# Dynamic tissue-specific H2Bub1 is required for human and mouse cardiogenesis

**Syndi Barish**
Harvard Medical School
**Jeffrey Drozd**
Yale University School of Medicine
**Kathryn Berg**
Yale University School of Medicine
**Isabella Berglund-Brown**
Brown University School of Medicine
**Labeeqa Khizir**
Rutgers Robert Wood Johnson School of Medicine
**Lauren Wasson**
Harvard Medical School
**Christine Seidman**
Harvard Medical School
**Jonathan Seidman**
Harvard University https://orcid.org/0000-0002-9082-3566
**Sidi Chen**
Yale University School of Medicine
**Martina Brueckner** ( martina.brueckner@yale.edu )
Yale University School of Medicine https://orcid.org/0000-0003-0347-5389

**Article**

**Keywords:**

**Posted Date:** April 1st, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1498907/v1

# Dynamic tissue-specific H2Bub1 is required for human and mouse cardiogenesis

Syndi Barish[1], Jeffrey Drozd[2], Kathryn Berg[1], Isabella Berglund-Brown[2], Labeeqa Khizir[2], Lauren K Wasson[3,4], Christine E Seidman[3,4,5], Jonathan G Seidman[3], Sidi Chen[1], Martina Brueckner[1,2,*]

Departments of Genetics[1] and Pediatrics[2]
Yale University School of Medicine,
333 Cedar Street,
New Haven, Connecticut 06510, USA

Department of Genetics[3],
Harvard Medical School
Boston, MA 02115, USA

Division of Cardiovascular Medicine[4]
Brigham and Women's Hospital
Boston, MA 02115, USA

Howard Hughes Medical Institute[5],
Harvard University, Boston, MA 02115, USA

*Correspondence to:
Martina.brueckner@yale.edu

**Abstract**

*De novo* variants affecting the core complex required for monoubiquitination of histone H2B (H2Bub1) are enriched in human congenital heart disease. H2Bub1 is an enigmatic chromatin modification required in stem cell differentiation, cilia function, post-natal cardiomyocyte maturation and transcriptional elongation. However, it is still unknown how H2Bub1 affects cardiogenesis (heart structure formation), which is distinct from cardiomyocyte maturation and underlies congenital heart disease. Here we show that the RNF20-core complex (RNF20-RNF40-UBE2B) is required for cardiogenesis in mouse embryos and is essential for differentiation of human iPSCs into cardiomyocytes. Mice with cardiac-specific deletion of *Rnf20* are e12.5 lethal, have thinned myocardium, a deficient ventricular septum, and abnormal cardiac sarcomere organization. We analyzed H2Bub1 marks during the time course of differentiation of human iPSCs into cardiomyocytes, and demonstrated that H2Bub1 marks are erased from a majority of genes at the transition from cardiac mesoderm to cardiac progenitor cells, but are preserved on a subset of long cardiac-specific genes. Sarcomeric gene expression is dependent on normal H2Bub1 both in mice and in human iPSC-derived cardiomyocytes. Finally, we identify an accumulation of H2Bub1 near the center of tissue-specific genes in human cardiomyocytes, mouse embryonic fibroblasts, and human fetal osteoblasts associated with transcriptional elongation efficiency that is absent in UBE2B knock-out H2Bub1-deficient cardiomyocytes. In summary, normal H2Bub1 distribution is required for cardiac morphogenesis and cardiomyocyte differentiation, and we propose that H2Bub1 regulates tissue-specific gene expression by increasing the efficiency of transcriptional elongation.

## Introduction

Congenital heart disease (CHD), a structural abnormality of the heart and/or great vessels, is the most common cause of mortality from congenital malformations. Whole-exome sequencing of CHD patients identified variants in a broad spectrum of chromatin modifier genes in 2.3% of cases[1-4] including genes affecting H3K4 methylation, H3K27 methylation and H2K120 monoubiquitination. Vertebrates deficient for *Kmt2d* (H3K4 methyltransferase) have a range of cardiac abnormalities[5,6], and studies in iPSC-derived cardiomyocytes show that many chromatin marks, including H3K4me3 and H3K27me3, are dynamic throughout cardiac lineage commitment[7]. However, how monoubiquitination of histone H2B on K120 (H2Bub1) affects structural cardiogenesis remains enigmatic. The H2Bub1 machinery was first discovered in yeast; in mammals, deposition of H2Bub1 requires a complex consisting of the E3 ubiquitin ligases RNF20 and RNF40 and the Ubiquitin Conjugating enzyme E2 B (UBE2B) in addition to interaction with the WW Domain-containing adaptor with coiled-coil (WAC)[8-14]. Unlike most histone marks, H2Bub1 is located on gene bodies, where it is enriched near the promoter and gradually decreases towards the 3' end[15]. It is postulated to function in both activating and repressing gene expression[16,17], and the effect of H2Bub1 on transcriptional regulation may be context-dependent. H2Bub1 has broad biological functions including differentiation, tumor suppression, and inflammation[18-24]. Constitutive deletion of *Rnf20* in mouse leads to failure of preimplantation development[25,26], conditional deletion in the mouse testes results in male infertility[26], and knockdown in *Xenopus* leads to abnormal embryonic left-right axis determination[27]. Interestingly, increased H2Bub1 levels, due to mutations affecting deubiquitination, result in mid-embryonic lethality in mice[28], suggesting that development is sensitive to the level of H2Bub1.

