Cooper, T. A. Dynamic antagonism between ETR-
966 3 and PTB regulates cell type-specific alternative splicing. *Mol Cell* **9**, 649–658 (2002).
967 59. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 1–14
968 (2011).
969 60. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation.
970 *Nature 2013 499:7457* **499**, 172–177 (2013).
971 61. Armaos, A., Colantoni, A., Proietti, G., Rupert, J. & Tartaglia, G. G. catRAPID omics
972 v2.0: going deeper and wider in the prediction of protein–RNA interactions. *Nucleic*
973 *Acids Res* **49**, W72–W79 (2021).
974 62. Liu, Z. *et al.* Single-cell transcriptomics reconstructs fate conversion from fibroblast to
975 cardiomyocyte. *Nature 2017 551:7678* **551**, 100–104 (2017).
976 63. Liu, X. *et al.* Long non-coding and coding RNA profiling using strand-specific RNA-
977 seq in human hypertrophic cardiomyopathy. *Scientific Data 2019 6:1* **6**, 1–7 (2019).

64. 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).
65. Jurgens, S. J. *et al.* Genome-wide association study reveals mechanisms underlying dilated cardiomyopathy and myocardial resilience. *Nature Genetics 2024 56:12* **56**, 2636–2645 (2024).
66. Garfinkel, A. C., Seidman, J. G. & Seidman, C. E. Genetic Pathogenesis of Hypertrophic and Dilated Cardiomyopathy. *Heart Fail Clin* **14**, 139–146 (2018).
67. Gakenheimer-Smith, L. *et al.* Expanding the phenotype of CACNA1C mutation disorders. *Mol Genet Genomic Med* **9**, e1673 (2021).
68. Malfatti, E. *et al.* A Premature Stop Codon in MYO18B is Associated with Severe Nemaline Myopathy with Cardiomyopathy. *J Neuromuscul Dis* **2**, 219 (2015).
69. Dattani, S., Samborska, V., Ritchie, H., Roser, M. Cardiovascular Diseases. Preprint at (2023).
70. Jimenez-Tellez, N. & Greenway, S. C. Cellular models for human cardiomyopathy: What is the best option? *http://www.wjgnet.com/* **11**, 221–235 (2019).
71. 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).
72. Micheletti, R. *et al.* The long noncoding RNA Wisper controls cardiac fibrosis and remodeling. *Sci Transl Med* **9**, (2017).
73. Cai, B. *et al.* The Long Noncoding RNA CAREL Controls Cardiac Regeneration. *J Am Coll Cardiol* **72**, 534–550 (2018).
74. Wu, J. *et al.* Dynamics of RNA localization to nuclear speckles are connected to splicing efficiency. *Sci Adv* **10**, (2024).
75. Barutcu, A. R. *et al.* Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol Cell* **82**, 1035-1052.e9 (2022).
76. Ding, J. H. *et al.* Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J* **23**, 885 (2004).
77. ER, P. *et al.* Transient regenerative potential of the neonatal mouse heart. *Science* **331**, 1078–1080 (2011).
78. 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).
79. Lam, N. T. & Sadek, H. A. Neonatal Heart Regeneration. *Circulation* **138**, 421–423 (2018).
80. 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).
81. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
82. 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).
83. Kolde, R. pheatmap: Pretty Heatmaps. Preprint at (2019).
84. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, (2021).
85. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 1–14 (2013).
86. Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, (2016).

1028 87. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for
1029 interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–
1030 15550 (2005).
1031 88. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation
1032 are coordinately downregulated in human diabetes. *Nature Genetics 2003 34:3* **34**, 267–
1033 273 (2003).
1034 89. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-
1035 aware quantification of transcript expression using dual-phase inference. *Nat Methods*
1036 **14**, 417 (2017).
1037 90. Robinson, M. D., Mccarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
1038 differential expression analysis of digital gene expression data. *BIOINFORMATICS*
1039 *APPLICATIONS NOTE* **26**, 139–140 (2010).
1040 91. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction
1041 of single-cell gene expression data. *Nature Biotechnology 2015 33:5* **33**, 495–502
1042 (2015).
1043 92. Marian, A. J. Molecular Genetic Basis of Hypertrophic Cardiomyopathy. *Circ Res* **128**,
1044 1533–1553 (2021).
1045 93. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**,
1046 764-779.e20 (2020).
1047 94. Schupp, J. C. *et al.* Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung.
1048 *Circulation* **144**, 286–302 (2021).
1049 95. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type
1050 deconvolution with multi-subject single-cell expression reference. *Nat Commun* **10**, 1–
1051 9 (2019).
1052 96. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
1053 reads. *EMBnet J* **17**, 10 (2011).
1054 97. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
1055 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
1056 98. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids*
1057 *Res* **34**, D590–D598 (2006).
1058 99. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res* **48**, D682–D688 (2020).
1059 100. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the
1060 integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat Protoc*
1061 **4**, 1184–1191 (2009).
1062 101. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python.
1063 *Bioinformatics* **36**, 2272–2274 (2020).
1064 102. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in
1065 genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**, (2009).
1066 103. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
1067 genomic features. *Bioinformatics* **26**, 841–842 (2010).
1068 104. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in
1069 multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
1070 105. Kuret, K., Amalietti, A. G., Jones, D. M., Capitanchik, C. & Ule, J. Positional motif
1071 analysis reveals the extent of specificity of protein-RNA interactions observed by CLIP.
1072 *Genome Biol* **23**, 1–34 (2022).
1073 106. Bertero, A. *et al.* Dynamics of genome reorganization during human cardiogenesis
1074 reveal an RBM20-dependent splicing factory. *Nature Communications 2019 10:1* **10**,
1075 1–19 (2019).

107. Buonaiuto, G., Taliani, V., Nicoletti, C., Desideri, F. & Ballarino, M. Preparation of Cardiac Extracts from Embryonal Hearts to Capture RNA–protein Interactions by CLIP. *Bio Protoc* **13**, (2023).
108. Sala, L. *et al.* Musclemotion: A versatile open software tool to quantify cardiomyocyte and cardiac muscle contraction in vitro and in vivo. *Circ Res* **122**, e5–e16 (2018).

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, Dr. Marcella Marchioni for technical help with cell culture and Dr. Maria Celardo and Marco Simula for helpful discussion. This research was funded by: 1) Sapienza University (RM12117A5DE7A45B and RM123188F6B80CE4), to M.B.; 2) Consiglio Nazionale delle Ricerche-CNR (projects DBA.AD005.225-NUTRAGE-FOE2021 and DSB.AD006.371-InvAt-FOE2022), to P.L.; 3) Avvio alla Ricerca Type 2 (Prot. AR223188B40CB2D0) to G.B.; 4) European Union - NextGenerationEU: National Center for Gene Therapy and Drug based on RNA Technology, CN3 - code: CN00000041; National Recovery and Resilience Plan (NRRP) MUR - M4C2 - Action 1.4 - Call "Potenziamento strutture di ricerca e di campioni nazionali di R&S" (Spoke 3 "Neurodegeneration", CUP: B83C22002870006, to M.B. and Spoke 6 "RNA Drug Development", CUP B83C22002860006, to P.L.); 5) by the European Union - Next-GenerationEU - National Recovery and Resilience Plan (NRRP) - M4C2, INVESTMENT N. 1.1, Call "PRIN 2022" (project 2022BYB33L, CUP: B53D23016090006), to M.B. and P.L.; 6) by the European Union - Next-GenerationEU - National Recovery and Resilience Plan (NRRP) - M4C2, INVESTMENT N. 1.1, Call "PRIN 2022 PNRR" (project P2022FFEWN RNA2FUN, CUP: B53D23026140001), to M.B.

Author contributions

M.B. designed, conceived and acquired funding for the study; M.B. and P.L. secured fundings; G.B developed the PA and RHE CRISPR-Cas9 edited cell lines and performed all the gene expression, morphological and functional analyses in hiPSC-derived cardiomyocytes; F.D. developed the *Δ*P cell line and provided support for the gene expression analyses; A.S. performed the alternative splicing analysis on the WT, *Δ*P and GSE200878 dataset, as well as the conservation and *in silico* interaction analyses; A.P. performed the computational analyses on the WT, *Δ*P, GSE130036 and GSE116250 datasets; A.D. carried out computational analyses on the syn7818379 datasets and the cellular deconvolution analysis; G.S. provided support with

the hiPSC handling and gene expression analyses; 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 P.L. and input from the other authors.

Competing interest

The authors declare no competing interests.

MAIN FIGURES

Fig. 1

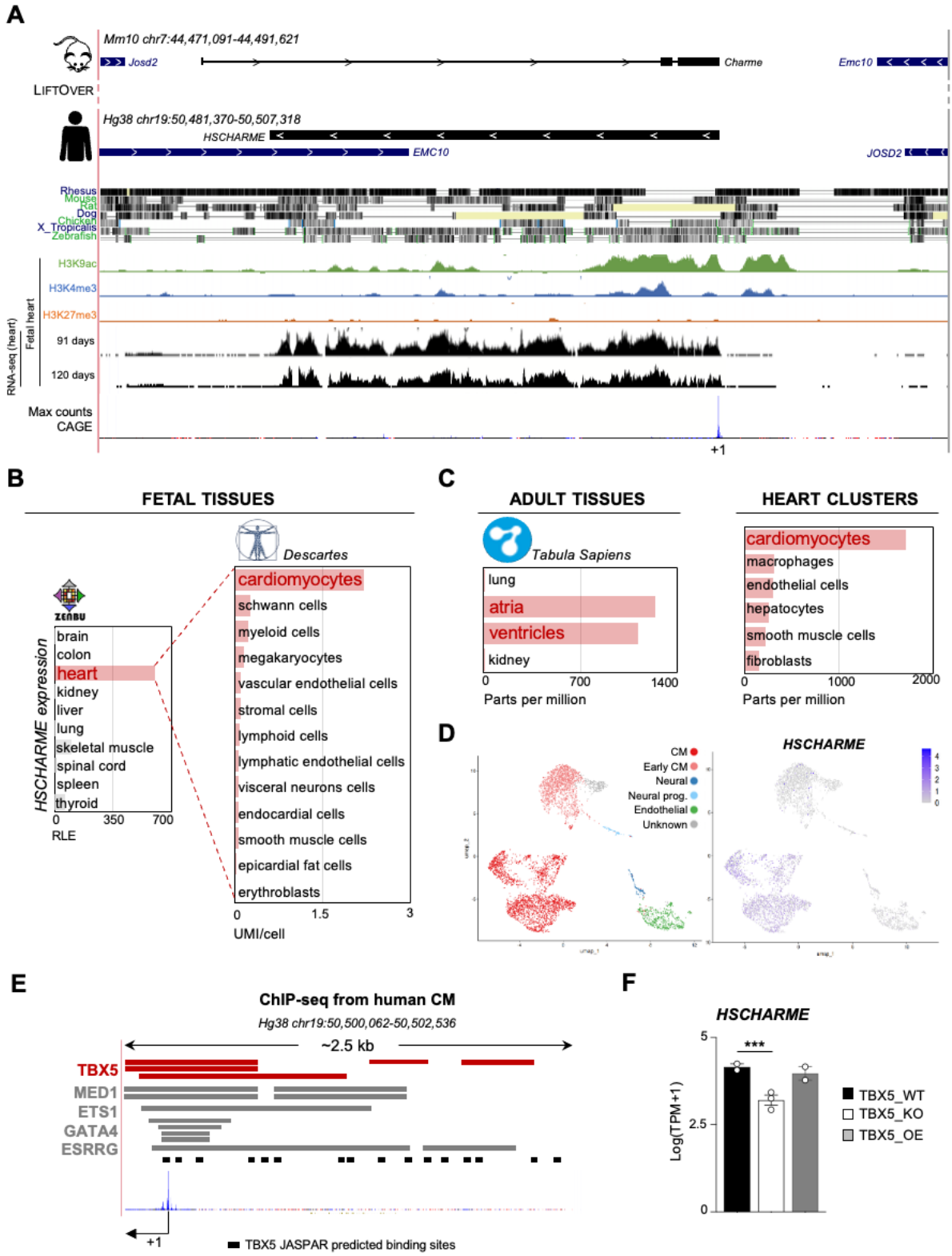


Fig. 1. Identification of conserved synteny of the *pCharme* locus in Homo sapiens. A) UCSC visualization showing the chromosome position and the genomic coordinates of *Charme* gene in mice and humans (*HSCHARME*). UCSC default tracks for Vertebrate Multiz Alignment & Conservation and Histone modifications (H3K9ac, green; H3K4me3, blue; H3K27me3; orange) are also shown, together with the RNA-seq reads taken from GEO and FANTOM5

