The epigenetic modifier DOT1L regulates gene regulatory networks necessary for cardiac patterning and cardiomyocyte cell cycle withdrawal

Paola Cattaneo[*,1,2,3,4], Michael G. B. Hayes[5], Nina Baumgarten[3,4], Dennis Hecker[3,4], Sofia Peruzzo[3,4], Paolo Kunderfranco[6], Veronica Larcher[3,4], Lunfeng Zhang[1], Riccardo Contu[5], Gregory Fonseca[7], Simone Spinozzi[5], Ju Chen[5], Gianluigi Condorelli[2,6,8], Marcel H Schulz[3,4], Sven Heinz[5], Nuno Guimarães-Camboa[1,3,4], Sylvia M. Evans[*,1,5]

*corresponding authors

Affiliations:

[1] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, 92093, USA.

[2] Institute of Genetic and Biomedical Research (IRGB), Milan Unit, National Research Council of Italy, Milan 20138, Italy

[3] Institute of Cardiovascular Regeneration, Goethe University, Frankfurt 60590, Germany

[4] German Center for Cardiovascular Research (DZHK), Partner Site Rhein Main, Frankfurt am Main, 60590, Germany

[5] Department of Medicine, University of California San Diego, La Jolla, CA, 92093, USA.

[6] Humanitas Clinical and Research Center – IRCCS, Rozzano (MI), 20089, Italy

[7] Department of Medicine, Meakins-Christie Laboratories, McGill University, Montreal, Quebec, Canada.

[8] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (MI), 20090, Italy

Address Correspondence to:

Dr. Paola Cattaneo

paola.cattaneo@irgb.cnr.it   or   cattaneo@med.uni-frankfurt.de

Dr. Sylvia M. Evans

syevans@health.ucsd.edu

ABSTRACT

Mechanisms by which specific histone modifications regulate distinct gene regulatory networks remain little understood. We investigated how H3K79me2, a modification catalyzed by DOT1L and previously considered a general transcriptional activation mark, regulates gene expression in mammalian cardiogenesis. Early embryonic cardiomyocyte ablation of *Dot1l* revealed that H3K79me2 does not act as a general transcriptional activator, but rather regulates highly specific gene regulatory networks at two critical cardiogenic junctures: left ventricle patterning and postnatal cardiomyocyte cell cycle withdrawal. Mechanistic analyses revealed that H3K79me2 in two distinct domains, gene bodies and regulatory elements, synergized to promote expression of genes activated by DOT1L. Surprisingly, these analyses also revealed that H3K79me2 in specific regulatory elements contributed to silencing genes usually not expressed in cardiomyocytes. As DOT1L mutants had increased numbers of postnatal mononuclear cardiomyocytes and prolonged cardiomyocyte cell cycle activity, controlled inhibition of DOT1L might be a strategy to promote cardiac regeneration post-injury.

Keywords: Cardiogenesis, Epigenetics, Cardiac Patterning, Cardiomyocyte proliferation

Nonstandard Abbreviations and Acronyms: DOT1L - disruptor of telomeric silencing 1, H3K79me – methylation of Lysine 79 of Histone 3, H3K27ac - acetylation of Lysine 27 of Histone 3

INTRODUCTION

Epigenetic enzymes play critical roles in organogenesis by defining chromatin structures required for cell type-specific transcriptional networks[1]. Yet, mechanisms by which genome-wide histone modifications result in activation or repression of specific genes remain little understood. DOT1L is the only epigenetic enzyme catalyzing methylation of lysine 79 of histone 3 (H3K79me)[2,3]. Originally identified in yeast due to its role in maintenance of telomeric regions[2], DOT1L has been extensively studied in MLL-rearranged leukemia, where it is

considered an emerging therapeutic target[4]. Initial genome-wide studies suggested that, rather than being associated with specific gene expression programs, H3K79me2/3 is ubiquitously found in all transcribed loci, leading to a model of H3K79me as a generic mark of active genes[5].

Recent studies have revealed that, in addition to gene body H3K79me, DOT1L can regulate expression of its targets via methylation of regulatory regions[6]. We have previously shown that, *in vitro*, DOT1L is required for proper differentiation of embryonic stem cells into cardiomyocytes[7], however it is currently unclear whether this enzyme is necessary for normal cardiogenesis *in vivo*. In mice, global ablation of *Dot1l* results in embryonic lethality[8] and selective *Dot1l* ablation in cardiomyocytes results in an adult lethal phenotype with perturbations in dystrophin expression[9]. However, owing to the nature of this knockout (using a Cre that is not active during early embryogenesis), it is not known whether DOT1L plays any role in cardiogenesis. Furthermore, our molecular understanding of the functioning of this enzyme in cardiogenesis *in vivo* is limited by the absence of RNA-seq or ChIP-seq datasets that would allow for unbiased identification of direct targets of DOT1L in embryonic cardiomyocytes.

Cardiac patterning, the process allowing differentiation of the distinct cardiac chambers during embryogenesis, is controlled by the asymmetric expression of a well-defined set of transcription factors[10-13]. Amongst these, *Hand1* (expressed in the left ventricle) and *Hand2* (predominantly expressed in the right ventricle) play a major role in defining systemic and pulmonary ventricular identity, respectively[14-16]. Epigenetic mechanisms contributing to the tightly regulated expression of patterning genes remain mostly unexplored, however, it is known that the H3K4 methyltransferase SMYD1/BOP1 is a critical regulator of *Hand2* expression[17]. To date, it is not known whether any particular epigenetic enzyme plays a similar role in left ventricle-specific regulation of *Hand1* expression.

Regulated cardiomyocyte proliferation is a major component of normal cardiogenesis. The rate of cardiomyocyte mitosis gradually decreases from midgestational stages until birth, reaching a state of complete mitotic withdrawal in the first week after birth, when the majority of cardiomyocytes become binucleated[18]. Finding epigenetic enzymes that play a role in this process might allow the identification of routes to promote adult cardiomyocyte proliferation for therapeutic purposes.

To test whether DOT1L plays a role in cardiogenesis, we ablated a floxed *Dot1l* allele using a Cre allele that is active in cardiomyocytes from early developmental timepoints and

3

observed a perinatal lethal phenotype with abnormal cardiac morphology. This phenotype resulted from perturbations in highly specific gene expression programs that drive two critical cardiogenic events: cardiac patterning and cardiomyocyte cell cycle withdrawal. In embryonic development, DOT1L emerged as a major regulator of the expression of several transcription factors involved in cardiac patterning, with left ventricle-specific genes being particularly sensitive to the absence of this enzyme. In the neonatal period, DOT1L promoted expression of genes involved in cardiomyocyte maturation and cell cycle exit. Integration of cardiomyocyte ChIP-seq and Hi-C data revealed that DOT1L regulated target genes via two mechanisms: methylation of H3K79 in gene bodies and methylation of H3K79 in regulatory elements (K79-REs). Our analyses identified two types of K79-REs: activating elements that synergized with gene body H3K79me2 to promote expression of targets activated by DOT1L, and K79-REs that contributed to silencing of genes normally not expressed in cardiomyocytes. In contrast to H3K79me2 being a general activator of gene transcription as previously thought, our results reveal H3K79me2 as an epigenetic mark that regulates specific gene programs by both activating and repressing target genes. Moreover, the increased proliferation of postnatal cardiomyocytes following loss of *Dot1l* suggests the use of DOT1L inhibitors, already in development for treatment of leukemia[19,20], as agents to promote cardiac regeneration.

RESULTS

Abnormal cardiogenesis in DOT1L cKOs

Following our previous observation that DOT1L is required for the differentiation of embryonic stem cells into cardiomyocytes[7], we decided to assess if this enzyme is required for normal cardiogenesis *in vivo*. Fluorescent in situ hybridization studies showed that the *Dot1l* gene is robustly expressed in cardiomyocytes at least from E10.5 onwards (Supplementary Figure 1a). This expression was observed throughout the heart, without being specific to any chamber. *Dot1l* expression was also observed in multiple tissues outside the heart, showing this enzyme is not cardiac-specific. However, at E16.5, *Dot1l* expression levels were higher in cardiomyocytes than in the developing valves (Supplementary Figure 1b). *Dot1l* has been previously ablated in the heart using an *αMHC-Cre* line, leading to an adult lethal phenotype

4

with reduced dystrophin expression[9]. We have shown that *αMHC-Cre* is not optimal for myocardial-restricted knockouts due to its expression in non-myocyte lineages inside and outside the heart, and because it does not flox-out in all cardiomyocytes at an early time point[21]. To avoid these limitations and study cardiomyocyte-specific roles of DOT1L from early embryonic timepoints, we ablated a floxed *Dot1l* allele (loxP sites flanking exon 2) using the *xMlc2-Cre* allele[22] that drives highly specific and efficient recombination in cardiomyocytes (Supplementary Figure 1c) from the cardiac crescent stage[22]. In our analyses, we compared *xMlc2-Cre+; Dot1l-WT/flox* mice (herein designated as Dot1L Ctrl or controls) with *xMlc2-Cre+; Dot1l-Δ/flox* littermate mice (herein designated as Dot1L cKO or mutants). Inclusion of a copy of the *Rosa26-tdTomato* reporter allele[23] in all crosses allowed highly specific labeling of cardiomyocytes by the red fluorescent protein tdTomato for downstream flow cytometry and confocal microscopy applications (Supplementary Figure 1d). Real time qPCR analyses revealed highly efficient ablation of the floxed allele in sorted E12.5 cardiomyocytes (Figure 1a).

DOT1L is the only enzyme catalyzing H3K79 methylation[3]. Consequently, cardiomyocyte ablation of *Dot1l* resulted in a strong reduction in H3K79me2 levels in E14.5 hearts (Figure 1b and 1c). Cardiomyocyte-specific ablation of *Dot1l* did not result in an embryonic lethal phenotype, as animals with the mutant genotype were observed at expected Mendelian ratios in all stages analyzed (Supplementary Figure 1e). Despite being born at expected numbers, Dot1L cKO mice started dying shortly after birth, with 50% mortality by postnatal day 9 (P9), and the longest-lived mutant surviving until P16 (Figure 1d). Macroscopic examination of organs at neonatal stages revealed that mutants had enlarged hearts with a rounded shape relative to controls (Figure 1e). This phenotype was exacerbated as animals aged and peaked at P10 in surviving animals (Figure 1e). This evident cardiac enlargement translated into increased heart weight and increased heart weight/body weight ratios (Figure 1f) without changes in body weight (Supplementary Figure 1f). To determine the timing of onset of this phenotype and assess for additional morphogenic malformations, we conducted a histological time course analysis spanning from midgestation to postnatal timepoints (Figure 1g). cKO hearts were completely indistinguishable from control littermates until E14.5 and started exhibiting the enlarged, rounded phenotype between E16.5 and E18.5. Postnatally, mutant hearts displayed ventricular walls with increased thickness and areas of moderate persistent trabeculation (Figure 1g). No additional major morphogenic abnormalities (septal or outflow defects) were observed.

Immunostaining for VIMENTIN revealed absence of valve malformations and absence of major foci of fibrosis in cKOs (Figure 1g).

DOT1L is Essential for Cardiac Patterning

To decipher gene expression networks misregulated in the absence of DOT1L, we used RNA-seq to assess the transcriptomes of FACS-sorted Ctrl and cKO cardiomyocytes just prior to the onset of clear phenotypic changes (E16.5). An average of 100,000 cardiomyocytes were sorted per E16.5 heart (including atrial and ventricular chambers), yielding enough RNA to prepare biological replicates from individual hearts (3 replicates per genotype group). Bioinformatics analyses revealed that from a total of 12110 expressed genes (RPKM $\geq$ 1 in either Ctrl or cKO), 1478 were misregulated in cKO versus Ctrl cardiomyocytes (log2FC $\leq$ -0.5 or $\geq$0.5, false discovery rate $\leq$0.05). From these, 439 genes were downregulated and 1039 upregulated in cKOs (Figure 2a and Supplementary File 1a). Transcripts upregulated in cKOs corresponded to genes that are normally expressed at low levels in control cardiomyocytes (62% of genes in the lower expression quartile), whereas downregulated genes corresponded to genes normally expressed at a medium/high level (Figure 2b). Notably, the gene displaying the highest downregulation (-5.4 log2FC) in mutant cardiomyocytes was the transcription factor *Hand1,* a central element in the establishment of left ventricular identity (Figure 2c and Supplementary File 1a). In addition, amongst genes significantly downregulated in mutants there were several other genes involved in cardiac patterning: the transcription factors *Irx4* (ventricular specific[24]), *Tbx5* (left ventricle and atrial-specific[25])*, Gata4 and Mef2c*, the epigenetic enzyme *Smyd1* and the left ventricular-specific genes *Gja5* (encoding Connexin40) and *Cited1* (Figure 2c and Supplementary File 1a). The transcription factor *Nkx2-5* also exhibited a trend toward downregulation (Supplementary File 1a). These transcriptomic differences revealed that DOT1L activity in embryonic cardiomyocytes is essential for establishment of gene expression networks that coordinate normal cardiogenesis. This requirement, however, was not generalized to all cardiac patterning genes, as transcript levels for *Hand2*, *Tbx20* and *Nppa* (ANF) were not altered in mutants (Supplementary File 1a).