CHD patients show enrichment in *de novo* damaging variants affecting the RNF20 interactome, compared to controls, implicating H2Bub1 in human CHD[2,27]. H2Bub1 is required for cilia gene expression at the left-right organizer to establish left-right asymmetry and direction of heart looping. However, only some patients with variants affecting H2Bub1 have laterality defects, suggesting that H2Bub1 affects cardiac morphogenesis outside of the determination of left-right asymmetry. Postnatal maturation of mouse cardiomyocytes is affected by mosaic deletion of both *Rnf20* and *Rnf40* at postnatal day 0, leading to immature cardiomyocytes at day 28 associated with downregulation of adult-biased metabolism genes[29]. However, this observation does not explain the cardiac defects observed in patients with CHD

and H2Bub1 defects, who have structural heart defects occurring prenatally, implying that there is an additional essential role for H2Bub1 in the entirely distinct process of cardiogenesis. Cardiogenesis is largely completed by mouse e15 and encompasses differentiation of cardiac progenitors, migration into the cardiac crescent, and formation of the heart structure, and is under tight transcriptional control[30]. It is likely that the role of H2Bub1 in cardiogenesis or CHD is distinct from the function of H2Bub1 in post-natal cardiomyocyte maturation.

Here we show that H2Bub1 is required for embryonic cardiomyocyte differentiation and cardiogenesis. Cardiac-specific deletion of *Rnf20* in mouse leads to embryonic lethality, abnormal compact myocardium, disorganized cardiac sarcomeres lacking an H zone, and a deficient ventricular septum. Developmental time-course analysis of H2Bub1 ChIP-seq during differentiation of human iPSCs to cardiomyocytes identifies abundant H2Bub1 marks in iPSCs that are lost on most genes at transition from cardiac mesoderm to cardiac progenitor cells, but are selectively maintained on a subset of genes significantly enriched in sarcomeric calcium genes. This same set of genes is misregulated when H2Bub1 is decreased (through complete deletion of *UBE2B*) *in vitro* and *in vivo*. Finally, we show that local accumulation of the H2Bub1 marks near the center of long genes with tissue-specific expression is correlated with transcriptional elongation efficiency. Together, our data indicate that H2Bub1 is essential for cardiac development through regulation of cardiac sarcomeric genes during cardiomyocyte differentiation and development in human and mouse.

## Results

### *Rnf20* is expressed during mouse heart development

Since *RNF20* variants have been linked to human CHD both with and without abnormal laterality, we asked how RNF20 affects cardiogenesis in mice. Analysis of RNF20 protein in mouse embryonic hearts demonstrated ubiquitous cardiac expression of nuclear RNF20 at e9.5 (**Fig. S1a).** At e11.5, RNF20 is expressed throughout the epicardium and endocardium, and forms an expression gradient in the myocardium with higher RNF20 at the myocardial surface compared to the lumen. Some cytoplasmic signal is observed at this stage, which could reflect a secondary role of RNF20, as seen in a different RING Finger protein, MURF-1 (**Fig. S1b**)[31]. To relate the dynamic nature of H2Bub1 during mouse heart

development to dynamic levels of the ubiquitination complex (RNF20-RNF40-UBE2B), we examined the expression of the complex components and H2Bub1 over time in the heart *in vivo.* Protein levels of RNF20-complex members and H2Bub1/H2B ratios are dynamic between embryonic day 9.5 (e9.5) and postnatal day 0 (P0) (**Fig. S1c)**.

***Rnf20* is required for mouse heart development**

The human CHD-associated variant in *RNF20* is an early stop codon, suggesting haploinsufficiency for *RNF20*, so we analyzed constitutively heterozygous *Rnf20* mice containing a targeted deletion of *Rnf20* by replacing all coding exons with a LacZ reporter (obtained from KOMP2) (**Fig. S2a**)[27,32]. The *Rnf20*$^{+/-}$ mice are phenotypically normal and survive to adulthood with no discernible cardiac abnormalities (**Fig. S2b**). However, no *Rnf20*$^{-/-}$ embryos are recovered post-blastocyst stage, consistent with previous data showing RNF20 is required for preimplantation development (**Fig. 1a, S2b**)[25,26].