CAGE datasets. TSS= transcriptional start site. **B)** Left panel: quantification of *HSCHARME* TSS usage (ROI coordinates: Hg38 chr19:50,476,155-50,510,700) in fetal tissues from CAGE data (Phase1 and 2 datasets) from ZENBU FANTOM5 Human hg19 promoterome. For each specific sample, bars represent the Relative Logaritmic Expression (RLE) of the Tag Per Million values of TSS usage. Right panel: quantification *HSCHARME* expression in cardiac cells subtypes from Descartes scRNA-seq atlas27. Bars represent UMI/cell values. **C)** Quantification of *HSCHARME* expression in adult tissues from scRNA-seq data of Tabula Sapiens28 organized by full details (left panel) or by tissue and cell type (right panel) showing *HSCHARME* expression. Bars represent the Parts Per Million values of *HSCHARME* TSS usage per sample. **D)** Left Panel: Integrated UMAP plot of single-cell transcriptomic profiles from hiPSC-CM (2857 cells from Day 14, 2321 cells from Day 45, as in [31],describing cell identity assignment to CM, Early CM, Neural, Neural progenitors, Endothelial, or Unknown (cells that could not be assigned to any specific identity) subpopulations. Right panel: *HSCHARME* expression at single-cell resolution over the UMAP representation of D14 and D45 hiPSC-CM dataset. **E)** UCSC visualization of ReMap ChIP-seq track across *HSCHARME* promoter in human CM. Analysis was performed by restricting the biotype selection to CM and focusing on the 2.5 Kb region upstream *HSCHARME* locus. The genomic coordinates, the *HSCHARME* TSS (+1, black arrow) and the TBX5 JASPAR binding sites (black squares) predicted with JASPAR 2022 (29) (relative profile score threshold=80%) are shown. **F)** Quantification of *HSCHARME* expression from RNA-seq analysis performed in n=2 wild type (WT), n=3 TBX5 knockout (KO) and n=2 TBX5 overexpressing (OE) hiPSC-derived CM (GSE8158530). Data are represented as Log(TPM+1) to avoid negative values with TPM= Transcript per Million mapped reads. Data information: ***$p < 0.001$, adjusted p-value.

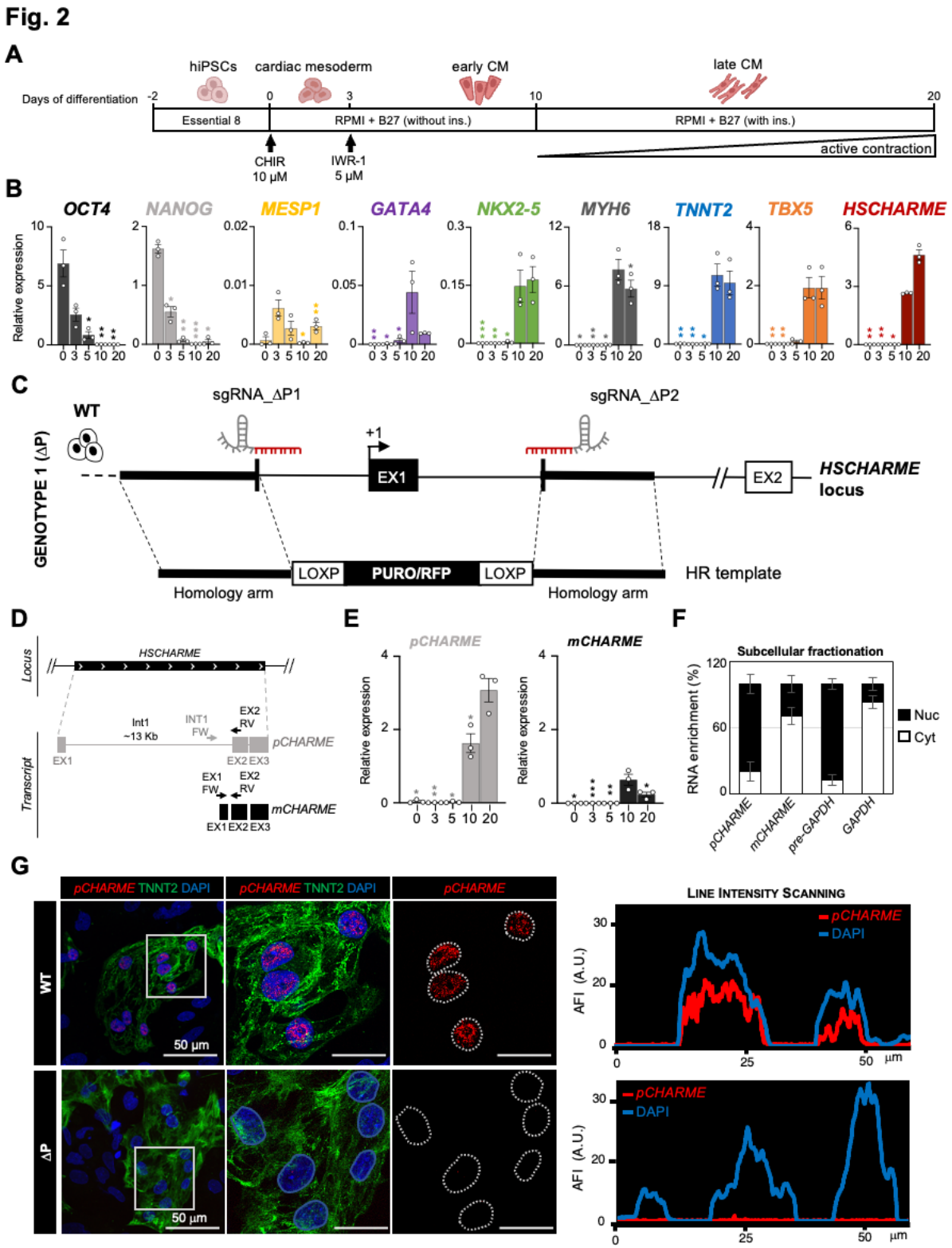


Fig. 2. *HSCHARME* characterization in hiPSC-derived human CM. A) Schematic representation of CM differentiation from hiPSCs. **B)** RT-qPCR amplification of readout mRNAs as well as *HSCHARME* from total RNA extracted from hiPSCs-derived CM at specific time points (D0-D20). Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 3 biological replicates. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (the timepoint with the highest expression). **C)** Schematic representation of the genome editing strategy

design followed to obtain ΔP hiPSC cell line using CRISPR/Cas9 technology. **D)** Schematic representation of *HSCHARME* isoforms as reconstructed by genome assembly. Arrows represent the position of primers for RT-qPCR. **E)** RT-qPCR amplification of *pCHARME* and *mCHARME* from total RNA extracted from hiPSC-derived CM at specific time points (D0-D20). Data were normalized to *ATP5O* mRNA and represent relative expression means (2^{-DCt}) ± SEM of 3 biological experiments. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (the timepoint with the highest expression). **F)** Quantification of the subcellular distribution of *pCHARME* and *mCHARME* isoforms from biochemical fractionation of D10 CM. The histogram shows the RT-qPCR quantification of the relative % of RNA abundance in cytoplasmic versus nuclear compartments. *GAPDH* and *pre-GAPDH* RNAs were used, respectively, as cytoplasmic and nuclear controls. **G)** Left panel: Representative RNA-FISH staining for *pCHARME* (red) combined with TNNT2 immunofluorescence (green) in WT and ΔP hiPSC-derived D20 CM. 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: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change).

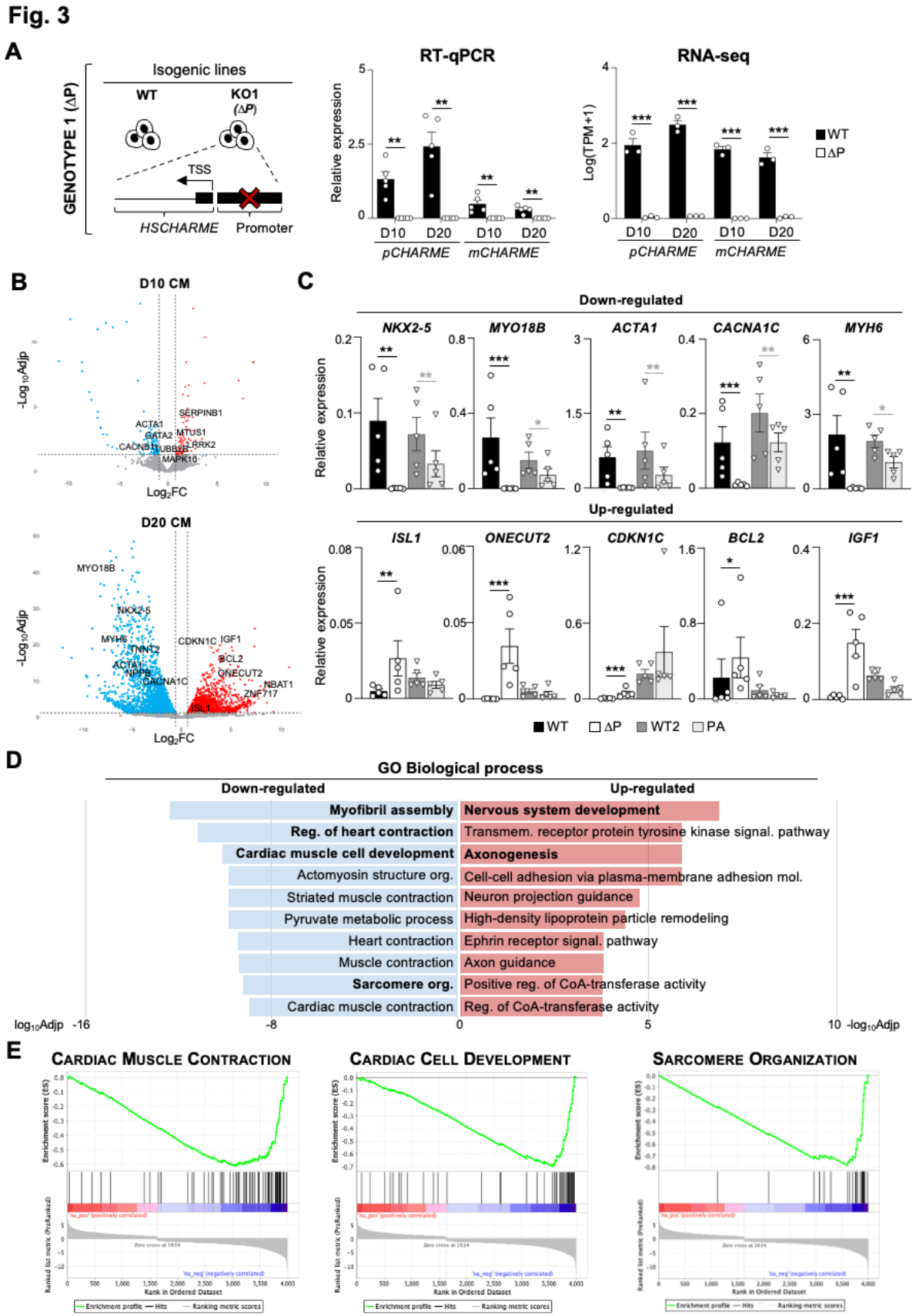


Fig. 3. Genome-wide profiling of WT and ΔP hiPSC-derived CM. A) Left panel: Schematic representation of the genome editing strategy design followed to obtain ΔP hiPSC cell line using CRISPR/Cas9 technology. Middle panel: RT-qPCR amplification of *pCHARME* and