To confirm the hypothesis of abnormal cardiac patterning in Dot1L cKOs, we performed fluorescent RNA in situ hybridization (RNA-scope) in histological sections of control and cKO

hearts. Strikingly, at E16.5, *Hand1* transcripts were completely undetectable in the left ventricle of mutants (consistent with the 0 RPKM values observed in RNA-seq studies, (Figure 2d, bottom panel). Downregulation of *Hand1* was already evident at E10.5 when levels of *Hand1* expression in control hearts were higher than at E16.5 (Figure 2d, top and middle panels), indicating that transcriptional consequences of DOT1L ablation preceded the first phenotypic manifestations by several days. Similarly, transcript levels for the ventricular-specific transcription factor *Irx4* were significantly reduced in cKO hearts both at E10.5 and E16.5 (Figure 2e), confirming RNA-seq results.

Our data revealed a previously unrecognized role for DOT1L in regulating the expression of transcription factors involved in cardiac patterning, in particular those involved in left ventricular identity. While a number of genes were downregulated in cKOs, *Hand1* was particularly sensitive to absence of DOT1L, with *Hand1* transcripts being completely absent from midgestational DOT1L-deficient cardiomyocytes. Consistent with these transcriptomic observations, cardiac-specific ablation of *Hand1* phenocopies DOT1L cKO, including an enlarged and round-shaped heart and peri-natal lethality[16].

Transcriptional Control Via Gene Body H3K79me2

Several core cardiac transcription factors are differentially expressed in DOT1L cKO cardiomyocytes. Thus, genes modulated in our transcriptomic analyses likely include a combination of targets directly regulated by DOT1L as well as those indirectly regulated owing to secondary effects. To identify genes directly regulated by DOT1L, we performed ChIP-seq assays for H3K79me2 in E16.5 control and cKO FACS-sorted cardiomyocytes. These analyses revealed a total of 31,895 H3K79me2 peaks significantly enriched in control versus DOT1L cKO CMs (Supplementary File 2). The fact that the vast majority of H3K79me2 peaks in Ctrls were lost in cKOs is consistent with DOT1L being the sole histone K79 methyltransferase in these cells and validated the purity of our sorted CMs (Figure 3a and Supplementary File 2). Genome-wide differential peak distribution analysis revealed that H3K79me2 is mostly an intragenic modification (introns, exons, UTR and promoter/TSS) with only 2% of differential peaks being located in intergenic regions (Figure 3b).

Genes downregulated in cKOs exhibited, in control CMs, maximal levels of H3K79me2 immediately after the transcription start site (TSS) that progressively decreased towards the transcription termination site (TTS) (Figure 3c). Gene body H3K79me2 (coverage≥50 reads and fraction of gene body≥0.20 in control) was present in 341 downregulated genes (77% of all downregulated genes) and 98 upregulated genes (9% of all upregulated genes) (Figure 3d, 3e and Supplementary File 1). This observation suggests that a significant part of gene downregulation observed in cKOs can be directly attributed to H3K79me2 in the gene body, whereas gene upregulation does not seem to be directly associated with gene body H3K79me2. This is consistent with the view of gene body H3K79me2 as an activation mark[5], however, it should be noted that this modification is not an absolute requirement for transcription, as multiple genes were expressed in controls without having significant levels of gene body H3K79me2 (Figure 3f). Within genes showing gene body H3K79me2 there were two clear categories: genes dependent on DOT1L (modulated in cKOs) and genes not significantly affected by the loss of DOT1L, suggesting a scenario in which other epigenetic modifications can compensate for the loss of gene body H3K79me2 in a subset of genes. Notably, the majority of genes whose transcription depended on gene body H3K79me2 belonged to the highest quartiles of expression, suggesting this modification might play a role in supporting high transcriptional levels.

Transcriptional Control Via H3K79me2 in Regulatory Elements

Recent studies in leukemia showed that there is a distinct subset of enhancers that depend on H3K79me2[6]. We conducted bioinformatic analyses to investigate whether cardiomyocytes have H3K79me2-dependent regulatory elements (REs) that, together with gene body H3K79me2, contribute to gene expression differences observed in cKOs. To predict interactions between candidate regulatory elements and target genes, we applied the Activity-by-Contact (ABC) model[26]. By integrating information on chromatin state (ChIP-seq) and genomic conformation (Hi-C), this method is more accurate at predicting actual interactions between REs and their target genes than methods simply based on physical proximity[26]. Because the promoter of a gene can serve as a RE for a distinct cis gene[27,28], our analyses also considered interactions between promoters and distal genes. We used the ABC algorithm to merge our own H3K79me2 ChIP-seq datasets with publicly available datasets of E16.5 heart H3K27ac ChIP-seq, for the identification

of regulatory regions[29] (ENCODE project, ENCFF153DSU), and of cardiomyocyte Hi-C[30] (GSM2544836). To ensure we focused on meaningful RE-target gene interactions, in all downstream analyses we only considered the top 5% most relevant interactions identified by the ABC method: 471645 interactions, corresponding to 63958 H3K27ac peaks (Supplementary File 3). From these, 25% (15991 peaks) overlapped with a differential H3K79me2 peak (Supplementary File 3). This large number of cardiomyocyte REs enriched in H3K27ac and H3K79me2 (from here on designated as K79-REs) could provide an alternative mechanism for gene expression regulation by DOT1L. To assess the relative contribution of H3K79me2 in gene body versus H3K79me2 in REs to the regulation of target genes, we quantified the effect of these variables in the cumulative distribution of log2FC values. These analyses revealed that genes with higher fraction of H3K79me2 in the gene body were more downregulated in mutants than genes with lower (less than 33%) fraction of H3K79me2 in the gene body (Figure 3g and Supplementary File 4). Considering all genes, those that interacted with K79-REs displayed a trend to being downregulated in mutant cardiomyocytes, whereas genes that did not interact with these REs were more likely to be upregulated (Figure 3h and Supplementary File 4). The overall contribution of K79-REs to the relative expression of target genes seemed less evident than the effect of gene body H3K79me2 (Figure 3g,h). However, a stratified analysis segregating genes according to the number of associated K79-REs revealed that the magnitude of downregulation was relatively low if the gene had 1 to 3 K79-REs, but increased if the gene had 4 or more K79-REs (Figure 3i and Supplementary File 4). Focusing exclusively on loci with gene body H3K79me2, those that interacted with K79-REs were more downregulated in mutant cardiomyocytes than those without these elements, indicating that gene body and regulatory element H3K79me2 synergized to potentiate expression of target genes (Figure 3j and Supplementary File 4).

The vast majority of the genes downregulated in Dot1L-deficient cardiomyocytes were associated with both gene body H3K79me2 and at least one K79-RE (Figure 3k). Interestingly, 18% of all downregulated genes were associated with K79-REs in the absence of gene body H3K79me2. Overall, only 5% of downregulated genes were not associated with H3K79me2 in gene body or regulatory elements, indicating most gene downregulation was a direct consequence of DOT1L cKO. On the other side of the scale, upregulated genes had reduced association with gene body H3K79me2 (as mentioned above), but, surprisingly, 51% of

9

upregulated genes were associated with K79-REs in the absence of gene body H3K79me2. This observation suggests a novel role for DOT1L in gene silencing via H3K79me2-dependent regulatory elements.

Functional annotation revealed that genes downregulated in cKO with H3K79me2 in gene bodies and REs were involved in muscle differentiation and morphogenesis (Figure 3l). On the other hand, genes upregulated in cKOs and associated with inhibitory K79-REs were related to non-myocyte functions: extracellular matrix organization and skeletal system development. To search for cues as to how DOT1L achieves both gene activation and silencing via methylation of REs, we screened for enrichment in binding sites for known transcriptional regulators in the K79-REs of genes downregulated versus genes upregulated in DOT1L cKOs. This analysis revealed that K79-REs associated with genes downregulated in DOT1L cKOs (activating K79-REs) were enriched in binding sites for TFs of the Forkhead box (FOX) family, known regulators of multiple steps in mammalian cardiogenesis[31], as well as multiple cardiac-specific TFs (Figure 4c). On the other side of the scale, the most enriched element in K79-REs associated with genes upregulated in cKOs (silencing K79-REs) corresponded to the binding site of MBD2, a member of the NuRD complex known to play an important role in transcriptional silencing[32] (Figure 4d).

Focusing on the cardiac patterning genes downregulated in DOT1L cKOs it became evident that these were directly targeted by DOT1L, with the majority belonging to the group of loci that had gene body H3K79me2 combined with K79-REs. This category included not only the most downregulated gene, *Hand1*, but also the other aforementioned central players in cardiogenesis: *Irx4, Gata4*, *Tbx5*, and *Mef2c* (Figure 4a and 4b and Supplementary File 4). Interestingly, *Smyd1*, encoding the H3K4 methyltransferase necessary for right ventricular expression of *Hand2[17],* was also a direct target of DOT1L, unveiling complex epigenetic mechanisms for regulation of cardiac patterning. Our results suggest DOT1L has particular importance in left ventricle patterning, as it regulates not only transcription factors specifying the identity of this chamber (*Hand1* and *Tbx5*), but also their direct downstream targets *Cited1* and *Gja5* (Cx40)[16]. On the other hand, *Nppa* (ANF), another classic target of both HAND1[16] and TBX5[33] was not altered in DOT1L cKOs, further highlighting the specificity and complexity of this epigenetic control (Figure 4e).

Impaired cardiomyocyte cell cycle withdrawal in Dot1L mutants

Defective cardiac patterning can account for the abnormal morphology of Dot1L cKO hearts, however, it does not explain the increased wall thickness observed in postnatal mutant hearts (Figure 1g). As increased myocardial wall thickening coincides with the developmental period in which cardiomyocytes withdraw from cell cycle[18], we hypothesized this process might be affected in Dot1L cKOs. Flow cytometry analyses of cells isolated from P1 hearts revealed that, at this stage, about 50% of all cardiac cells were cardiomyocytes (labeled by expression of the reporter gene tdTomato, Figure 5a). Assessment of EdU incorporation rates (24 hours EdU pulse) revealed that, in control hearts, 19% of all cardiomyocytes were proliferative, whereas this number increased to 26% in mutant hearts (Figure 5a), representing a statistically significant increase in cardiomyocyte proliferation (Figure 5b). In mice, by P10 most cardiomyocytes are withdrawn from cell cycle and 80% of cardiomyocytes are binucleated[18]. Consistently, fluorescent microscopy analyses of cardiomyocytes isolated from P10 hearts revealed that less than 3% of control cardiomyocytes were EdU+ (24 hours EdU pulse, Figure 5c and 5d), compared to 10% in mutant hearts (Figure 5c and 5d). Nucleation analysis revealed that DOT1L cKO hearts had almost twice as many mononucleated cardiomyocytes (34% in cKOs versus 19% in controls) at the expense of binucleated cardiomyocytes (62% in cKOs versus 80% in Ctrls) (Figure 5e). Quantification of proliferation ratios across different nucleation categories (Figure 5f) revealed that Dot1L cKOs had a moderate increase in the percentage of EdU+ mononucleated cardiomyocytes (8.1% in cKOs versus 5.2% in Ctrls) and a significant 4.9-fold increase in the percentage of EdU+ binucleated cardiomyocytes (9.1% in cKOs versus 1.86% in controls).

To validate these analyses, we assessed proliferation of control and Dot1L cKO cardiomyocytes using the *Rosa26-Fucci2a* cell cycle reporter[34]. Expression from this allele is Cre-dependent (thus, restricted to cardiomyocytes in our analyses) and labels G1 cardiomyocytes in red, and actively proliferating cells in yellow (G1/S) or green (S/G2/M)[34]. Analysis of histological sections of P10 hearts (Figure 5g, 5h and Supplementary Figure 2a and 2d) revealed that in controls the majority of cardiomyocytes (67%) were in G1. G1/S cardiomyocytes could also be detected (32%), but S/G2/M cells were extremely rare (1%). On the other hand, DOT1L mutant cardiomyocytes showed significantly different cell cycle distribution across all cycling phases: 34% in G1, 61% in G1/S and 5% in S/G2/M (Figure 5g, 5h and Supplementary Figure 2a

11

and 2d). In addition, increased proliferation of mutant cardiomyocytes was also confirmed by phospho-Histone 3 (pH3) antibody staining in histological sections (Figure 5g and 5i and Supplementary Figure 2b and 2c), as well as EdU analysis (2 hours EdU pulse, Supplementary Figure 2d-g) of Dot1L mutants and controls in the *Rosa26-Fucci2a* background.

To gain insight into gene expression networks underlying the sustained proliferation of Dot1L cKO cardiomyocytes, we used neonatal (P1) FACS-sorted cardiomyocytes to perform mechanistic analyses similar to those described for E16.5: RNA-seq, H3K79me2 ChIP-seq, followed by bioinformatic analysis employing the ABC method to predict interactions between regulatory elements and their target genes (Supplementary Figure 3 and Supplementary Files 1-4). These analyses revealed that, similar to E16.5, the majority of genes downregulated in cKO cardiomyocytes were directly regulated by DOT1L via H3K79me2 both in gene body and REs (Figure 5j). On the other hand, the vast majority of genes upregulated in cKOs (92%) did not have gene body H3K79me2, however, 66% had interactions with K79-REs (Figure 5j). Functional annotation revealed that genes upregulated in cKOs and associated with K79-REs were enriched in genes involved in neuronal differentiation and vesicle-mediated transport in synapse, whereas genes downregulated in cKOs with H3K79me2 in gene body and REs were enriched in genes involved in muscle cell differentiation and myofibril organization (Figure 5k). Importantly, within this category of genes there was the gene encoding the cell cycle regulator p27 (*Cdkn1b*) (Figure 5l). Interestingly, p27 knockouts have a phenotype of increased heart size due to defective cardiomyocyte cell cycle arrest that resembles the one of Dot1L cKOs[35]. Altogether these observations suggest that in the neonatal period, DOT1L directly promotes expression of genes involved in cardiomyocyte maturation and cell cycle withdrawal whilst repressing expression of genes associated with neuronal functions.

DISCUSSION

Devising a blueprint for transcriptional and epigenetic mechanisms regulating normal heart development is of critical importance for our understanding of congenital heart disease and to pave the way for regenerative therapies of the adult heart post injury. Our previous studies using embryonic stem cells led to the hypothesis that DOT1L might play a relevant role in

cardiogenesis, which we confirmed with the *in vivo* experiments reported here. In addition, we deciphered novel mechanisms of action of this epigenetic modifier. *Dot1l* is ubiquitously expressed in the heart and its cardiac-specific conditional ablation resulted in a fully penetrant phenotype of an enlarged and rounded heart that culminated in neonatal lethality (Figure 1). Genome-wide transcriptomic and ChIP-seq assays provided detailed mechanistic insight into gene expression programs underlying this phenotype, highlighting DOT1L as a direct regulator of several genes involved in two processes that take place at distinct developmental stages: embryonic cardiac patterning (Figures 2, 3 and 4) and neonatal CM maturation and cell cycle withdrawal (Figures 5). Detailed bioinformatic analyses revealed that DOT1L regulated expression of its targets via methylation of H3K79 in gene body, regulatory regions, or a combination of both.

In embryonic cardiac patterning, DOT1L directly regulated several genes involved in defining distinct cardiac chambers. Amongst these, there was a clear enrichment in genes required for left ventricular identity (*Hand1*, *Tbx5*, *Cited1*, *Gja5*). From these, *Hand1* emerged as a gene particularly sensitive to absence of DOT1L activity. Likely reflecting this tight regulation, the dilated and rounded heart phenotype with neonatal lethality displayed by *Dot1l* mutants strongly resembles the phenotype of *Hand1* conditional mutants[16]. *Hand1* cardiac cKOs also display ventricular septal and outflow defects, whereas *Dot1l* mutants do not. This is likely a consequence of residual *Hand1* transcripts at earlier embryonic stages – *Hand1* transcripts were completely absent from E16.5 DOT1L cKO cardiomyocytes but could be detected at E10.5 (although at much lower levels than in stage-matched controls, Figure 2). The dependence of *Hand1* expression on DOT1L is of particular relevance for our understanding of epigenetic regulation of cardiogenesis. Previous studies have shown that SMYD1 (that catalyzes H3K4 methylation) is critical for activation of *Hand2*[17], but an epigenetic mechanism specifically regulating *Hand1* has remained elusive. Together with these previous observations, our results suggest a model in which DOT1L and SMYD1, two enzymes expressed throughout the heart, are essential for left/right asymmetric gene expression by serving as regulators of critical patterning transcription factors (Figure 4c).

DOT1L depletion resulted in absence of *Hand1* expression without affecting expression of *Hand2*, whereas absence of SMYD1 results in blunted activation of *Hand2* without affecting expression of *Hand1*[17]. DOT1L also regulated expression of *Smyd1* itself, revealing a complex

13

epigenetic mechanism controlling normal cardiogenesis. It is unclear if SMYD1 is involved in regulating *Dot1l* expression. Interestingly, despite the fact that *Smyd1* transcripts were significantly downregulated in Dot1L cKOs, its critical target *Hand2* was not altered in these mutants, suggesting that the levels of SMYD1 protein present in Dot1L cKOs are sufficient to sustain mechanisms of transcriptional regulation dependent on this enzyme. Despite divergent roles in the regulation of left/right genes, DOT1L and SMYD1 also share some common targets. For example, transcription of *Irx4* is dependent on the action of both enzymes (our own data and [17]). IRX4 is required for ventricular myocyte identity, and it is conceivable that DOT1L and SMYD1 act in a coordinated way to ensure a rheostatic control of *Irx4* transcription. How DOT1L and SMYD1, two enzymes that are expressed throughout the heart, achieve regulation of chamber-specific genes (ventricular versus atrial, left versus right) is not currently known and will be the subject of future studies. Given that several of DOT1L targets are also targeted by critical regulators of cardiogenesis, such as NKX2.5 and TBX5, it is possible that the specificity of these epigenetic enzymes is derived from their interactions with distinct transcription factors.

Initial studies suggested that H3K79 methylation is a generic marker of active transcription[5], but our datasets clearly revealed that, in cardiomyocytes, H3K79me2 methylation is not a basic requirement for gene transcription, as in control animals, multiple genes are active without having gene body H3K79 methylation. From the group of genes that bear this modification in controls and lose it in mutant hearts, two major categories emerge: those that are modulated upon DOT1L loss and those that are unaffected by the absence of this enzyme. The latter may reflect mechanisms of epigenetic redundancy. Our bioinformatic analyses integrating differential H3K79me2 ChIP-seq peaks with a mark of regulatory regions[29] (H3K27ac ChIP-seq) and 3-dimensional genomic conformation (cardiomyocyte Hi-C) suggested a model in which, in addition to gene body H3K79me2, DOT1L mediates expression of target genes via H3K79me2 in cis-regulatory elements. In loci with gene body H3K79me2, those associated with K79-REs were more downregulated in Dot1L cKOs. The magnitude of downregulation correlated with two variables: extent of H3K79me2 gene body coverage and the number of K79-REs interacting with the gene, suggesting these two forms of H3K79me2 synergize to promote expression of target genes. Interestingly, our analyses suggested that K79-REs could also be of an inhibitory nature, blocking transcription of target genes that lack gene body H3K79me2. To our knowledge

this is the first report of an involvement of DOT1L/H3K79me2 in this form of gene expression regulation.

In mice, after the first week of life, most cardiomyocytes are binucleated and completely withdrawn from the cell cycle[18]. Resistance of mature cardiomyocyte to proliferation is a major hurdle for cardiac regeneration[36,37]. Thus, devising strategies to promote reacquisition of proliferative potential in cardiomyocytes is a major priority in cardiac regenerative medicine[36,37]. FACS and histological studies using biochemical and genetic strategies to identify cardiomyocyte-specific proliferation revealed that Dot1L cKO cardiomyocytes retain proliferative potential at P10, a stage at which control counterparts are already withdrawn from the cell cycle (Figure 5). Notably, Dot1L cKO hearts also had a higher percentage of mononucleated cardiomyocytes than control hearts. Transcriptome analyses suggested that DOT1L depleted cardiomyocytes are less mature than control cardiomyocytes and fail to activate expression of p27, a known repressor of cell cycle. Permanent ablation of DOT1L results in severe consequences and lethality, but our results revealing sustained proliferation of DOT1L-depleted cardiomyocytes suggest that temporary inhibition of DOT1L in postnatal heart might be a strategy to promote re-acquisition of mitotic potential in cardiomyocytes. This prospect is particularly exciting considering that pharmacological inhibitors of DOT1L have been developed and are currently in clinical trials for leukemia[19,20], and will be explored in follow up studies.

METHODS

Mouse Strains and Experiments

Animal experiments were conducted according to protocols approved by the Institutional Animal Care and Use Committee at University of California, San Diego. All transgenic lines used were kept on an outbred background (Black-Swiss, Charles River laboratories). Mice were maintained in plastic cages with filtered air intake ports (Techniplast) on a 12 hr light cycle and fed Teklad LM-485 irradiated diet (Harlan Laboratories, catalog number 7912). For analyses conducted in embryonic stages, embryos were staged according to the embryonic day (E) on which dissection took place, with noon of the vaginal plug day being considered as E0.5 and birth typically occurring at E19. *Dot1l*flox mice were obtained from the KOMP Repository (CSD29070)

https://www.komp.org/geneinfo.php?geneid=54455. Floxed-out *Dot1l^A* mice were generated by crossing the *Dot1l*^flox^ allele with the epiblastic *Meox2-Cre* allele obtained from JAX laboratories (Stock No: 026858). *xMlc2-Cre* mice[22] were gently provided by Timothy Mohun. The *Rosa26-tdTomato (Ai9) (tdTom)* indicator allele[23] was purchased from JAX (Stock No: 007905). The Rosa26-*Fucci2A* cell cycle reporter allele was gently provided by Ian James Jackson[34]. All experiments were performed using littermate cKOs and Ctrls. Images shown are representative examples of experiments with n≥3 biological replicates. In experiments assessing proliferation, mice received an injection of EdU (at P1 25μL of a 3g/L stock and at P10 50μL of a 3g/L stock) 2h or 24h prior to euthanasia.

For genotyping, genomic DNA was extracted by adding 250μL of 50mM NaOH to a tail tip biopsy and heating at 95°C for 30 minutes. The solution was then neutralized by adding 50μL of 1M Trish-HCl (pH 8.0). Genotyping PCR reactions (36 amplification cycles) were performed using Taq DNA Polymerase with ThermoPol Buffer (New England Biolabs, M0267L), dNTPs from Promega (U1511) and 1μL of DNA solution. The following genotyping primers were used:
XMlc2Cre-Fw 5'-TAGGATGCTGAGAATCAAAATGT-3';
XMlc2Cre-Rev 5'-TCCCTGAACATGTCCATCAGGTTC-3';
Dot1L-Fw 5'-CCATATTAGTGTTCAAGGGCTACT-3';
Dot1L-fl/wt-Rev 5'-AGCATAAGGATGCCAACTACTAAC;
Dot1L-null-Rev 5'-AAGGAGGTCCTACTCATAGTCCTT-3';
Rosa26-tdTomato-Fw 5'-CTGTTCCTGTACGGCATGG-3';
Rosa26-tdTomato-Rev 5'-GGCATTAAAGCAGCGTATCC-3';
Rosa26-wt-Fw 5'-AAGGGAGCTGCAGTGGAGTA-3';
Rosa26-wt-Rev 5'-CCGAAAATCTGTGGGAAGTC-3';
Rosa26-wt-Fw 5'-CAAAGTCGCTCTGAGTTGTTATCAG-3';
Rosa26-wt-Rev 5'-GGAGCGGGAGAAATGGATATGAAG-3';
Rosa26-Fucci2a-Rev 5'-TCACCCAGGAGTCATTTGAT-3'.

Immunofluorescence

Tissues were isolated in cold PBS and fixed in 4% paraformaldehyde at 4°C overnight. Tissues were dehydrated in a sucrose gradient (5% -1 hour, 12% - 1 hour, 20% - overnight) and embedded in a 1:1 mix of 20% Sucrose and Tissue-Tek Optimal Cutting Temperature compound

(OCT, Sakura). 10 μm thick histological sections were cut using a cryostat (Leica CM3050) and kept at -20°C (short-term) or -80°C (long-term) until being processed for immunofluorescence. Sections or isolated cardiomyocytes were blocked in 10% donkey serum before incubation with antibodies. Primary antibodies were incubated at 4°C overnight. The following primary antibodies were used: anti-α Sarcomeric Actinin (1:400 Abcam #ab68167), anti-Vimentin (1:100 Abcam#ab45939), anti-TNNT (1:100 Thermo Fisher #MA5-12960), anti-PDGFR-α (1:200 R&D Systems #AF1062), anti-tdTomato (1:100 Sicgen #ab8181-200), anti-GFP (1:600 Abcam #ab13970), anti-phosphoH3 (1:100 Millipore #06-570). EdU incorporation was detected using a Click-iT® EdU kit (Thermo Fisher Scientific; #C10340). Immunofluorescence images were acquired using an Olympus FV1000 or a Leica SP8 scan confocal microscope. Image processing was performed with Fiji and Volocity software.