mCHARME from WT vs ΔP D10 and D20 CM. Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 5 biological replicates. Data information: **p < 0.01; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change). Right panel: Quantification by RNA-seq (TPM) of *pCHARME* and *mCHARME* expression in WT vs ΔP hiPSC-derived D10 and D20 CM. Data are represented as Log(TPM+1) to avoid negative values. Data information: ***p < 0.001, FDR. **B)** Volcano plots showing differentially expressed genes (DEGs) from the RNA-seq analysis of WT vs ΔP hiPSC-derived CM at D10 vs CTRL at the same timepoint (upper panel) and D20 vs CTRL at the same timepoint (lower panel). Significantly up-regulated (FDR<0.05; $log_2FC>1$) and down-regulated (FDR<0.05; $log_2FC<-1$) genes are represented in red and blue respectively. **C)** RT-qPCR quantification of down-regulated (upper panel) and up-regulated (lower panel) DEGs in WT vs ΔP (black) and WT2 vs PA (grey) hiPSC-derived D20 CM. Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 5 biological experiments. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (WT for ΔP-CM and WT2 for PA-CM). Data information: *p < 0.05, **p < 0.01, ***p < 0.001; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change). **D)** Gene ontology (GO) enrichment analysis performed with EnrichR on down-regulated (left) and up-regulated (right) DEGs in WT vs ΔP hiPSC-derived D20 CM. Bars indicate +/–log10 adjusted p-value (log10Adjp and –log10Adjp) 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 resulting down-regulated in WT vs ΔP D20 CM.

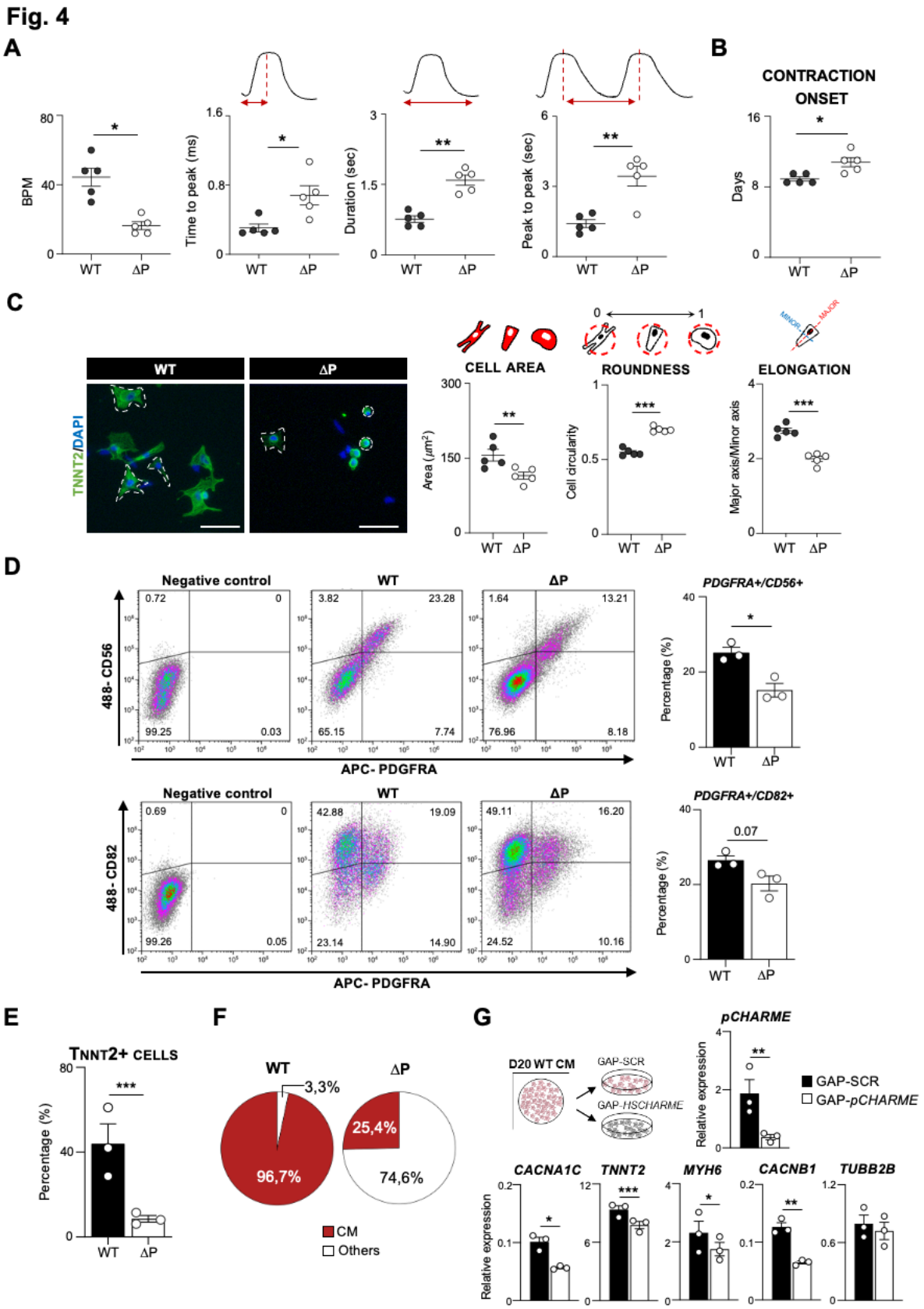


Fig. 4. Functional implication of *HSCHARME* in cardiomyogenesis. A) Quantitative analysis of the beat rate (expressed as beats per minute=BPM) and contraction parameters in WT and ΔP D20 CM. Representative images of the specific parameter are shown above each graph. Data represent means ± SEM of 5 biological replicates. **B)** Onset (days) of spontaneous

contraction of WT and ΔP hiPSC-derived CM. Data represent means ± SEM of 5 biological replicates. **C)** Left panel: Representative images of WT and ΔP hiPSC-derived D20 CM selected for morphological analyses (white dashed outlines) after TNNT2 (green) and DAPI (blue) staining. Right panel: Population measurements of CMs morphological features. A schematic representation of the specific measure quantified is shown above each plot. Black bars represent means ± SEM of 5 biological replicates. **D)** Left panels: Representative flow-cytometry density plot of WT and ΔP hiPSC-CM analyzed at day 10 for cardiac progenitors (PDGFRA/CD56) and cardiac-committed progenitors (PDGFRA/CD82) marker expression. Percentages of each cell population are shown. Right panels: PDGFRA+/CD56+ and PDGFRA+/CD82+ positive cells were set by analysing negative control samples stained with secondary antibody only. Percentages are plotted on the right and represent mean ± SEM of 3 biological replicates. Data information: *$p < 0.05$; one-sample two-tailed Student's t-test on LogFC against the null hypothesis of zero (no change). **E)** Flow cytometry quantification of cardiac troponin-T (TNNT2) positive cells percentage in WT and ΔP hiPSC-derived D20 CM. Data represent mean ± SEM of 3 biological replicates. Data information: ***$p < 0.001$; one-sample two-tailed Student's t-test on LogFC against the null hypothesis of zero (no change). **F)** Pie-chart displays the estimated proportion of CM (red) and other cell types (white, **Supplementary Table 1**) based on deconvolution analysis of WT and ΔP CM (D20). **G)** RT-qPCR quantification of *pCHARME* and selected mRNA in CM (D20) transfected with control (GAP-SCR) or *pCHARME*-targeting (GAP-*pCHARME*) antisense LNA-GapmeRs. See '**Materials and methods**' for details. Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 3 biological experiments. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (GAP-SCR). Data information: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change).

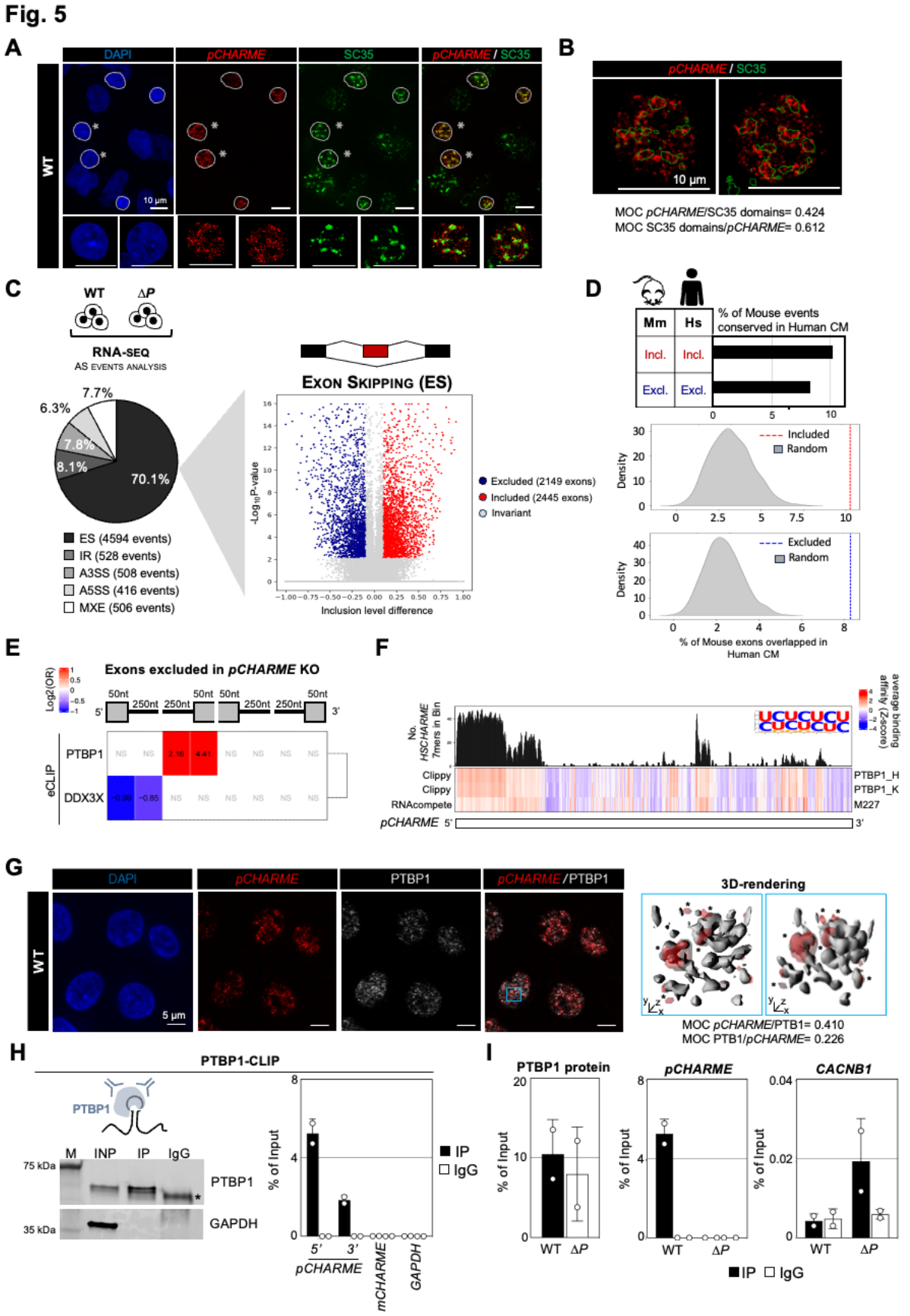


Fig. 5. *pCHARME* regulates the alternative splicing of cardiac-specific pre-mRNAs. **A)** Representative maximum 2D projection of the full-size images of confocal caption for RNA-FISH *pCHARME* (red) staining combined with SC35 (green) and DAPI (blue) IF in WT

D20 CM. The white lines indicate *pCHARME*-positive nuclei. Digital magnifications of the nuclei highlighted by white asterisks are reported below. **B)** Graphic visualization of *pCHARME*/SC35 domains colocalization signals highlighting the boundary of SC35 domains overlaid with *pCHARME*. Colocalization indexes between *pCHARME* and SC35 domains signals are indicated by Mander's overlap coefficients (MOC). **C)** Left panel: Pie Chart depicting the portion of significantly altered (FDR<0.05) splicing events detected by rMATS comparing WT vs ΔP D20 RNA-seq samples. All classical splicing alterations were investigated, such as Exon skipping (ES); Intron retention (IR); alternative 3' splice site (A3SS); alternative 5' splice site (A5SS) and Mutually exclusive exons (MXE). Right panel: Volcano plot depicts significant (FDR<0.05; Inclusion level variations stronger than 10%) ES events in ΔP D20 CM. X-axis represent exon inclusion ratio while y-axis represent –log10 of P-value. A schematic representation of the investigated event is shown above. **D)** Upper panel: Barplot shows the proportion of exon skipping events identified in murine *pCharme* KO hearts and regulated by *pCHARME* in human CM. Events are stratified by the regulatory outcome: exon inclusion (Incl.) and exon exclusion (Excl.). Lower panel: Density plots represent distribution of conserved events expected by chance (grey distribution). The observed conservation is marked by vertical dashed lines, red for inclusion and blue for exclusion. **E)** Upper panel: Schematic representation of the pre-mRNA regions analyzed in the RBP binding enrichment analysis. Lower panel: Heatmap displaying the $\log_2$(Odds Ratio) of eCLIP binding site (enrichment or depletion) in *pCHARME*-regulated skipping exon (SE) events, compared to unaffected control exons, across the analyzed pre-mRNA regions (from left to right: exonic-upstream, intronic-upstream, intronic-center(5'SS), exonic-center(5'SS), exonic-center(3'SS), intronic-center(3'SS), intronic-downstream and exonic-downstream). Statistically significant enrichment or depletion is indicated by red or blue colors, respectively, while white denotes the absence of statistical significance (FDR > 0.05), also marked by the letters "NS" (Not Significant). Statistical significance was determined using Fisher's exact test, and p-values were corrected for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) method. **F)** Heatmap displaying the PTBP1 binding affinity for *pCHARME* RNA sequence. The sequence is represented from 5'-end (left) to 3'-end (right) in segments of 50 nucleotides, with a 10-nucleotide sliding step. Bar plot in the top panel displays the total number of occurrences of *pCHARME*-specific 7-mers within each defined segment of the transcript. On the right, the sequence logo generated from the 16 distinct 7-mer types that constitute the *pCHARME*-specific signature is shown. Heatmap below shows the row-scaled (Z-score) PTBP1 binding affinity scores for *pCHARME* segments, based on data from two eCLIP experiments (HepG2 and K562) and one RNAcompete experiment (M227_0.6). Binding affinity scores were calculated as the average PEKA score of all 5-mers within each 50-nucleotide segment for the Clippy-analyzed eCLIP datasets, and as Z-scores for the RNAcompete experiment. **G)** Left panel: Representative confocal images showing RNA-FISH staining for *pCHARME* (red) combined with IF for PTBP1 (white) and DAPI (blue) in WT (D20) CM. Right panel: composite image of *pCHARME* (red) and PTB1 (grey) as volume view of the region depicted with in the blue square. Asterisks point out *pCHARME* regions overlaid with PTBP1. Colocalization indexes between *pCHARME* and PTBP1 signals are indicated by Mander's overlap coefficients (MOC). **H)** Left panel: Western blot analysis of PTBP1 protein from PTBP1-CLIP assay in WT D20 CM. GAPDH protein serves as a loading control. Input (Inp) samples represent 10% of the total protein extracts. Right panel: RT-qPCR quantification of *pCHARME* (amplified at both 5′ and 3′ intron-1 ends) and *mCHARME* recovery in PTBP1 IP and IgG samples. *GAPDH* RNA serves as negative control. Values are expressed as percentage of input and represent mean ± SD of 2 biological replicates. I) Left panel: PTBP1 protein recovery in WT and ΔP D20 CM. Quantification of the chemiluminescent signal was performed with the ImageJ tool. Values are expressed as percentage of input and represent

mean ± SD of 2 biological replicates. Middel panel: RT-qPCR quantification of *pCHARME* (amplified at 5′ intron-1 end) recovery in PTBP1 IP and IgG samples from WT and ΔP D20 CM. Values are expressed as percentage of input and represent mean ± SD of 2 biological replicates. Right panel: RT-qPCR quantification of *CACNB1* RNA recovery in PTBP1 IP and IgG samples from WT and ΔP D20 CM. Values are expressed as percentage of input and represent mean ± SD of 2 biological replicates.

Fig. 6

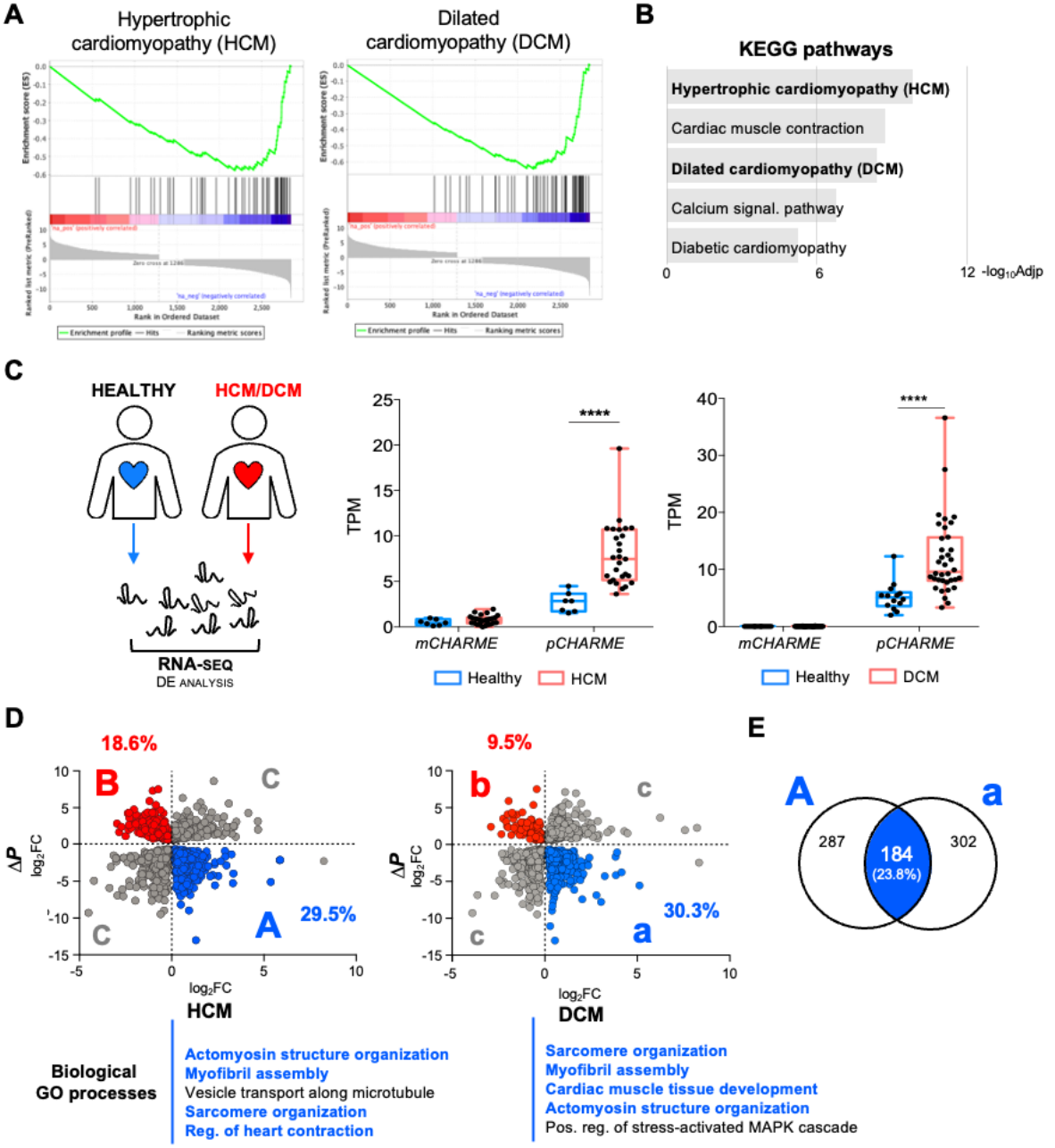


Fig. 6. Identification of *pCHARME* targets in hypertrophic and dilated cardiomyopathies. **A)** GSEA plot showing the enrichment of the Human Phenotype Ontology terms "Hypertrophic Cardiomyopathy" and "Dilated Cardiomyopathy" that resulted strongly down-regulated in the analysis of ΔP vs WT D20 CM. **B)** KEGG pathways enrichment analysis performed with EnrichR on significant DEGs from ΔP vs WT D20 CM. Bars indicate the top categories of biological processes in decreasing order of –log10 adjusted p-value (–log10Adjp). All the represented categories show an Adjp value <0.05. **C)** Quantification of *mCHARME* and

pCHARME expression from RNA-seq analysis performed in healthy (CTRL) vs patients of hypertrophic (left panel) and dilated (right panel) cardiomyopathies. Data information: *p < 0.05; ***p < 0.01; ****p<0.0001; FDR **D)** Scatterplots representing the log2FC comparison between DEGs in ΔP 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. **E)** Venn diagrams depicting the intersection between the "A" (ΔP vs WT down-regulated DEGs crossed with Healthy vs HCM up-regulated DEGs) and "a" (ΔP vs WT down-regulated DEGs crossed with Healthy vs DCM up-regulated DEGs) categories.