RNA scope

RNAscope fluorescent in situ hybridizations (ISH) were conducted using the RNA-scope Multiplex Fluorescent Reagent Kit v.2 (Advanced Cell Diagnostics, 323100) following standard protocols provided by the manufacturer, using the following RNA-scope probes (ACDbio): Mm-Dot1L-C2 (#533431-C2); Mm-Hand1-C2 (#429651-C2); Mm-Irx4 (#504831)-

Cardiomyocyte isolation

Embryonic and neonatal cardiomyocytes were isolated using a modified version of the a previously described protocol[38]. Briefly, embryonic or postnatal day 1 hearts were isolated and transferred into ice cold HBSS. Embryonic single cell suspensions from were obtained by performing eight rounds of enzymatic digestion (5 minutes each) with Collagenase type II (0.7mg/ml Worthington) at 37°C under agitation. Postnatal day 1 single cell suspensions were obtained by performing an overnight digestion with trypsin (0.5mg/ml) at 4°C followed by eight rounds of enzymatic digestion (5 minutes each) with Collagenase type II (1mg/ml Worthington) at 37°C. Cells were collected in cold medium containing fetal calf serum to stop the enzymatic reaction and centrifuged at 110 rcf to allow initial separation of cardiomyocytes from other cardiac cells. Cell preps were resuspended in 500μL of FACS buffer (HBSS, 5%FBS, 2.5mM EDTA) and 1μL of DAPI was added immediately prior to flow cytometry. Embryonic and postnatal day 1 live (DAPI negative) control and mutant cardiomyocytes were sorted based on

the red fluorescence emitted by the Cre reporter tdTomato using an Influx Cell sorter (BD Biosciences) and collected in TRIZol reagent for RNA extraction or cross-linked as described below for chromatin analysis.

Postnatal day 10 cardiomyocytes were isolated using a Langerdoff system using Collagenase type II (0.7mg/ml Worthington). After perfusion cells were dissociated from the heart, collected in conical tubes and allowed to settle by gravity in order to obtain separation of viable, rod-shaped cardiomyocytes from dead cardiomyocytes and from interstitial cells of the heart. For histological analysis, isolated cardiomyocytes were fixed in 4%PFA and processed for immunostaining.

Protein isolation and Western Blot analysis

Total protein extracts were prepared by lysing samples in RIPA buffer. Protein lysates in Laemmli buffer were separated by electrophoresis on 12% SDS-PAGE gels and transferred for 2 hours at 4°C on to a PVDF membrane (BioRad). After blocking for an hour in 5% dry milk, membranes were incubated overnight at 4°C with the primary antibody in blocking buffer. The following primary antibodies were used: H3 dimethyl Lys79 (abcam #ab3594) and anti-H3 (abcam #ab1791). Blots were washed and incubated with a horseradish peroxidase (HRP)-conjugated secondary antibody generated in Rabbit (1:2000; Cell Signaling Technology #7074) for 1.5 hours at room temperature. Immunoreactive protein bands were visualized using an enhanced chemiluminescence reagent (Thermo Fisher Scientific). Protein quantification was achieved using ImageJ software.

Quantification of proliferation by Flow Cytometry

FACS quantification of rates of EdU incorporation was carried out using littermate Dot1L cKOs and Ctrls. EdU detection was done in cell suspensions using the Click-iT® EdU Alexa 647 kit (Thermo Fisher Scientific; C10340), according to the manufacturer's instructions. TdTomato signal was used to discriminate cardiomyocytes from other lineages of cardiac cells. Stained cells were analyzed using a FACS Canto II flow cytometer (BD Bioscience). DIVA and FlowJo software (BD Pharmingen) were used for data acquisition and analysis.

RNA extraction, qRT-PCR and RNA-seq.

RNA was extracted using Tryzol (Invitrogen #15596026) and Direct-zol RNA Kits (Zymo Research #R2061) following instructions provided by the manufacturers. All transcriptomics analyses (qPCR or RNA-seq) were performed on FACS-sorted cardiomyocytes. For E16.5 and P1, biological replicates were prepared from single hearts. For E12.5, preparation of each biological replicate required pooling of hearts with same genotypes. For each stage analyzed, a minimum of 3 biological replicates, prepared from littermate Dot1L Ctrl and Dot1L cKO hearts, were used. For qRT-PCR experiments, cDNA was produced using the SuperScript VILO cDNA Synthesis Kit (Invitrogen #11754050). qRT-PCR was performed using SYBR Select Master Mix for CFX (Applied biosystems #4472952) on a Bio-Rad CFX96 Real-Time PCR system using the following primers: DOT1L-Ex2-Fw 5'-TGCTGCTCATGAGATTATTGAGA-3' (primer hybridizing within the loxP-flanked exon2 of the floxed *Dot1l* allele); DOT1L-Ex4-Rev 5'-ATGGCCCGGTTGTATTTGTC-3'; 18s-Fw 5'-AAATCAGTTATGGTTCCTTTGGTC-3'; 18s-Rev 5'-GCTCTAGAATTACCACAGTTATCCAA-3'

For RNA-seq experiments libraries were generated from 25ng of RNA using the TruSeq Stranded mRNA library Prep kit (Illumina# 20020594) and sequenced on a HiSeq 4000 System (Illumina) using a single read 50 protocol.

ChIP-seq

ChIP-seq was essentially performed as described[39]. Briefly, cells were fixed in in 1% formaldehyde/PBS for 10 minutes at room temperature, the reactions quenched by adding 2.625 M glycine to 125 mM final 20% BSA to 0.5% final and cells pelleted by centrifugation for 5 minutes at 1,000 g, 4°C. Cells were washed twice with ice-cold 0.5% BSA/PBS and cell pellets were snap-frozen in liquid nitrogen and stored at −80°C. Fixed cells were thawed on ice, resuspended in 500 µl ice-cold buffer L2 (0.5% Empigen BB, 1% SDS, 50 mM Tris/HCl pH 7.5, 1 mM EDTA, 1 × protease inhibitor cocktail) and chromatin was sheared to an average DNA size of 300–500 bp by administering 7 pulses of 10 s duration at 13 W power output with 30 s pause on wet ice using a Misonix 3000 sonicator. The lysate was diluted 2.5-fold with ice-cold L2 dilution buffer (20 mM Tris/HCl pH 7.4@20°C, 100 mM NaCl, 0.5% Triton X-100, 2 mM EDTA, 1 × protease inhibitor cocktail), and one percent of the lysate was kept as ChIP input. For each immunoprecipitation, aliquots of diluted lysate equivalent to 150,000 to 1 million cells, 20

μl of Dynabeads Protein A (for rabbit polyclonal antibodies) and 2 μg anti H3K79me2 antibody (Abcam #ab3594) were combined and rotated overnight at 8 RPM and 4°C. The following day, beads were collected on a magnet and washed three times each with wash buffer I (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.1% SDS, 2 mM EDTA), wash buffer III (10 mM Tris/HCl pH 7.5, 250 mM LiCl, 1% IGEPAL CA-630, 0.7% Deoxycholate, 1 mM EDTA) and twice with ice-cold TET (10 mM Tris/HCl pH7.5, 1 mM EDTA, 0.2% Tween-20). Libraries were prepared directly on the antibody/chromatin-bound beads: after the last TET wash, beads were suspended in 25 μl TT (10 mM Tris/HCl pH7.5, 0.05% Tween-20), and libraries were prepared using NEBNext Ultra II reagents according to the manufacturer's protocol but with reagent volumes reduced by half. DNA was eluted and crosslinks reversed by adding 4 μl 10% SDS, 4.5 μl 5 M NaCl, 3 μl EDTA, 1 μl proteinase K (20 mg/ml), 20 μl water, incubating for 1 h at 55°C, then overnight at 65°C. DNA was cleaned up by adding 2 μl SpeedBeads 3 EDAC in 61 μl of 20% PEG 8000/1.5 M NaCl, mixing and incubating for 10 minutes at room temperature. SpeedBeads were collected on a magnet, washed twice by adding 150 μl 80% EtOH for 30 s each, collecting beads and aspirating the supernatant. After air-drying the SpeedBeads, DNA was eluted in 25 μl TT and the DNA contained in the eluate was amplified for 12 cycles in 50 μl PCR reactions using NEBNext High-Fidelity 2X PCR Master Mix or NEBNext Ultra II PCR master mix and 0.5 μM each of primers Solexa 1GA and Solexa 1GB. Libraries were cleaned up as above by adding 36.5 μl 20% PEG 8000/2.5 M NaCl and 2 μl Speedbeads, two washes with 150 μl 80% EtOH for 30 s each, air-drying beads and eluting the DNA into 20 μl TT. ChIP library size distributions were estimated following 2% agarose/TBE gel electrophoresis of 2 μl library, and library DNA amounts measured using a Qubit HS dsDNA kit on a Qubit fluorometer. ChIP input material (1 percent of sheared DNA) was treated with RNase for 15 min at 37°C in EB buffer (10 mM Tris pH 8, 0.5% SDS, 5 mM EDTA, 280 mM NaCl), then digested with Proteinase K for 1 h at 55°C and crosslinks reversed at 65°C for 30 min to overnight. DNA was cleaned up with 2 μl SpeedBeads 3 EDAC in 61 μl of 20% PEG 8000/1.5 M NaCl and washed with 80% ethanol as described above. DNA was eluted from the magnetic beads with 25 μl of TT and library prep and amplification were performed as described for ChIP samples. Libraries were sequenced single-end for 75 cycles (SE75) to a depth of 20-25 million reads on an Illumina NextSeq 500 instrument.

Bioinformatic analyses

RNA-seq

Sequencing reads were processed to remove Illumina barcodes and aligned to the UCSC *Mus musculus* reference genome (build mm10) using STAR v.2.5.1b with default parameters[40]. ReadsPerGene.out.tab files were then processed with edgeR[41]. RNA expression was calculated in reads per kilobase per million mapped reads (RPKM) considering the sum of exon length. Differentially expressed coding genes were selected based on the following parameters: FDR ≤0.05, RPKM ≥1 in at least one biological condition, and log2 Fold Change ≤ -0.5 for genes downregulated and ≥ 0.5 for genes upregulated in Dot1L cKOs vs Dot1L Ctrls. Pathway analysis was performed using METASCAPE[42].

ChIP-seq alignment and peak-calling

All analyses were performed using the mouse reference genome GRCm38 (mm10) and the gencode gene annotation version vM25. Bowtie2[43] was applied to align the fastq files to the mouse reference genome. First, the required index structure was built using: *bowtie2-build -f --seed 123 --threads 20 Mus_musculus.GRCm38.dna.primary_assembly.fa mouse_GRCM38_mm10.* Since the reads where not paired, we ran for each fastq file: *bowtie2 -x mouse_GRCM38_mm10 -U <fastq-file> -S <output-file-name>.sam -q -t -p 30.* Conversion of the resulting files (sam) to bam format was done using samtools (version samtools 1.10)[44]. To allow easy visualization in a genome browser, bam files were converted to bigWig (bw) files using deeptools bamCoverage function. Next peak-calling was performed with MACS2 (version macs2 2.2.7.1)[45]: *macs2 callpeak -t <treatment>.bam -c <input>.bam -n <prefix-name> --outdir <output-dir> -f BAM -g 1.87e9 -B,* where <treatment> is either the aligned reads of the Dot1L Ctrl or Dot1L cKO and <input> the corresponding input signal. For all following analyses we used the narrowPeak files.

To compute differential H3K79me2 peaks between Ctrl and Dot1L cKOs, we applied DiffBind (version 2.10.0 bioconductor-diffbind)[46] using the resulting narrow peak files from the peak calling of Ctrl and cKO cardiomyocytes (both replicates) and the corresponding bam files of the H3K79me2 ChIP-seq and input.

Determining genomic distribution of H3K79me2 peaks

To determine the number of H3K79me2 peaks overlapping with distinct genomic regions, the gene annotation file was split into exon, UTR and gene regions. The most 5' transcription start site (TSS) for all mouse genes was downloaded from biomart[47] and a promoter region of length 400 bp, centered around the most 5' TSS, was determined. The number of H3K79me2 peaks overlapping either with exon, UTR or promoter regions was computed using the intersect function (used flags -f 0.1 -F 0.1 -e) from bedtools (v2.25.0)[48]. Subsequently, based on this information, peaks overlapping with introns and intergenic regions were deduced. Peaks were assigned as introns when they overlapped a gene but not an exon. Peaks that did not intersect with any of the aforementioned regions (exons, UTRs, promoters and introns) were classified as intergenic. Results from these analyses were plotted in the diagrams shown in Figure 3b (corresponding to E16.5) and Supplementary Figure 3b (corresponding to P1).

Computing coverage and fraction of the gene body covered by H3K79me2

The profiles of the H3K79me2 signal at the gene body were computed using deeptools[49]. We used the .bam files resulting from the bowtie2 analysis and the up- and downregulated genes in bed-file format (with strand information).

First, we used deeptools bamCompare to determine the mean signal of the two replicates of Ctrl and cKO for each time point: *bamCompare -b1 <treatment_rep1>.bam -b2 <treatment_rep2>.bam -o <treatment>_mean_rep1_rep2.bw -of bigwig --scaleFactorsMethod None --operation mean --effectiveGenomeSize 2652783500 -p 25 --normalizeUsing RPKM -- binSize 20 --skipNonCoveredRegions*, where <treatment> is either Ctrl or cKO.

Next, we visualized the data using deeptools computeMatrix and plotProfile functionalities: *computeMatrix scale-regions -S Ctrl_mean_rep1_rep2.bw cKO_mean_rep1_rep2.bw -R downregulated_genes_FDR_0.05.bed upregulated_genes_FDR_0.05.bed   -o inputProfile.mat.gz --endLabel TTS --beforeRegionStartLength 2000 --afterRegionStartLength 2000 --regionBodyLength 5000 -p 20 --skipZeros*

plotProfile -m inputProfile.mat.gz -out profile.pdf --perGroup --startLabel TSS --endLabel TTS --plotTitle "Average distribution of H3K79me2 relative to the distance from TSS and TTS" --samplesLabel "Ctrl (upregulated genes)" "cKO (upregulated)" --regionsLabel "downregulated genes" "upregulated genes" --plotFileFormat pdf

To compute the fraction of the gene body which is covered by H3K79m2, we applied bedtools (bedtools v2.25.0) coverage function. Given a set of genomic regions and a .bam file, the function computes per genomic region the number of reads that overlap and the fraction of bases that have a non-zero coverage based on the .bam file. We determined the number of overlapping reads and the fraction for all annotated mouse genes for Ctrl and cKO. Next, the mean fraction (of rep1 and rep2) of the up- and downregulated genes was computed. The density plots are based on all up- and downregulated genes.

Calling regulatory interactions with H3K27ac ChIP-seq and Hi-C and differential H3K79me2 enhancer analysis

For the identification of enhancer regions H3K27ac ChIP-seq data from C57BL/6 heart tissue was downloaded from ENCODE for both timepoints E16.5 (ENCFF153DSU) and P1 (ENCFF463XJT). A Hi-C matrix of mouse cardiomyocytes was downloaded from GEO (GSM2544836)[30]. The Hi-C matrix was normalized using Knight-Ruiz normalization with the Juicebox dump command (version 1.22.01)[50] with a resolution of 5000 bp for each chromosome. *java -jar juicer_tools_1.22.01.jar dump observed KR <hic-file> chr$ chr$ BP 5000 <output-file-name>*

The regulatory interactions between promoter and candidate enhancers were assessed with a self-implemented version of the ABC-score of Fulco et al.[26]. For all genes on the autosomes (GRCm38p6) the most 5' TSS was taken and all candidate enhancers in a 5Mb window centered on the TSS were scored according to the following equation:

$$ABCscore_{E,G} = \frac{A_E \times C_{E,G}}{\sum\limits_{\text{all } e \text{ within 5 Mb window}} A_e \times C_{e,G}}$$

The interaction of an enhancer (E) with a gene (G) is described by the activity of the enhancer (A_E) and the contact to the gene ($C_{E,G}$). The ABC-score returns the relative contribution of an interaction in relation to all other interactions of that gene. The enhancer activity is the signal of the H3K27ac ChIP-seq peak and is scaled so that the maximum value for all enhancers within a gene window is 100. The contact ($C_{E,G}$) between an enhancer and a gene's TSS is the contact frequency of the respective bins in the normalized Hi-C matrix and is also scaled to a maximum

23

of 100 for all enhancers within the window. All scored candidate enhancer-gene interactions were ranked according to the ABC-score and the top 5% were used for further analyses. Chromosomes X and Y where excluded from the analysis. An enhancer was considered to be differential for H3K79me2 if ≥ 10% of the enhancer's length was covered by a differential H3K79me2 peak. Statistical differences between the cumulative log2(fold-change of gene expression) distributions were calculated pairwise with a two-tailed Kolmogorov-Smirnov test (SciPy 1.4.1, stats module, ks_2samp function).

Motif enrichment analysis

To identify transcription factors likely regulating the activity of K79-REs associated to genes differentially expressed in Dot1L cKOs, a motif enrichment analysis was performed. To this end, REs overlapping with a H3K79me2 peak and associated to differentially expressed genes were divided into two groups: those associated with genes downregulated in Dot1L cKOs (activating K79-REs, N=1355) and those associated with genes upregulated in Dot1L cKOs (silencing K79-REs, N=1047). TF binding motifs (total of 515) were downloaded from the JASPAR database[51] and only those significantly expressed (RPKM ≥ 1 in our RNA-seq dataset) in E16.5 cardiomyocytes (a total of 367 TFs) were considered in subsequent analyses. Using TRAP[52], a TF-affinity value per TF for each RE sequence was computed. The value is defined as the sum over all binding site probabilities of the given TF for the current sequence. The TRAP analysis was performed separately for the RE sequences of the up- and downregulated genes. A one-tailed Mann-Whitney test (using R's Wilcox test function, confidence level of 0.975) to identify TFs enriched in the K79-REs of upregulated genes versus K79-REs of downregulated genes and vice versa was performed. The resulting p-values were adjusted by applying the Benjamini–Hochberg procedure. All TFs with an adjusted p-value of 0.05 were considered significant.

Quantification Experiments and Statistical Analysis

Statistical significance of differences in the survival of *Dot1L* mice was assessed using Kaplan-Meier survival analysis with the log-rank method of statistics. In all other graphs, data are expressed as mean ± SEM, with a minimum of 3 biological replicates (the exact replicate number is described in the legend to each figure). Statistical significance of differences among groups

was tested by 2-tailed Student's t-test. A value of P ≤0.05 was considered statistically significant. * P≤0.05, **P≤0.01. Analyses were performed with GraphPad Prism software.

Data Availability

All RNA-seq and ChIP-seq data that support the finding of this study have been deposited in the Gene Expression Omnibus (GEO) under accession number code GSE184192. Previously published ChIP-seq and Hi-C data that were re-analyzed here are available under accession codes ENCFF153DSU and ENCFF463XJT from ENCODE and GSM2544836 from GEO. Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Sources of Funding

Acknowledgments

Disclosures

None to declare.

Contributions

References:

1   Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-432, doi:10.1038/nature05918 (2007).

2   Singer, M. S. *et al.* Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae. *Genetics* **150**, 613-632 (1998).

3   Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* **12**, 1052-1058 (2002).

4   McLean, C. M., Karemaker, I. D. & van Leeuwen, F. The emerging roles of DOT1L in leukemia and normal development. *Leukemia* **28**, 2131-2138, doi:10.1038/leu.2014.169 (2014).

5   Steger, D. J. *et al.* DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol* **28**, 2825-2839, doi:10.1128/MCB.02076-07 (2008).

6   Godfrey, L. *et al.* DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat Commun* **10**, 2803, doi:10.1038/s41467-019-10844-3 (2019).

7   Cattaneo, P. *et al.* DOT1L-mediated H3K79me2 modification critically regulates gene expression during cardiomyocyte differentiation. *Cell Death Differ* **23**, 555-564, doi:10.1038/cdd.2014.199 (2016).

8   Jones, B. *et al.* The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS Genet* **4**, e1000190, doi:10.1371/journal.pgen.1000190 (2008).

9   Nguyen, A. T. *et al.* DOT1L regulates dystrophin expression and is critical for cardiac function. *Genes Dev* **25**, 263-274, doi:10.1101/gad.2018511 (2011).

10   Paige, S. L., Plonowska, K., Xu, A. & Wu, S. M. Molecular regulation of cardiomyocyte differentiation. *Circ Res* **116**, 341-353, doi:10.1161/CIRCRESAHA.116.302752 (2015).

11   Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922-1927, doi:10.1126/science.1132292 (2006).

12   Harvey, R. P. Patterning the vertebrate heart. *Nat Rev Genet* **3**, 544-556, doi:10.1038/nrg843 (2002).

13   Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* **126**, 1037-1048, doi:10.1016/j.cell.2006.09.003 (2006).

14   Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* **16**, 154-160, doi:10.1038/ng0697-154 (1997).

15   Thomas, T., Yamagishi, H., Overbeek, P. A., Olson, E. N. & Srivastava, D. The bHLH factors, dHAND and eHAND, specify pulmonary and systemic cardiac ventricles independent of left-right sidedness. *Dev Biol* **196**, 228-236, doi:10.1006/dbio.1998.8849 (1998).

16   McFadden, D. G. *et al.* The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development* **132**, 189-201, doi:10.1242/dev.01562 (2005).

17   Gottlieb, P. D. *et al.* Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat Genet* **31**, 25-32, doi:10.1038/ng866 (2002).

18   Soonpaa, M. H., Kim, K. K., Pajak, L., Franklin, M. & Field, L. J. Cardiomyocyte DNA synthesis and binucleation during murine development. *Am J Physiol* **271**, H2183-2189, doi:10.1152/ajpheart.1996.271.5.H2183 (1996).

19   Stein, E. M. *et al.* The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood* **131**, 2661-2669, doi:10.1182/blood-2017-12-818948 (2018).

20   Dickins, R. A. Rerouting DOT1L inhibitors in leukemia. *Blood* **136**, 1900-1901, doi:10.1182/blood.2020007352 (2020).

21   Guimaraes-Camboa, N. *et al.* HIF1alpha Represses Cell Stress Pathways to Allow Proliferation of Hypoxic Fetal Cardiomyocytes. *Dev Cell* **33**, 507-521, doi:10.1016/j.devcel.2015.04.021 (2015).

22   Breckenridge, R., Kotecha, S., Towers, N., Bennett, M. & Mohun, T. Pan-myocardial expression of Cre recombinase throughout mouse development. *Genesis* **45**, 135-144, doi:10.1002/dvg.20275 (2007).

23   Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133-140, doi:10.1038/nn.2467 (2010).

24   Bao, Z. Z., Bruneau, B. G., Seidman, J. G., Seidman, C. E. & Cepko, C. L. Regulation of chamber-specific gene expression in the developing heart by Irx4. *Science* **283**, 1161-1164, doi:10.1126/science.283.5405.1161 (1999).

25   Bruneau, B. G. *et al.* Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. *Dev Biol* **211**, 100-108, doi:10.1006/dbio.1999.9298 (1999).

26   Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669, doi:10.1038/s41588-019-0538-0 (2019).

27   Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98, doi:10.1016/j.cell.2011.12.014 (2012).

28   Dao, L. T. M. *et al.* Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat Genet* **49**, 1073-1081, doi:10.1038/ng.3884 (2017).

29   Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

30   Rosa-Garrido, M. *et al.* High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure. *Circulation* **136**, 1613-1625, doi:10.1161/CIRCULATIONAHA.117.029430 (2017).

31   Zhu, H. Forkhead box transcription factors in embryonic heart development and congenital heart disease. *Life Sci* **144**, 194-201, doi:10.1016/j.lfs.2015.12.001 (2016).

32   Lai, A. Y. & Wade, P. A. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer* **11**, 588-596, doi:10.1038/nrc3091 (2011).

33   Bruneau, B. G. *et al.* A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**, 709-721, doi:10.1016/s0092-8674(01)00493-7 (2001).

34   Mort, R. L. *et al.* Fucci2a: a bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice. *Cell Cycle* **13**, 2681-2696, doi:10.4161/15384101.2015.945381 (2014).

35   Poolman, R. A., Li, J. M., Durand, B. & Brooks, G. Altered expression of cell cycle proteins and prolonged duration of cardiac myocyte hyperplasia in p27KIP1 knockout mice. *Circ Res* **85**, 117-127, doi:10.1161/01.res.85.2.117 (1999).