Fig. 7

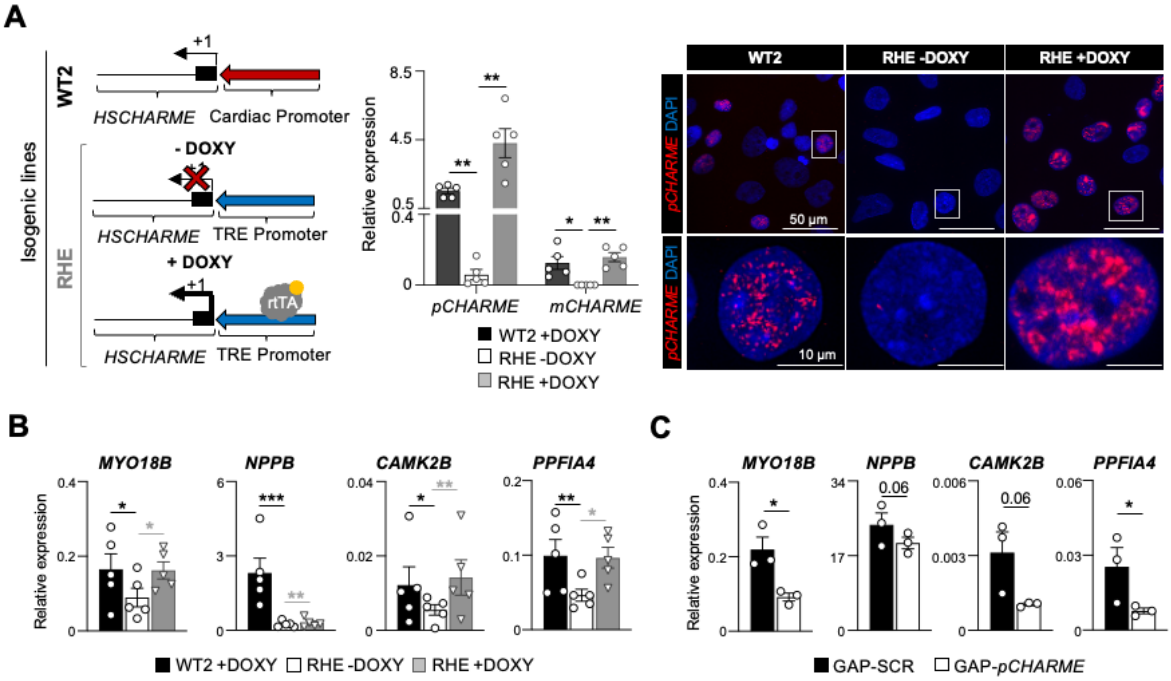


Fig. 7. *pCHARME* dosage impacts the expression of disease-linked genes. A) Left panel: Schematic representation of the CRISPR/Cas9 genome editing strategies used to obtain the RHE-hiPSCs. Middle panel: RT-qPCR quantification of *pCHARME* and *mCHARME* expression in WT2, RHE (-DOXY) and RHE (+DOXY) hiPSC-derived CM. Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 5 biological replicates. Right panel: Representative maximum 2D projection of the full-size image of confocal caption for *pCHARME* RNA-FISH (red) and DAPI (blue) staining in WT2, RHE (-DOXY) and RHE (+DOXY) CM (D20). Digital magnifications of the nuclei highlighted by white squares are reported below. **B)** RT-qPCR quantification of pathology-linked transcripts in WT2, RHE (-DOXY) and RHE (+DOXY) hiPSC-derived CM (D20). Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 5 biological replicates. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (RHE -DOXY). **C)** RT-qPCR quantification of pathology-linked transcripts in WT CM (D20) transfected with control (GAP-SCR) or *pCHARME*-targeting (GAP-*pCHARME*) antisense LNA GapmeRs. See '**Materials and methods**' for details. Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 3 biological experiments. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (GAP-SCR). Data

information: *p < 0.05; ***p < 0.01; ****p<0.0001; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change).

Fig. 8

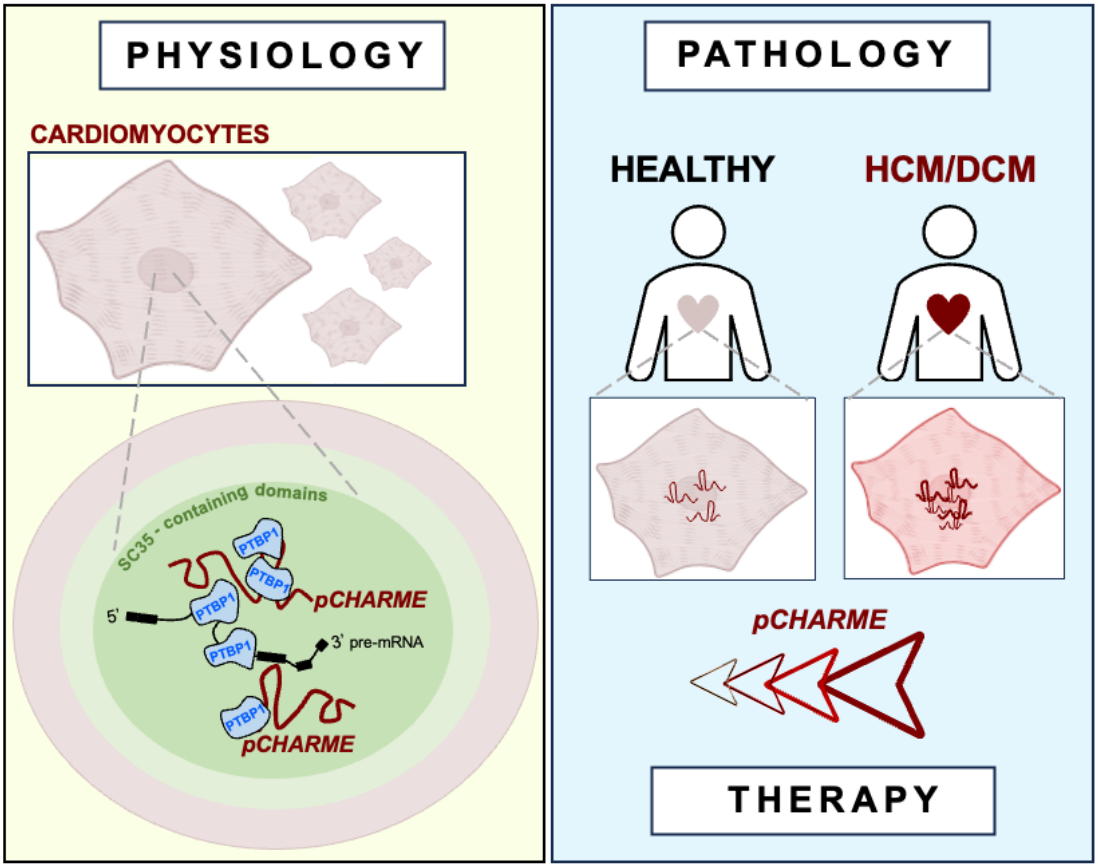
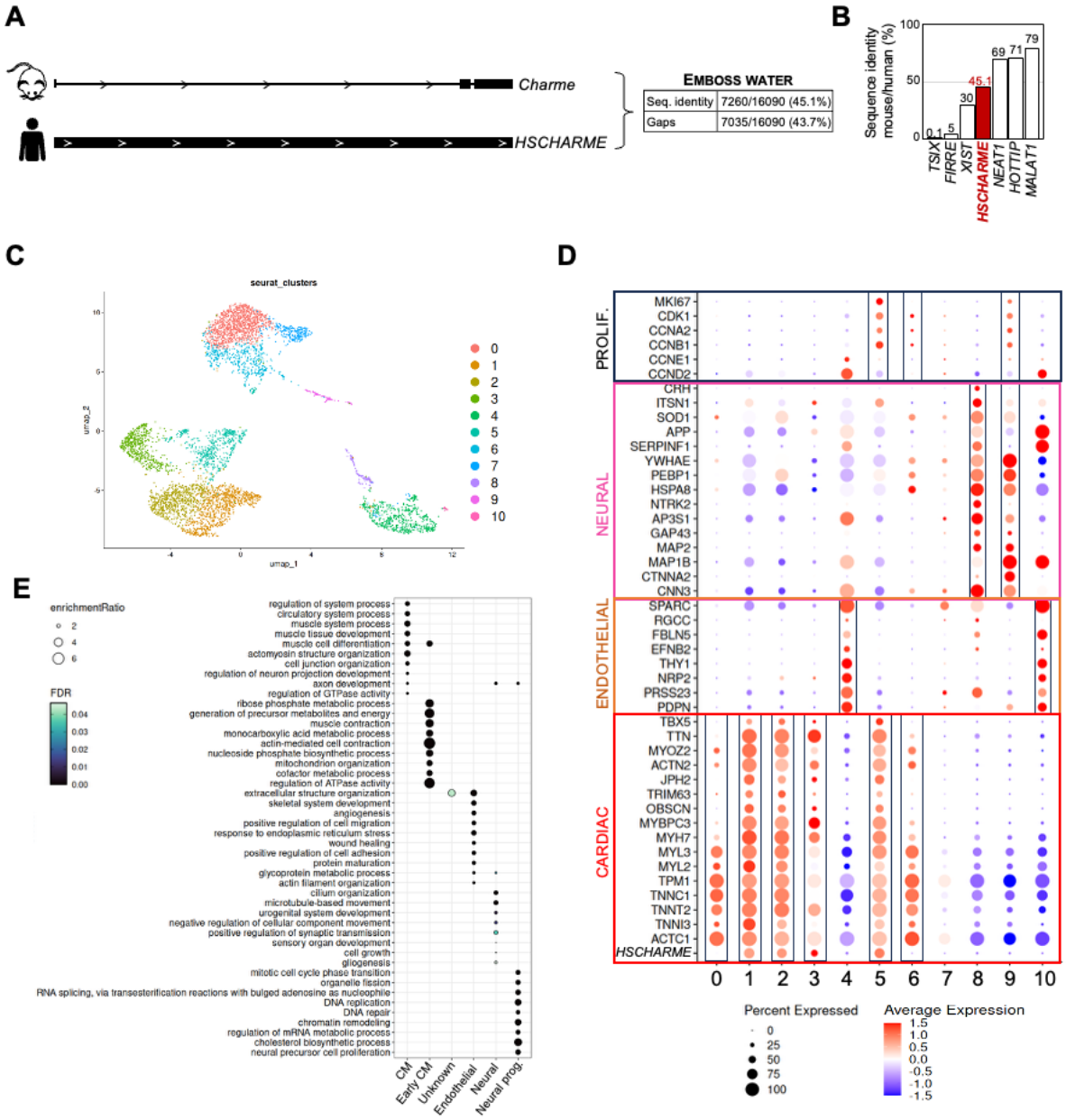
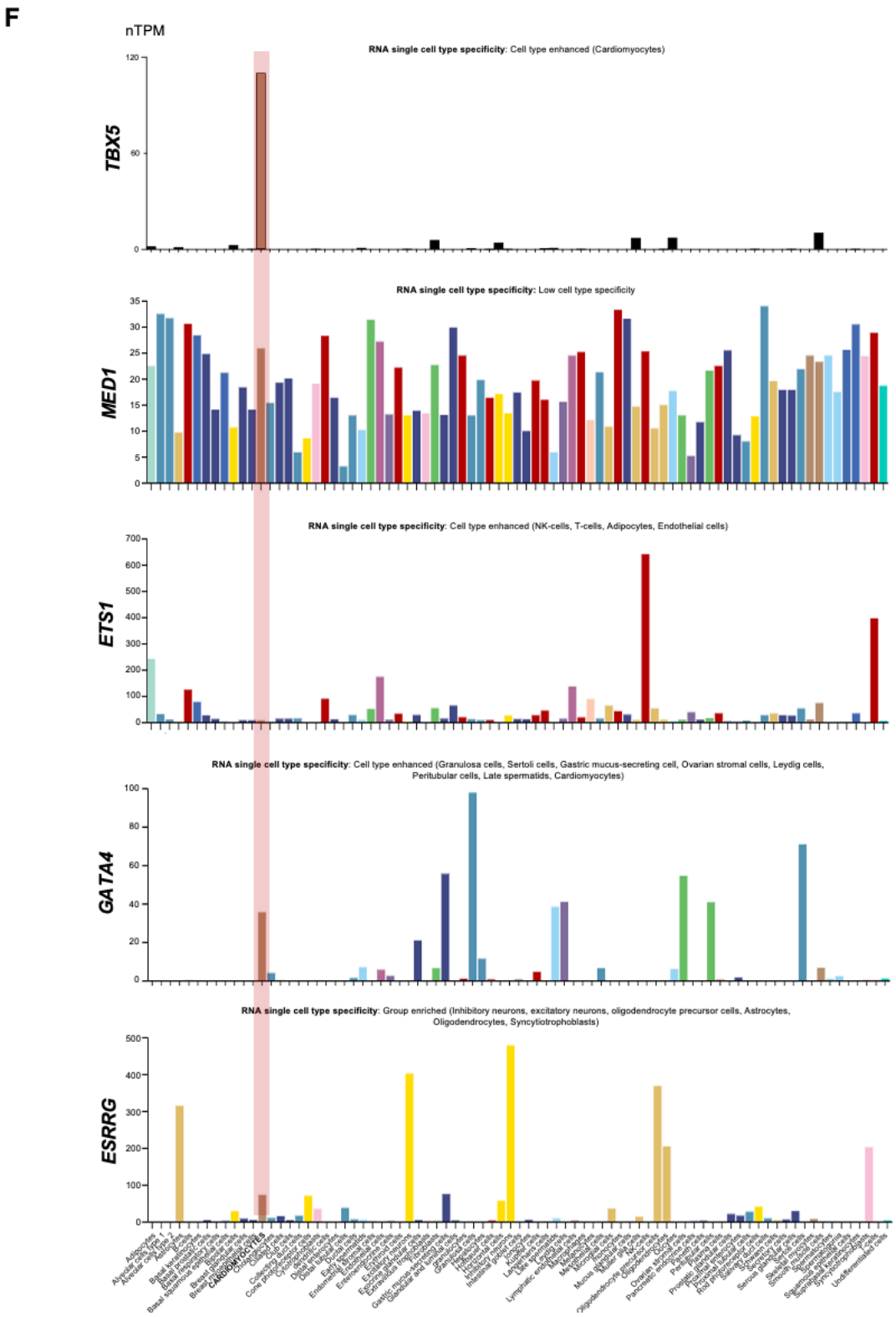