36   Guo, Y. & Pu, W. T. Cardiomyocyte Maturation: New Phase in Development. *Circ Res* **126**, 1086-1106, doi:10.1161/CIRCRESAHA.119.315862 (2020).

37   Eschenhagen, T. *et al.* Cardiomyocyte Regeneration: A Consensus Statement. *Circulation* **136**, 680-686, doi:10.1161/CIRCULATIONAHA.117.029343 (2017).

38   Louch, W. E., Sheehan, K. A. & Wolska, B. M. Methods in cardiomyocyte isolation, culture, and gene transfer. *J Mol Cell Cardiol* **51**, 288-298, doi:10.1016/j.yjmcc.2011.06.012 (2011).

39   Heinz, S. *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**, 1522-1536.e1522, doi:10.1016/j.cell.2018.07.047 (2018).

40   Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

41   Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

42   Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).

43   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

44   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

45   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

46   Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393, doi:10.1038/nature10730 (2012).

47   Hunt, S. E. *et al.* Ensembl variation resources. *Database (Oxford)* **2018**, doi:10.1093/database/bay119 (2018).

48   Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

49   Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

50   Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).

51   Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).

52   Sobreira, T. J., Durham, A. M. & Gruber, A. TRAP: automated classification, quantification and annotation of tandemly repeated sequences. *Bioinformatics* **22**, 361-362, doi:10.1093/bioinformatics/bti809 (2006).

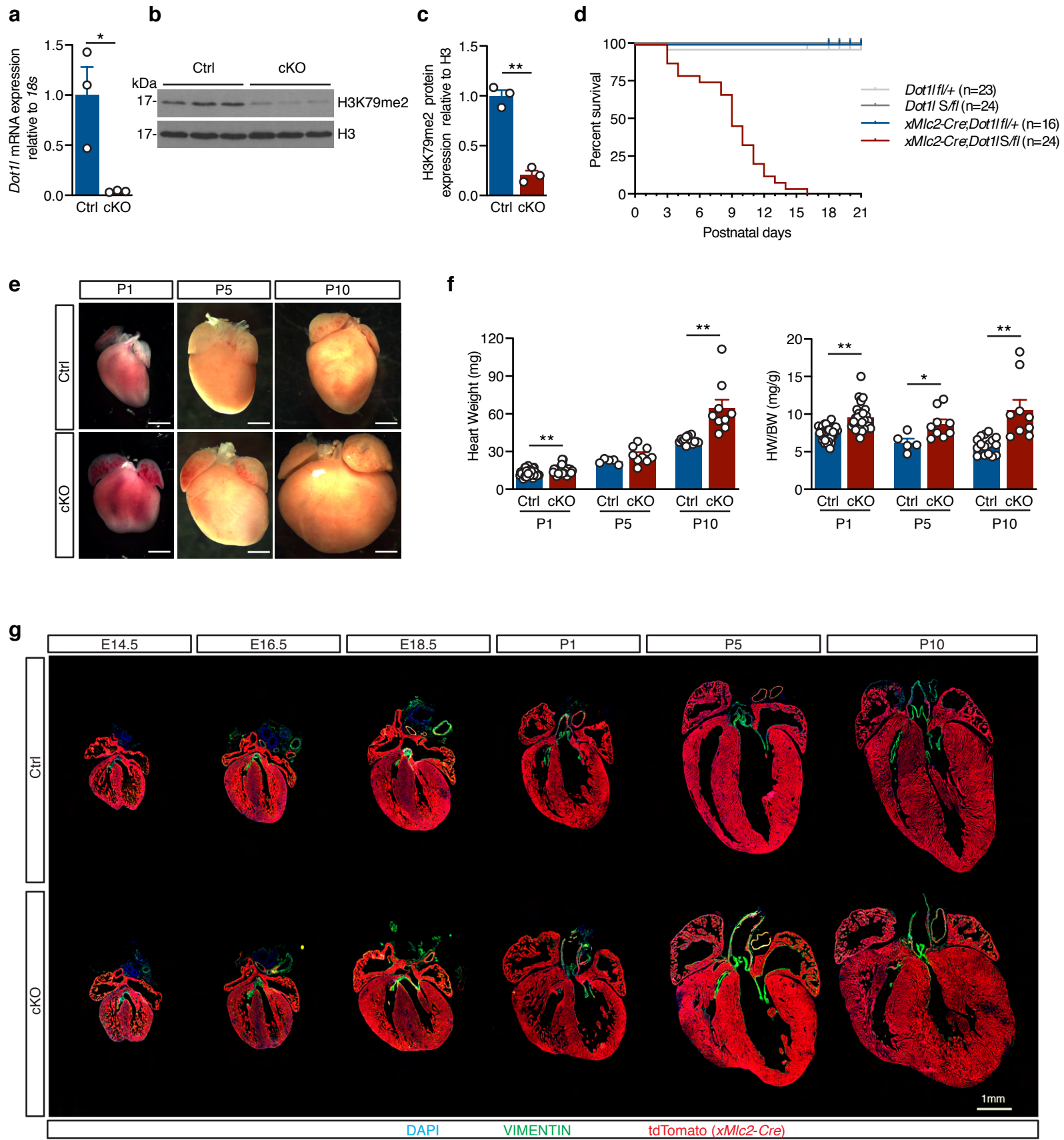**Figure 1 - Cattaneo et al.**

DAPI VIMENTIN tdTomato (*xMlc2-Cre*)

Figure 1: Cardiomyocyte-specific ablation of DOT1L from early developmental timepoints results in enlarged hearts and peri-natal lethality. a) qPCR analysis using a primer within the floxed exon of *Dot1l* mRNA showing efficient ablation of this gene in E12.5 cKO cardiomyocytes (N=3 biological replicates). **b-c)** Western blot (b) and respective quantification (c) showing strongly reduced H3K79me2 levels in E14.5 hearts upon ablation of Dot1L (N=3 biological replicates). **d)** Kaplan-Meier survival curves showing postnatal lethality of Dot1L cKOs. **e)** Whole mount images of postnatal day (P) 1, P5 and P10 hearts in Ctrls and Dot1L cKOs representative of the enlarged heart phenotype of Dot1L cKOs (scale bars = 1mm). **f)** Graphs representing a significant increase in heart weight (mg) and heart weight/body weight ratio (HW/BW (mg/g)) in Dot1L cKO vs Ctrl in all stages analyzed. **g)** Immunofluorescence time course depicting the dynamics of phenotypic manifestations in Dot1L cKOs in embryonic (E) and postnatal (P) stages. DAPI (4',6-diamidin-2-phenylindol) in blue, VIMENTIN in green and lineage traced XMLC2-Cre;tdTomato cardiomyocytes in red (scale bar = 1mm). In all graphs Ctrl indicates control mice (*XMlc2-Cre;Dot1L fl/+),* cKO indicates mutant mice (*XMlc2-Cre;Dot1L Δ/fl*). Data is presented as mean ± SEM; * P≤0.05, **P≤0.01.
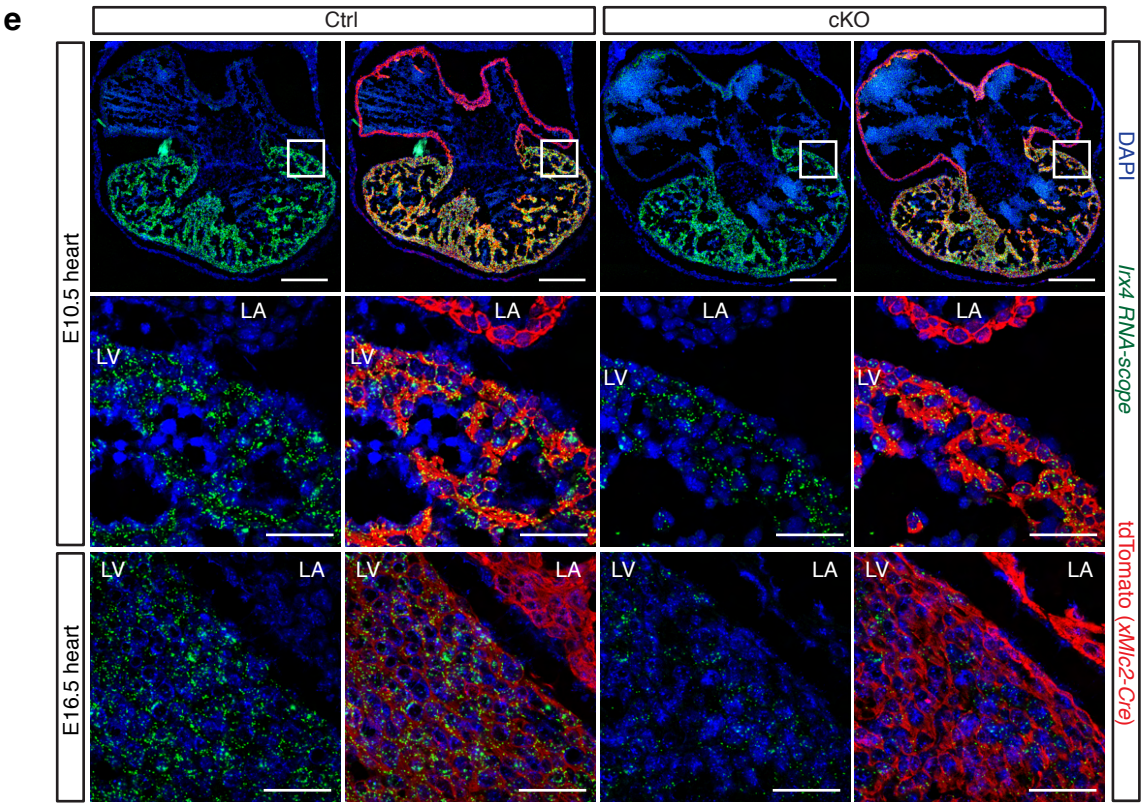
Figure 2 - Cattaneo et al.

a Genes expressed at E16.5 (12110)

439 genes (4%) DOWN in cKO
1039 genes (8%) UP in cKO
10632 genes (88%) unchanged in cKO

b Quartile distribution (%)

	Down	Unch	Up
4th quartile	41	25	7
3rd quartile	27	26	11
2nd quartile	21	26	20
1st quartile	11	23	62

c Top 25 transcription regulators down regulated in cKO

Row Z score +1.5 / −1.5

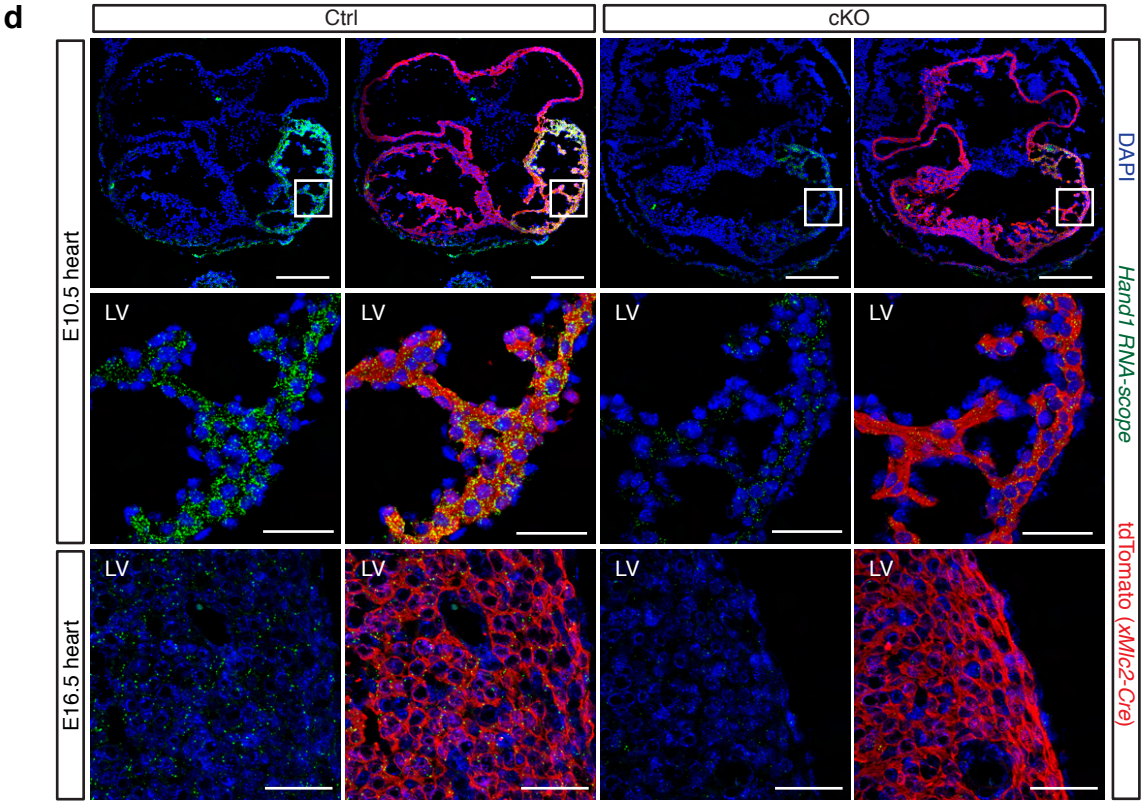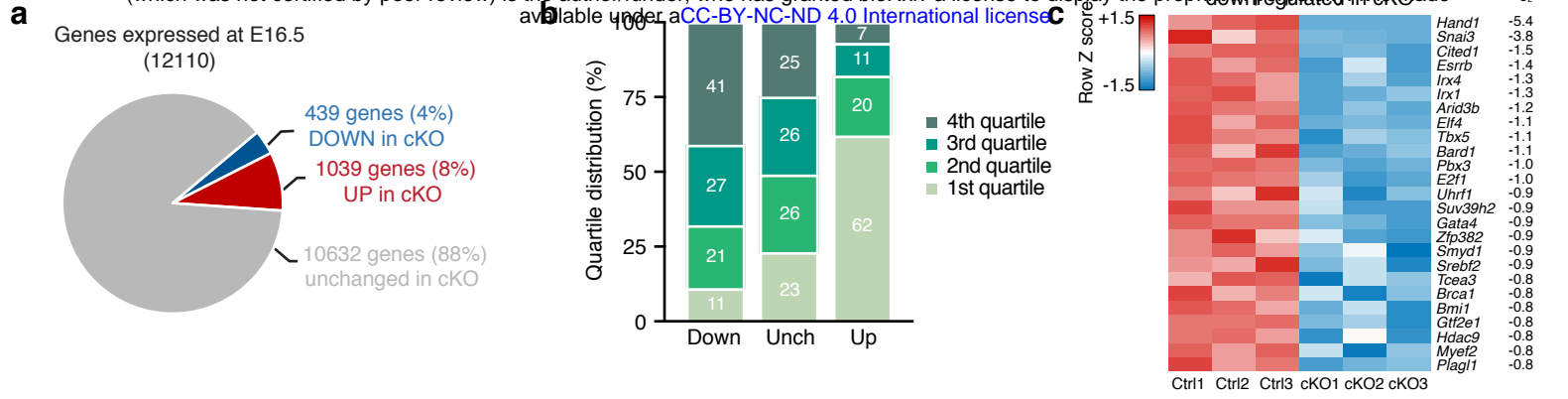| Gene	log₂FC
Hand1	−5.4
Snai3	−3.8
Cited1	−1.5
Esrrb	−1.3
Irx4	−1.3
Irx1	−1.3
Arid3b	−1.2
Tbx5	−1.1
Bard1	−1.1
Pbx3	−1.0
Elf4	−1.0
Uhrf1	−0.9
Suv39h2	−0.9
Gata4	−0.9
Zfp382	−0.9
Smyd1	−0.9
Srebf2	−0.9
Tcea3	−0.8
Brca1	−0.8
Bmi1	−0.8
Gtf2e1	−0.8
Hdac9	−0.8
Myef2	−0.8
Plagl1	−0.8

Ctrl1 Ctrl2 Ctrl3 cKO1 cKO2 cKO3

Figure 2: DOT1L is required in cardiomyocytes for normal expression of cardiac patterning genes. a) Pie chart representing the number of genes down- (log2FC ≤ -0.5; FDR ≤ 0.05) and up-regulated (log2FC ≥ 0.5; FDR ≤ 0.05) in Dot1L cKO cardiomyocytes at E16.5. **b)** Quartile distribution of gene expression in cardiomyocytes at E16.5. Genes downregulated (Down) in Dot1L cKO were expressed at a high level in control cardiomyocytes, whereas the majority of upregulated genes (Up) belonged to the bottom quartile of expression. Genes not significantly modulated (Unch) were evenly distributed across quartiles of expression. Data are shown as stacked percentage bar graph. **c)** Heatmap showing the expression of the top 25 transcription regulators downregulated in Dot1L cKO cardiomyocytes at E16.5, highlight that multiple transcription regulators with known roles in patterning of the cardiac chambers were significantly downregulated. **d-e)** RNA-scope analyses validating blunted expression of *Hand1* (d) and reduced levels of *Irx4* (e) in Dot1L cKOs both at E10.5 and E16.5. Scale bars = 250μm for low magnification panels and 50μm for high magnification.
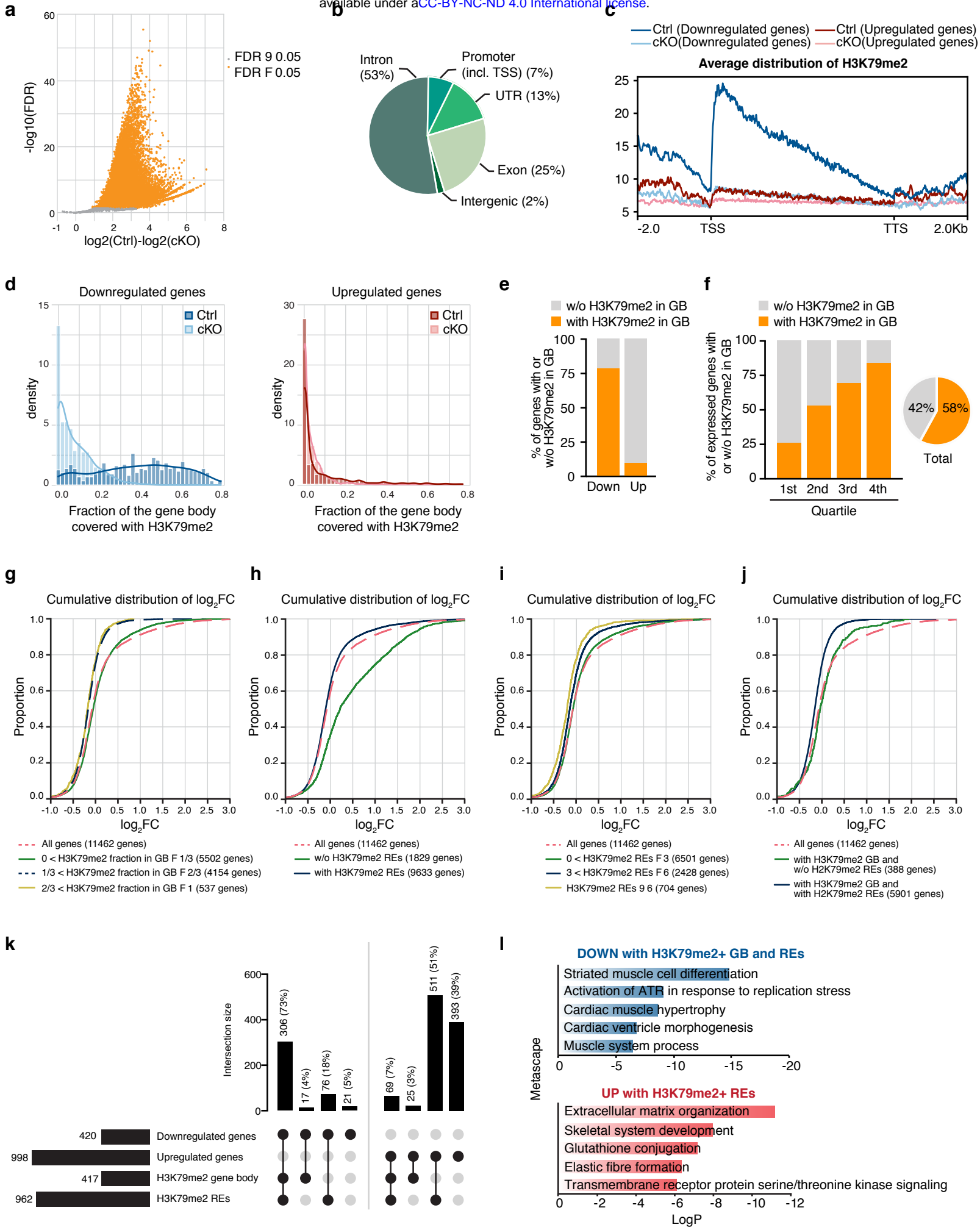
Figure 3 - Cattaneo et al.

Figure 3: DOT1L controls transcription of target genes via a combination of H3K79me2 in gene bodies and regulatory elements. a) Volcano plot displaying H3K79me2 ChIP-seq peaks significantly enriched in E16.5 Dot1L Ctrl vs cKO cardiomyocytes. **b)** Pie chart indicating the genomic distribution of differential H3K79me2 ChIP-seq peaks in E16.5 cardiomyocytes. **c)** Metagene profiles showing the average distribution of H3K79me2 input-normalized tag density relative to Transcription Start Site (TSS) and Transcription Termination Site (TTS) with ±2Kb flanking regions. Genes downregulated in Dot1L cKOs had, in control cardiomyocytes, high levels of H3K79me2 in the vicinity of the TSS that progressively decreased towards the TTS. Genes upregulated in Dot1L cKOs did not show significant levels of H3K79me2 in Ctrl or cKO cardiomyocytes. **d)** Fraction of gene body covered with H3K79me2 in downregulated genes (left graph) and in upregulated genes (right graph) indicating that most genes downregulated in E16.5 Dot1L cKOs had, in Ctrl cardiomyocytes, H3K79me2 covering the majority of the gene body, whereas the majority of genes upregulated in cKOs didn't have H3K79me2 gene body coverage in Ctrl cardiomyocytes. **e)** Graph representing the percentage of down- and up-regulated genes in E16.5 cKO cardiomyocytes with (Coverage ≥ 50 reads and Fraction of gene body ≥ 0.2) or without (Coverage < 50 reads and Fraction of gene body < 0.2) gene body H3K79me2 in E16.5 Ctrl cardiomyocytes. **f)** Percentage of genes with H3K79me2 in the gene body (GB) (Coverage ≥ 50 reads and Fraction of gene body≥ 0.2) or without H3K79me2 in the gene body (Coverage <50 reads and Fraction of gene body< 0.2). This modification was abundant amongst highly expressed genes (4th quartile of RNA expression) and progressively decreased towards the lower quartiles of expression. Globally more than half (58%) of all genes expressed in E16.5 cardiomyocytes had gene body H3K79me2. **g-h-i-j)** Cumulative log2FC distribution analyses in different categories of genes in E16.5 cardiomyocytes: genes with different levels of H3K79me2 fraction in the gene body (GB) (g); genes with and without H3K79me2/H3K27ac positive regulatory elements (REs) (h); genes with increasing numbers of H3K79me2/H3K27ac positive regulatory elements (REs) (i); genes with a combination or none of H3K79me2 in GB and REs (j). DOT1L regulated expression of target genes via H3K79me2 in GB and REs and genes associated with both types of H3K79me2 (GB and RE) were more downregulated in cKOs than those containing exclusively gene body H3K79me2. A Kolmogorov-Smirnov test was applied to assess statistical significance of differences between distribution of gene groups. **k)** UpSet plot indicating the number and percentage of genes up and downregulated in cKOs vs Ctrls with and/or without H3K79me2 in gene body (GB) and regulatory elements (REs) in E16.5 cardiomyocytes. **l)** Metascape pathway analysis of genes downregulated with H3K79me2 in GB and REs (top) and upregulated with H3K79me2 in REs (bottom) in Dot1L cKO vs control E16.5 cardiomyocytes. Top 5 enriched categories are shown, sorted by Log P value.
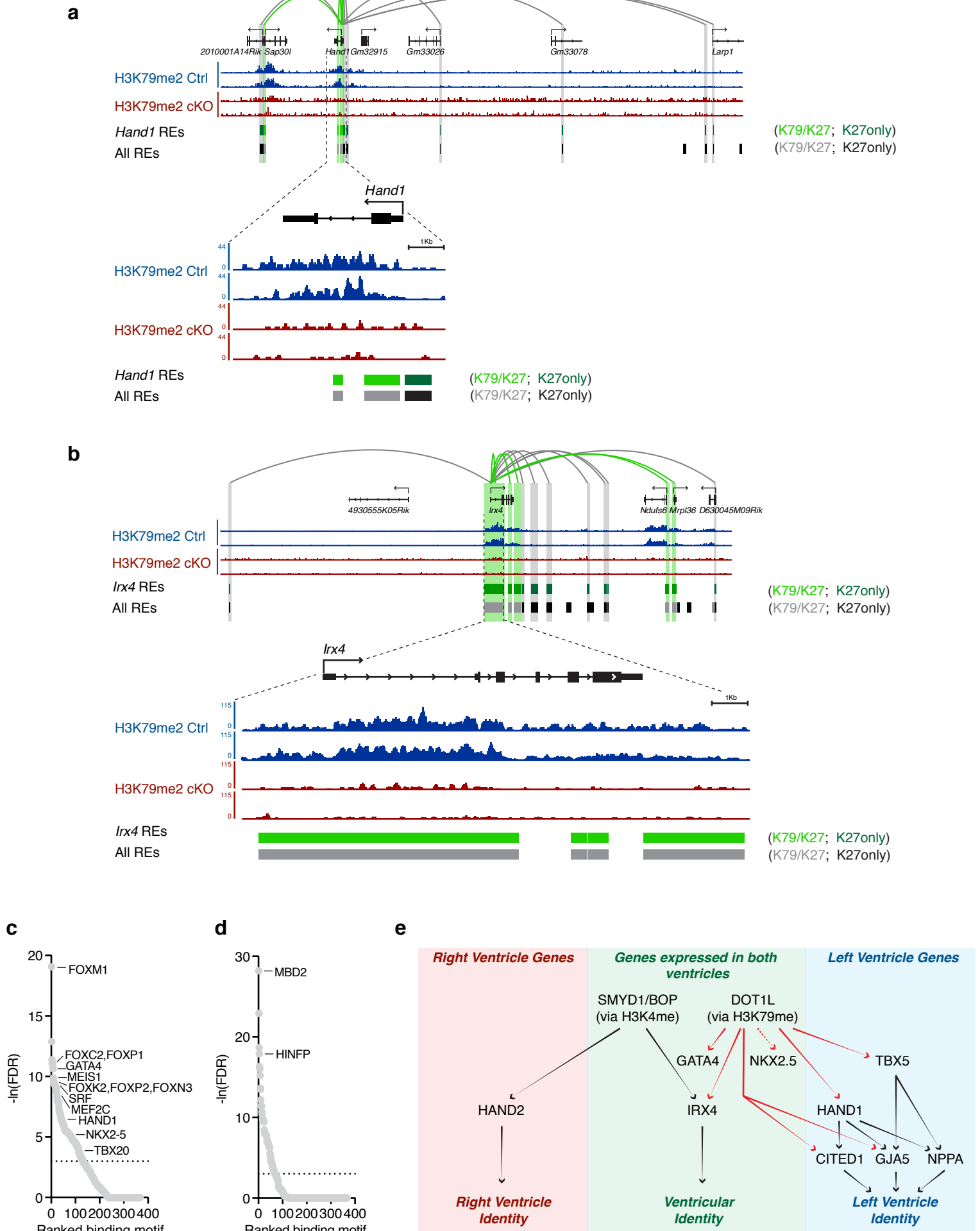
Figure 4 · Cattaneo et al.