Fig. 8. Overview of *pCHARME* role in cardiomyocyte physiology and disease. Under physiological conditions (Physiology) *pCHARME* localises within S35-domains and acts as a decoy for the splicing repressor PTBP1, thereby positively regulating the expression of genes involved in the differentiation of human cardiomyocytes. In DCM and HCM cardiomyopathy hearts (Pathology), *pCHARME* is aberrantly upregulated, and its target genes -many of which are associated with the diseases- are also dysregulated. This scenario envisages novel therapeutic approaches (Therapy) aimed at regulating disease-linked genes through *pCHARME* modulation.

SUPPLEMENTARY FIGURES

Supplementary Fig. 1

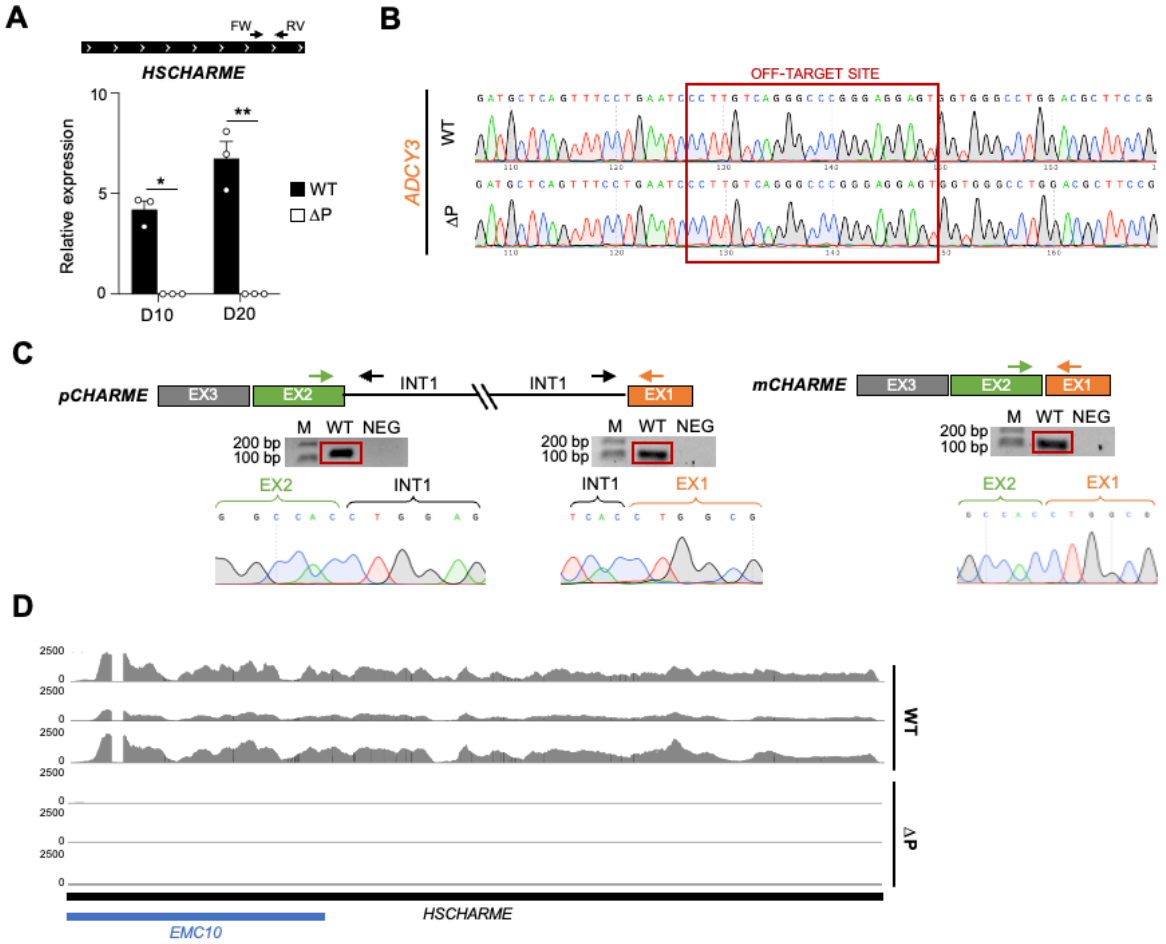




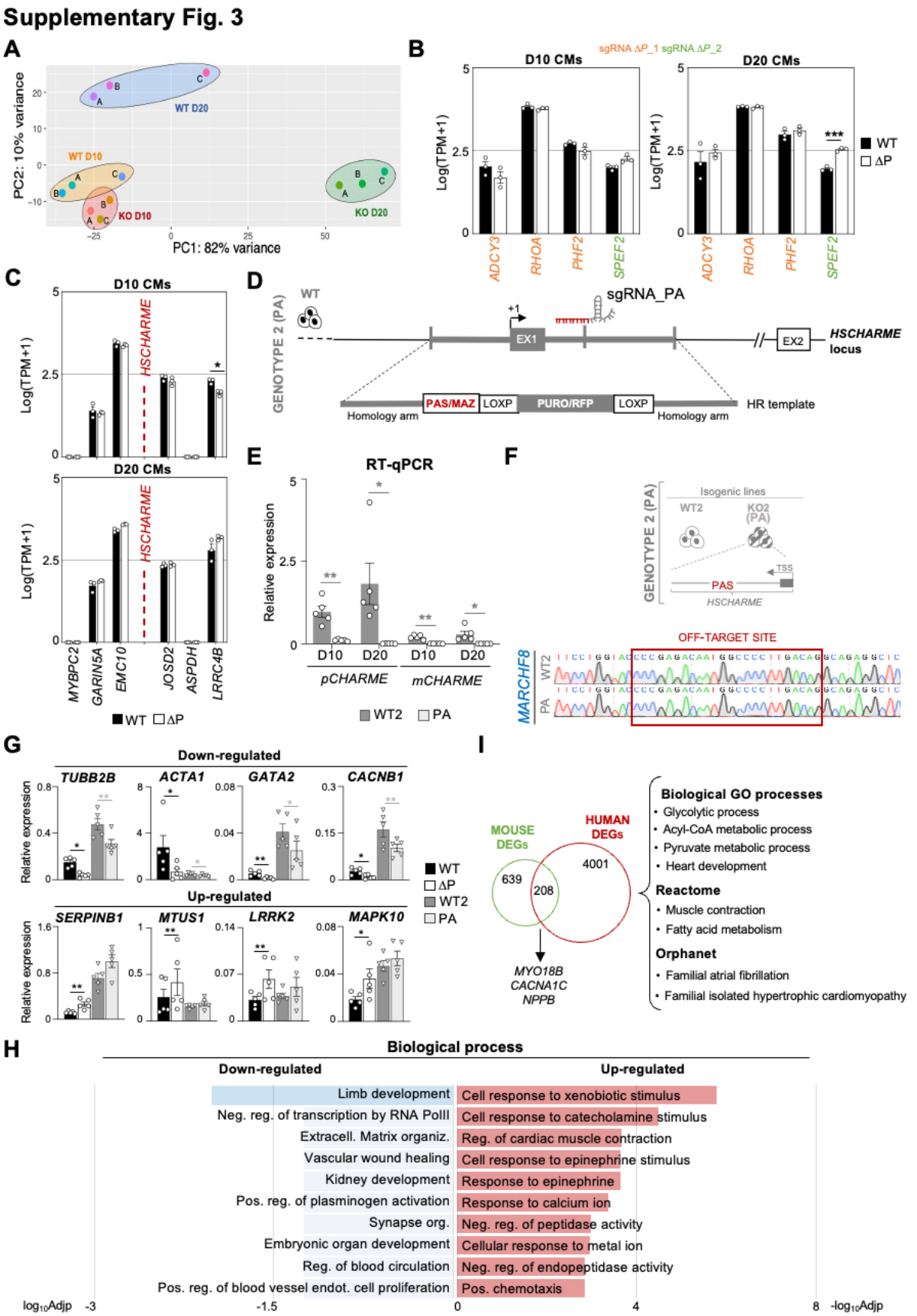
Supplementary Fig. 1. A) Table reports the sequence conservation values between the mouse and human *Charme* genes. The data were obtained by local sequence using the Smith–Waterman algorithm available at Emboss water (https://www.ebi.ac.uk/jdispatcher/psa/emboss_water). **B)** Barplot reporting the percentage of sequence identity between mice and humans for a subset of well- known lncRNAs. C) UMAP visualization of single-cell transcriptomic profiles from hiPSC-CM (2857 cells from Day 14, 2321 cells from Day 45 as in [31]. Cells are colored based on the clusters identified through

unsupervised clustering by Seurat. **D)** Dot plot of scaled average gene expression across cell clusters of neural, endothelial, cardiac and proliferative cell markers. Dot size corresponds to the proportion of cells within a cluster expressing the indicated gene, and color indicating the scaled expression level. Boxes indicates the correspondence between gene markers and cell types. **E)** Dot plot of gene ontology categories across cell populations. The plot displays only the weighted set categories. Dot size corresponds to the enrichment ratio of terms, and color indicates the statistical significance (FDR). **F)** Quantification of *ESRRG, MED1, ETS1, GATA4 and TBX5* expression levels at single cell resolution as reported in the Human Protein Atlas Portal. RNA expression is quantified in nTPM (nTPM = normalized Transcripts Per Million).