Figure 4: Genomic interactions underlying H3K79me2-dependent regulation of *Hand1* and *Irx4* transcription. a, b) Browser tracks displaying H3K79me2 ChIP-seq profiles of Ctrls (blue) and Dot1L cKOs (red) E16.5 cardiomyocytes in the genomic regions containing the *Hand1* (a) and *Irx4* (b) loci. Loops display all significant regulatory interactions between REs and the *Hand1* or *Irx4* loci, as identified by the AbC analysis. Gray loops identify interactions with REs that have H3K27ac but not H3K79me2, whereas green loops identify interactions with REs enriched in both epigenetic marks. For reference, all REs (defined as H3K27ac+ peaks) are displayed in the bottom lane, regardless of their regulatory association with the *Hand1* and *Irx4* loci. **c, d)** Motif enrichment analysis ranking transcription factors (TFs) with TRAP values enriched in REs associated with downregulated genes (c) and upregulated genes (d). **e)** Diagrammatic representation of key cardiac patterning genes directly (red arrows) or indirectly (black arrows) regulated by DOT1L (via H3K79me2).
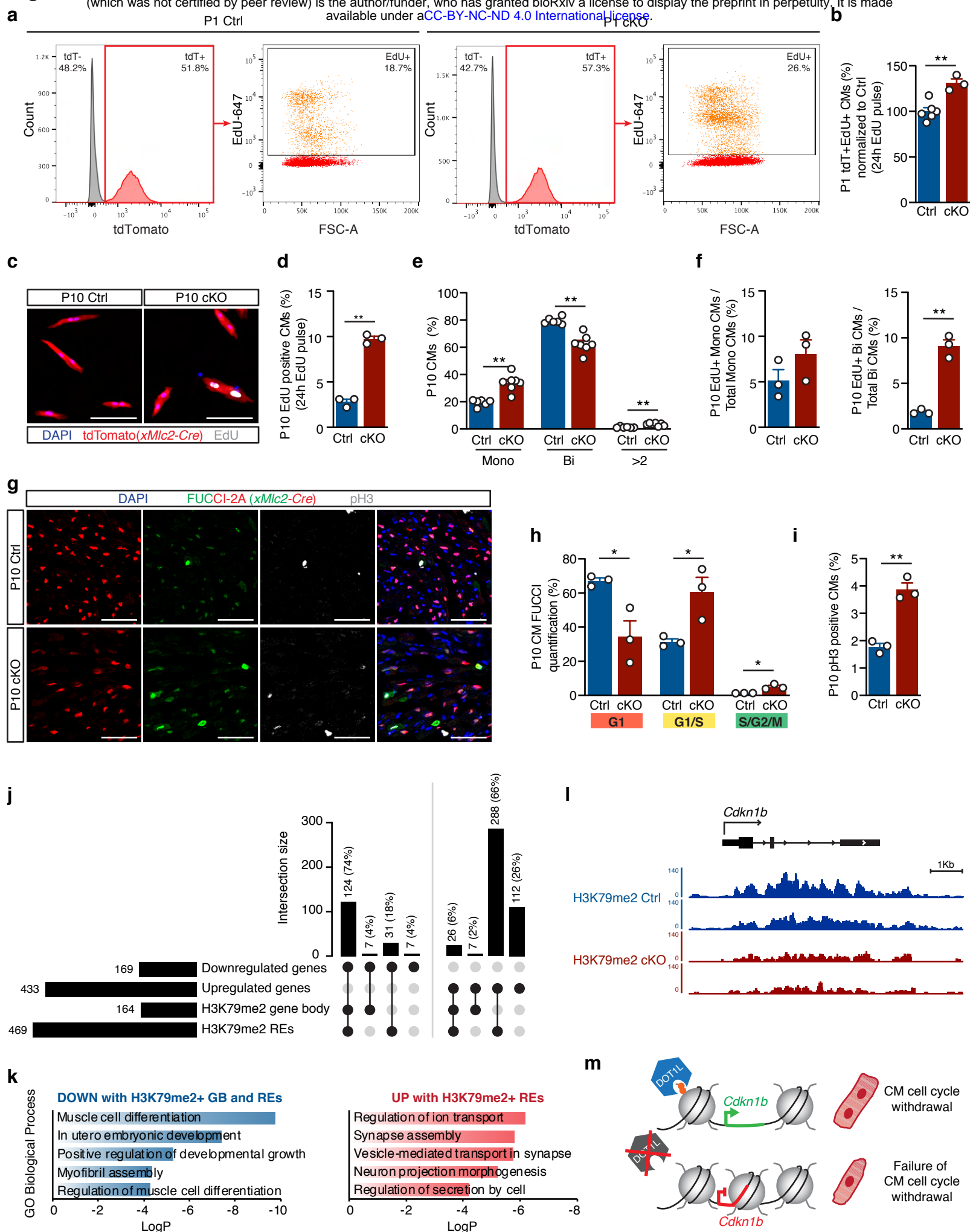
Figure 5 - Cattaneo et al.

Figure 5: Neonatal Dot1L cKO cardiomyocytes fail to undergo cell cycle withdrawal. a-b) Representative FACS analysis (a) and respective quantification (b) showing significantly increased EdU incorporation within P1 cardiomyocytes (tdTomato+ cells) of Dot1L cKO vs Ctrl hearts (N≥3 biological replicates). **c-d)** Representative immunofluorescence images (c), and respective quantification (d) showing significantly increased rates of EdU incorporation in P10 cardiomyocytes isolated from Dot1L cKO vs Ctrl hearts. DAPI in blue, endogenous tdTomato signal driven by XMlc2-Cre;Rosa26-tdTomato in red and EdU in white. (Scale bar =100μm; Mean of 755 CMs counted per heart from N=3 biological replicates). **e)** Quantification of relative percentage of mononuleated (Mono), binuclated (Bi) or multinucleated (>2) cardiomyocytes in P10 Dot1L Ctrls and cKOs. At P10, Dot1L cKO hearts had more mononucleated (Mono) and less binucleated (Bi) cardiomyocytes than littermate Ctrls (mean of 697 CMs counted per heart from N≥6 biological replicates). **f)** Quantification of percentage of EdU+ cardiomyocytes within mononucleated (left graph) and binucleated (right graph) cardiomyocytes of P10 Dot1L Ctrls and cKOs (Mean of 755 CMs counted per heart from N=3 biological replicates). **g,h,i)** Immunofluorescence images (g) and respective quantification (h,i) of P10 Dot1L Ctrl and cKO hearts on a *Rosa26*-FUCCI2A background (Cre-dependent cell cycle indicator). Red-only nuclei represent cardiomyocytes in G1; red+green (yellow) nuclei represent cardiomyocytes in G1/S; green-only nuclei correspond to cardiomyocytes in S/G2/M, pH3 staining in white. Dot1L cKO hearts had a significantly higher percentage of cardiomyocytes in G1/S and in S/G2/M compared to Ctrls, (h, mean of 2070 CMs counted per heart from N=3 biological replicates) and of phophoHistone3+ (pH3) cardiomyocytes (i, mean of 1019 CMs counted per heart from N=3 biological replicates). (Scale bar = 50μm; sections have been quantified from all the compartment of the heart). **j)** UpSet plot indicating the number and percentage of genes up and downregulated in cKOs vs Ctrls with and/or without H3K79me2 in gene body (GB) and regulatory elements (REs) in P1 cardiomyocytes. **k)** Gene Ontology pathway analysis of genes downregulated with H3K79me2 in GB and REs (left) and upregulated with H3K79me2 in REs (right) in Dot1L cKO vs Ctrl P1 cardiomyocytes. Top 5 enriched categories are shown, sorted by Log P value. **l)** Browser tracks of H3K79me2 ChIP-seq in Ctrls (blue) and Dot1L cKOs (red) P1 cardiomyocytes in the genomic regions in the vicinity of *Cdkn1b* locus. **m)** Diagrammatic representation of mechanism of defective cardiomyocyte cell cycle withdrawal in the absence of DOT1L. In all graphs from (b) to (i) data is presented as mean ± SEM; * P≤0.05, **P≤0.01.
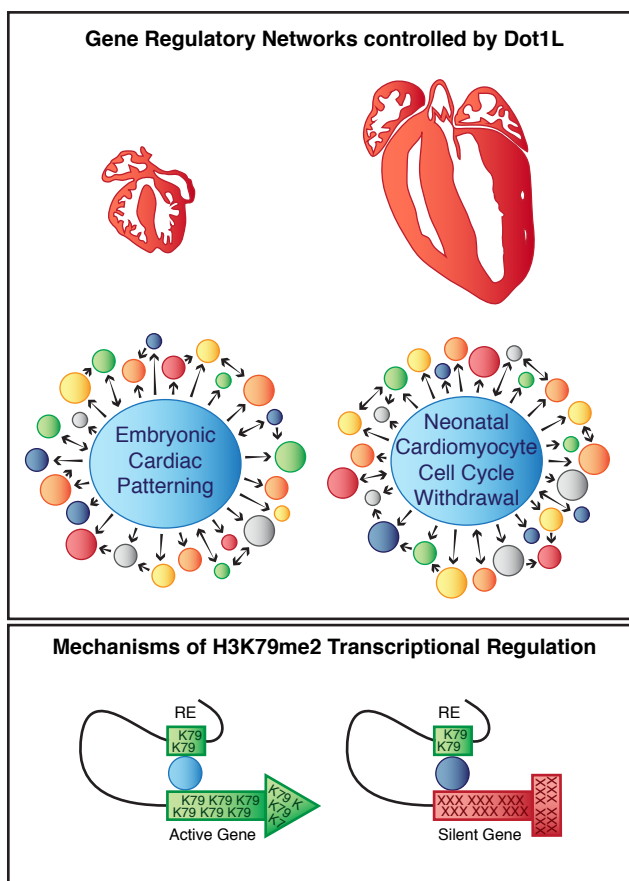
Figure 6

Graphical model of gene regulatory networks in embryonic and postnatal hearts controlled by Dot1L (top). Mechanisms of H3K79me2 transcription regulation (bottom).