Supplementary Fig. 2

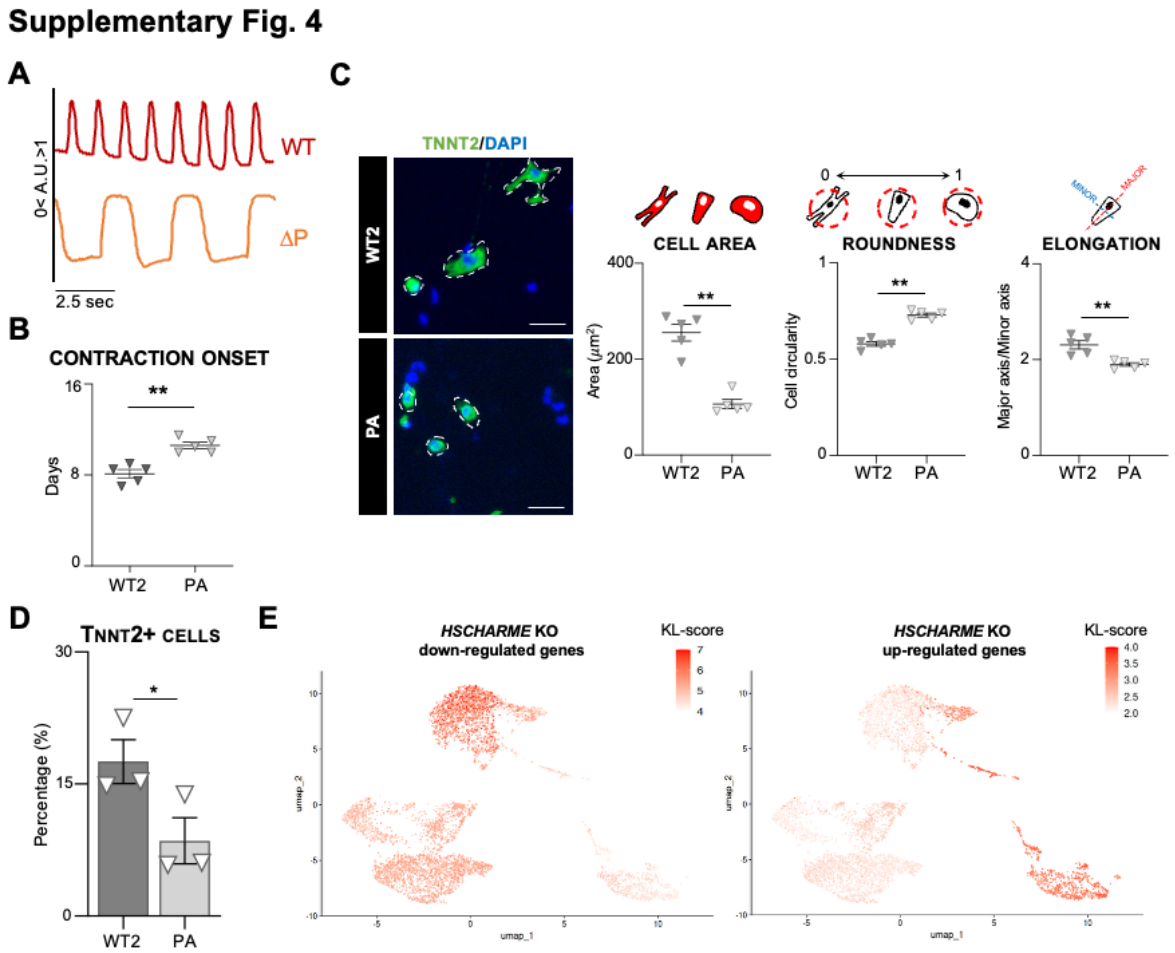


Supplementary Fig. 2. A) RT-qPCR amplification of *HSCHARME* from WT vs ΔP D10 and D20 CM. Data were normalized to *ATP5O* mRNA and represent relative expression means (2^{-DCt}) ± SEM of 3 biological replicates. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (WT). Data information: *p < 0.05; ***p < 0.01; ****p<0.0001; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change). **B)** Representative example of sequencing results from WT and ΔP on the predicted off-target *ADCY3* gene. The red square indicates the region interested in the off-targeting. **C)** sqRT-PCR amplifications of *pCHARME* (5' and 3'-end) and *mCHARME* from WT-CM. Bands in the red square were extracted and subjected to Sanger sequencing as reported below. **D)** RNA-seq normalized read coverage tracks (TPM) across *HSCHARME* locus from WT and ΔP-CM (D20).



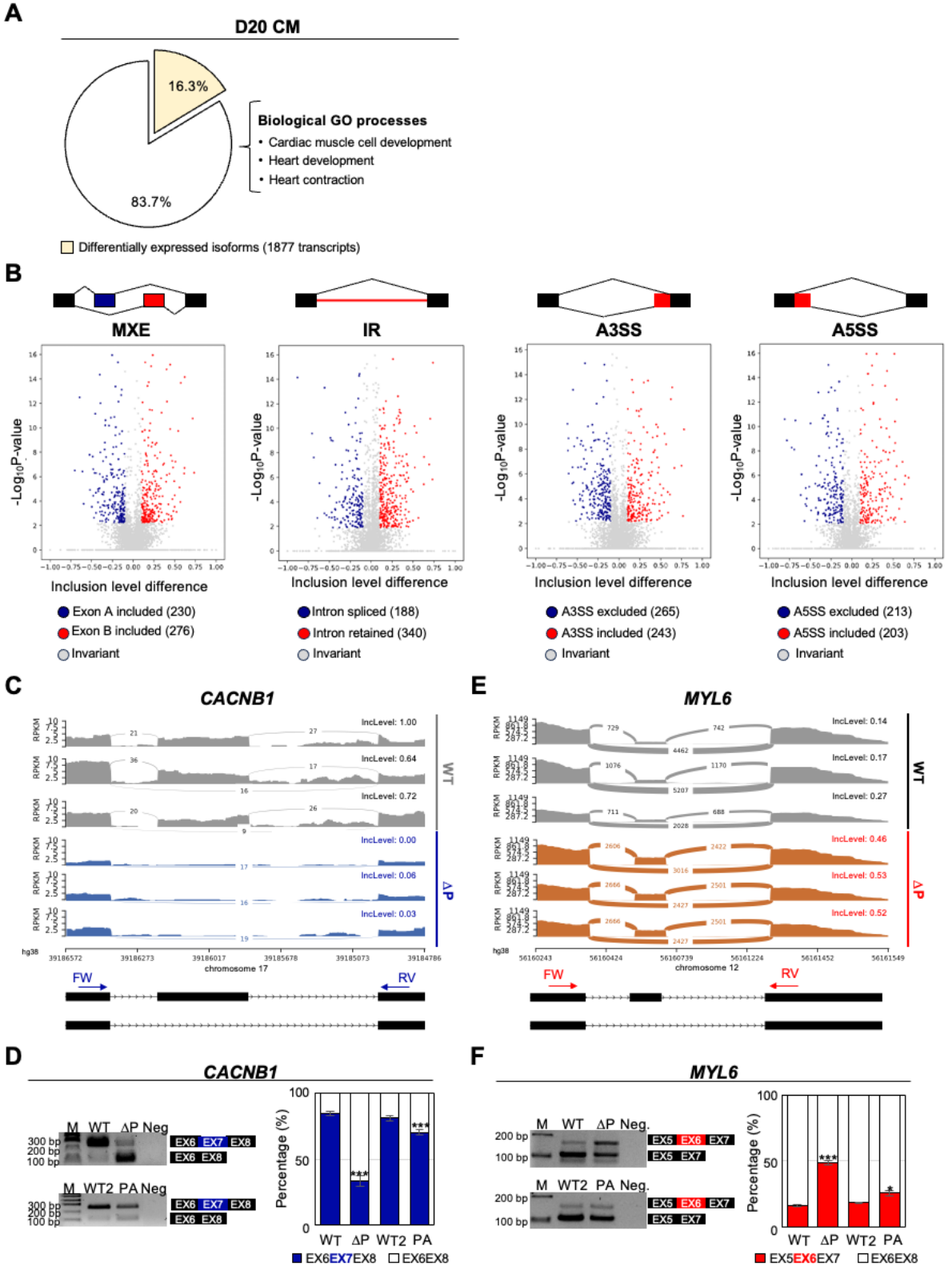
Supplementary Fig. 3. A) Principal Component Analysis (PCA) performed on WT and ΔP RNA-seq datasets at D10 and D20 hiPSC-derived CM. **B)** Quantification by RNA-seq (TPM) of sgRNAs off-targets expression in WT vs ΔP hiPSC-derived D10 (left panel) and D20 (right

panel) CM. Data are represented as Log(TPM+1) to avoid negative values. Data information: ***p<0.001, FDR. **C)** Quantification by RNA-seq (TPM) of *HSCHARME*-neighbouring genes expression in WT vs ΔP hiPSC-derived D10 (upper panel) and D20 (lower panel) CM. Data are represented as Log(TPM+1) to avoid negative values. **D)** Representation of the genome editing strategy design followed to obtain the PA hiPSC cell line by CRISPR/Cas9. **E)** RT-qPCR amplification of *pCHARME* and *mCHARME* from WT2 *vs* PA (D10 and D20) CM. Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 5 biological replicates. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (WT2). **F)** Representative example of sequencing results from WT2 and PA on the predicted off-target *MARCHF8.* The red square indicates the region interested in the off-targeting. A schematic representation of the genome editing strategy is represented on the top. **G)** RT-qPCR quantification of down-regulated (upper panel) and up-regulated (lower panel) DEGs in WT *vs* ΔP (black) and WT2 vs PA (grey) hiPSC-derived CM (D10). Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 5 biological replicates. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (WT for ΔP-CM and WT2 for PA-CM). **H)** Gene ontology (GO) enrichment analysis performed with EnrichR on down-regulated (left) and up-regulated (right) DEGs in WT *vs* ΔP hiPSC-derived CM (D10). 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. **I)** Venn diagrams depicting the intersection between mouse and human DEGs in response to *Charme* ablation. The top GO categories for biological process, Reactome and Orphanet are shown on the right. Notable genes are reported below. Data information: *p < 0.05; ***p < 0.01; ****p<0.0001; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change).

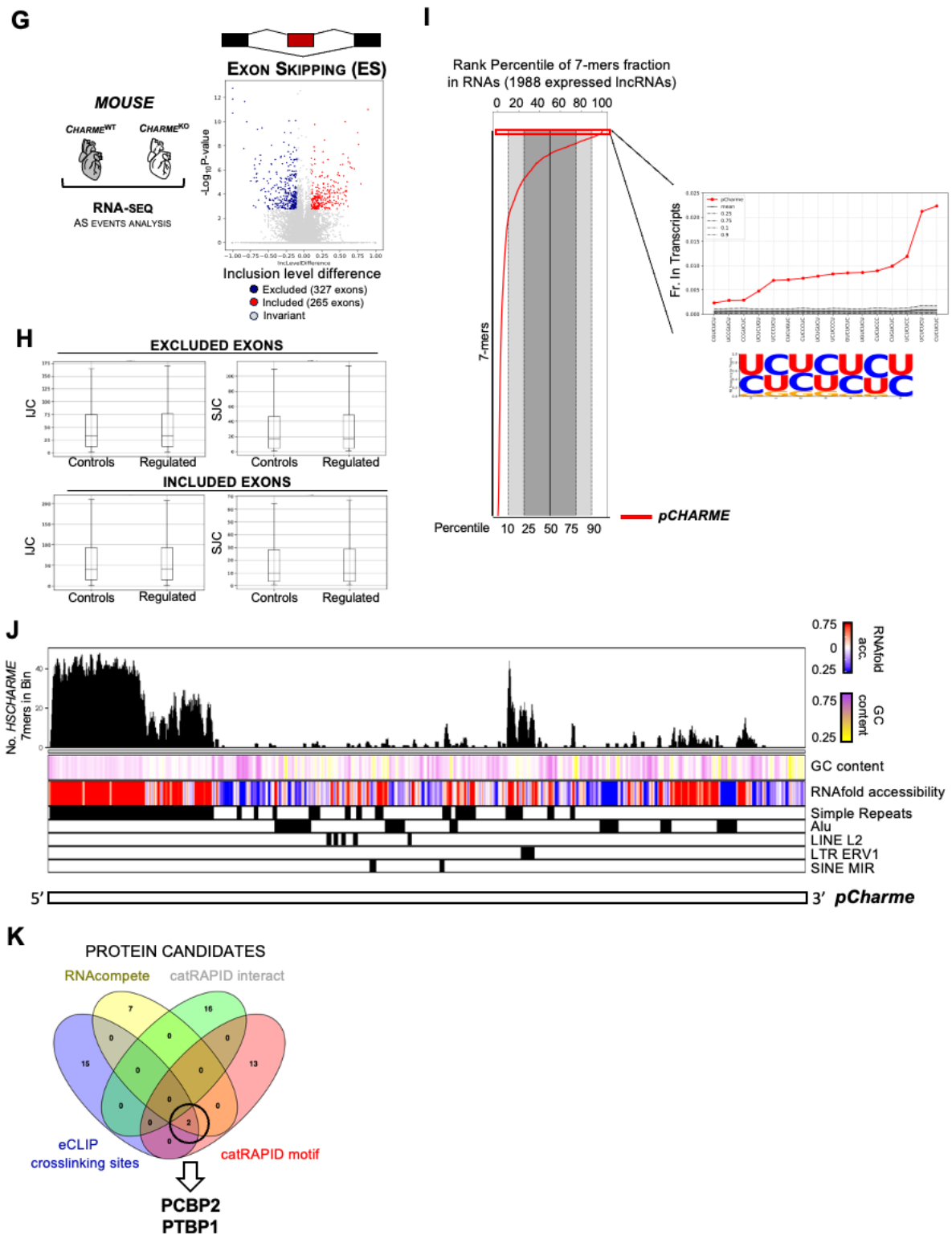


Supplementary Fig. 4. A) Representative traces of normalized spontaneous contraction profile of WT (red) and ΔP (orange) hiPSC-derived CM (D20). Traces derive from the analisys of brightfield 2D videos with the MUSCLEMOTION plugin. **B)** Onset (days) of spontaneous contraction of WT and ΔP hiPSC-derived CM. Data represent means ± SEM of 5 biological replicates. **C)** Left panel: Representative images of WT2 and PA hiPSC-derived CM (D20) selected for morphological analyses (white dashed outlines) after TNNT2 (green) and DAPI (blue) staining. Scale bar=50 µm. Right panel: Population measurements of CM morphological features. A schematic representation of the specific measure quantified is shown above each plot. Black bars represent means ± SEM of 5 biological replicates. **D)** Flow cytometry quantification of cardiac troponin-T (TNNT2) positive cells (%) in WT2 and PA hiPSC-derived CM (D20). Data represent mean ± SEM of 3 biological replicates. Data information: *p < 0.05; one-sample two-tailed Student's t-test on LogFC against the null hypothesis of zero (no change). **E)** UMAP visualization of KL (Kullback–Leibler) score distribution across single cells (see **"Material and Methods"** for details). The left panel ("KL Score Down") displays the KL enrichment score of downregulated genes in *HSCHARME* ΔP at day 20. The right panel ("KL Score Up") displays the KL enrichment score of upregulated genes in *HSCHARME* ΔP (D20). Data information: *p < 0.05, **p < 0.01, ***p < 0.001, paired two-tailed Student's t-test.

Supplementary Fig. 5



1427

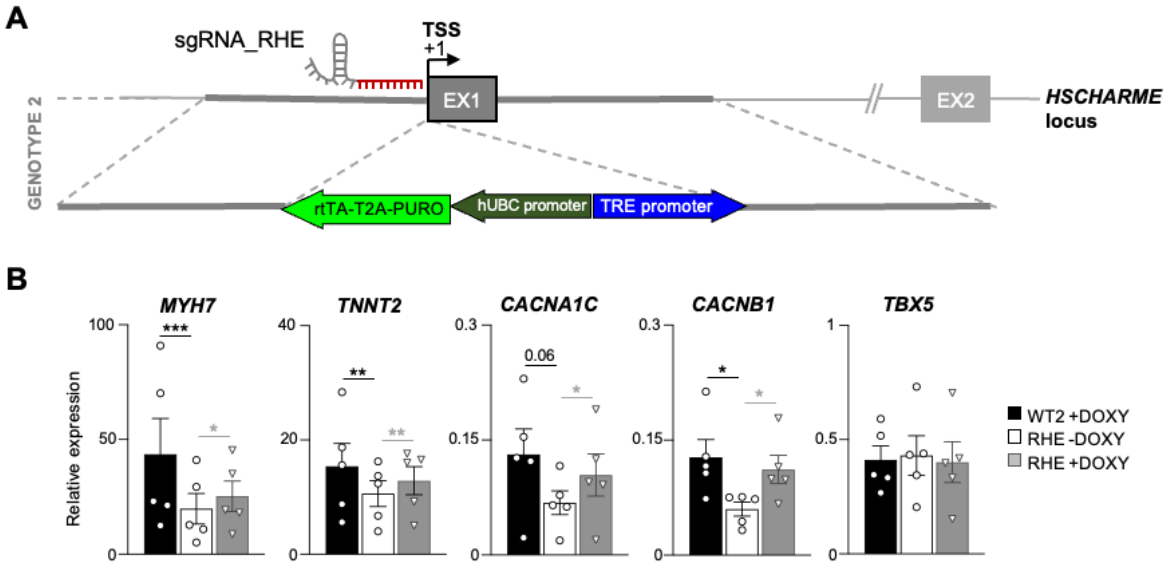


Supplementary Fig. 5. A) Pie Chart representing the fraction of significant (FDR<0.05) differentially expressed transcript isoforms in ΔP *vs* WT CM (D20). The top 3 GO categories for biological process and molecular function are shown on the right. **B)** Volcano plots depicting significantly altered splicing events (FDR<0.05; Inclusion level variations stronger than 10%) in ΔP-CM (D20). All classical splicing alterations are reported, such as Exon skipping (ES); Intron retention (IR); alternative 3' splice site (A3SS); alternative 5' splice site (A5SS) and Mutually exclusive exons (MXE). X-axis represent inclusion ratio while y-axis represent $-\log_{10}$ of P-value. A schematic representation of the specific investigated event is

shown above each plot. **C)** Sashimi plots of *CACNB1* exon skipping events in WT and ΔP hiPSC-derived CM (D20). The X-axis indicates genomic location while the Y-axis reports expression levels as Reads Per Kilobase of transcript per Million mapped reads (RPKM). The lines connecting different exons (bridges) indicate reads on the specific splicing junction. Quantification of splicing reads is reported in each bridge while Inclusion levels (IncLevel) are reported above each sample. The exonic structure of the event is shown below together with oligonucleotides position used for sqRT-PCR validation. **D)** Left panel: Representative electrophoresis analyses of *CACNB1* exon skipping events by sqRT-PCR from WT *vs* ΔP or WT2 *vs* PA CM (D20). Right panel: Quantification of sqRT-PCR amplifications calculated by the intensity of the exon-including and the exon-excluding band and plotted as percentage. Each plot reports the mean percentage ± SEM of 5 biological replicates. **E)** Sashimi plots of exon skipping events in the *MYL6* transcripts in WT and ΔP hiPSC-derived CM (D20). The X-axis indicates genomic location while the Y-axis reports expression levels as Reads Per Kilobase of transcript per Million mapped reads (RPKM). The lines connecting different exons (bridges) indicate reads on the specific splicing junction. Quantification of splicing reads is reported in each bridge while Inclusion levels (IncLevel) are reported above each sample. The exonic structure of the event is shown below together with oligonucleotides position used for sqRT-PCR validation. **F)** Left panel: Representative electrophoresis analyses of sqRT-PCR for the exon skipping event in *MYL6* in WT vs ΔP and WT2 vs PA D20 CM. Right panel: Quantification of sqRT-PCR amplifications calculated by the intensity of the exon-including and the exon-excluding band and plotted as percentage. Each plot reports the mean percentage ± SEM of 5 biological replicates. **G)** Volcano plot depicts significant (FDR<0.05; Inclusion level variations stronger than 10%) ES events in murine *Charme* WT and KO post-natal (PN) hearts. X-axis represents exon inclusion ratio, y-axis represents –log10 of P-value. A schematic representation of the investigated event is shown above. **H)** Boxplots display the distribution of Intron Junction Counts (IJC) and Splice Junction Counts (SJC) for exon exclusion events (upper panel) and exon inclusion events (lower panel) following *pCHARME* depletion, along with their corresponding matched control events. Statistical significance was assessed using Mann-Whitney U-test. All the reported comparisons showed no significant differences (p-value > 0.05). **I)** Left panel: Percentile rank distribution of 7-mer frequencies in *pCHARME* compared to a reference set of 1,987 lncRNAs expressed in CM. Each point on the red line represents the rank percentile of an individual 7-mer based on its frequency in *pCHARME* relative to the lncRNA reference set. The x-axis denotes the percentile scale, while the y-axis refers to each of the possibile 7-mer. The light-gray shaded area represents the 10th–90th percentile range of the background distribution, and the dark-gray area highlights the interquartile range (25th–75th percentiles). The median (50th percentile) is marked by a black line. Right panel: Absolute frequency of each of the 16 *pCHARME*-specific 7-mers (red line) compared to their distribution observed in the background lncRNAs. Gray areas again denote the 10th–90th (light) and 25th–75th (dark) percentile bands, with the mean value across background lncRNAs shown in black. The sequence logo generated from the 16 *pCHARME*-specific 7-mers, illustrating their consensus UC-rich composition, is shown below. **J)** Heatmap showing key features of *pCHARME* sequence. The transcript is represented from the 5' end (left) to the 3' end (right) in consecutive 50-nucleotide segments with a 10-nucleotide sliding step. The top panel displays a bar plot reporting the total number of *pCHARME*-specific 7-mer occurrences within each segment. The heatmap below illustrates three distinct features along the transcript. The first is the GC content where segments with higher GC content are shown in violet, while those with lower GC content appear in yellow. The second feature is RNA accessibility, computed using RNAfold software and reported as the fraction of unpaired nucleotides in each segment; segments with higher accessibility are colored in red, and those with lower accessibility in blue. The third feature is the annotation of repetitive elements,

computed using RepeatMasker software, where each segment is marked in black to indicate the presence of the specified repeat elements, or in white if absent. **K)** Venn Diagram depicting the overlap among proteins predicted to interact with *pCHARME* sequence using RNAcompete, eCLIP crosslinking sites, catRAPID interact and catRAPID motif approaches (see **"Material and Methods"** for details).

Supplementary Fig. 6



Supplementary Fig. 6. A) Schematic representation of the CRISPR/Cas9 genome editing design used to obtain the *RHE* hiPSCs. **B)** RT-qPCR quantification of selected mRNA in WT2, RHE (-DOXY) and RHE (+DOXY) hiPSC-derived CM (D20). Data were normalized to *ATP5O* mRNA and represent relative expression means ($2^{\wedge -DCt}$) ± SEM of 5 biological replicates. Statistical tests were conducted on log_2-transformed fold change (LogFC) values compared to the control sample (RHE -DOXY). Data information: *p < 0.05; ***p < 0.01; ****p<0.0001; one-sample two-tailed Student's t-test against the null hypothesis of zero (no change).