To test whether RNF20 is required in cardiogenesis, we generated mice with cardiac-specific deletion of *Rnf20* using a conditional *Rnf20*$^{fl}$ allele containing loxP sites flanking exons 2-4 of the *Rnf20* gene and the *Nkx2.5*Cre driver (**Fig. S3a**)[26,33]. We evaluated timing and specificity of Cre expression by mating with ROSA$^{mt/mg}$ and evaluating Cre expression. Only males that correctly transmitted the Cre were used (**Images in Fig. 1a**). *Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{+}$ embryos are found at Mendelian ratios until e12.25, but none were recovered after e12.5. Further, the *Rnf20*$^{fl/+}$::*Nkx2.5*Cre$^{+}$ control mice are found at Mendelian ratios through birth and appear phenotypically normal, indicating that the phenotype is not secondary to *Nkx2.5* haploinsufficiency resulting from the *Nkx2.5*Cre allele (**Fig. 1a, S3b).** At e11.5, the embryos and heart appear normal by H&E, including 100% normal heart looping direction (**Fig. S3c)**. By e12.25, the compact myocardium is significantly thinned in *Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{+}$ embryos irrespective of comparable diameters to wildtype, and the ventricular septum is deficient (left ventricle compact myocardium thickness p-value = 0.009, right ventricle compact myocardium thickness p-value = 0.004, septum length p-value = 0.0007) (**Fig. 1c).** We verified the *Nkx2.5*-specific deletion of *Rnf20* at e11.5; nuclear RNF20 protein is in both myocardial and epicardial cells in the *Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{-}$ embryos, but is deleted in myocardial cells and retained only in the NKX2.5-negative epicardial cells in *Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{+}$

embryos (**Fig. 1b**). We also investigated the effect of *Rnf20* deletion using a different cardiac-specific driver (*Isl1*Cre)[34] that is mainly expressed in epicardial and endocardial cells. Since these mice have morphologically normal hearts and survive to adulthood in Mendelian ratios, we conclude that in cardiogenesis, the primary function of RNF20 is in cardiomyocyte development (**Fig. 1a, S3d, S3e, S3f, S3g)**. Thus, RNF20 functions in the development of compact myocardium and ventricular septum, indicating an essential role of RNF20 in cardiogenesis consistent with the structural CHD observed in human patients with variants affecting H2Bub1.

**H2Bub1 is dynamically distributed during human cardiomyocyte differentiation**

The majority of NKX2.5-expressing cells become cardiomyocytes, which is consistent with our data showing abnormal development of compact myocardium in response to *Nkx2.5*Cre-mediated *Rnf20* deletion. To investigate H2Bub1 during cardiomyocyte differentiation we utilized *in vitro* differentiation of human induced pluripotent stem cells (iPSCs) into cardiomyocytes (CMs) (**Online Methods)**[35]. The genome-wide H2Bub1 profile and corresponding transcriptional changes during CM development were analyzed by ChIP-seq of H2Bub1 and bulk RNA-seq at five stages of CM differentiation: iPSCs, mesoderm (M), cardiac mesoderm (CMes), cardiac progenitor (CP) and CM (**Fig. S4a**). The stages of differentiation were verified by validating that the RNA-seq replicates cluster together and stage-specific marker genes are expressed at the correct time (**Fig. S4b, Supplemental Data 2).** H2Bub1 ChIP-seq peaks from iPSCs were grouped into four clusters based on H2Bub1 occupancy (high, moderate, low, and none) (**Fig. S5a,b, Supplemental Data 1**). The general profile of H2Bub1 is as previously reported, with very low-occupancy at the transcription start site (TSS) and coverage over the entire gene-body with gradual diminution from 5' to 3'[15]. The moderate-occupancy cluster has a distinct profile compared to the high and low-occupancy clusters. While the high and low-occupancy clusters have constant H2Bub1 throughout the 5' region of the gene-body, H2Bub1 occupancy in the moderate-occupancy cluster decreases more proximal to the 5' end (**Fig. S5b**). Further, H2Bub1 occupancy increases between iPSC and M near the TSS, but remains constant between M and CMes. It then decreases between CMes and CP, and again remains constant between CP and CM (**Fig. S5a).**

We next performed DAVID gene ontology (GO) enrichment analysis on the genes in each cluster for each cell type to obtain their biological context. The high and moderate-occupancy clusters (Clusters 1 and 3) are enriched for general cell maintenance terms, suggesting a role for H2Bub1 in maintaining homeostasis. The no-occupancy cluster (Cluster 4) is enriched for both general cell maintenance and chromatin assembly terms. Most interesting to us is the low-occupancy cluster (Cluster 2), which is enriched in development GO terms. Cluster 2 in iPSCs contains multiple types of developmental genes (muscle, renal, eye, heart, etc.), consistent with previous studies demonstrating RNF20 is required to exit pluripotency[19,20]. Strikingly, as cells commit to a cardiac fate transitioning from CMes to CP and CM, H2Bub1 is retained only on developmental genes related to heart (such as RYR2 and TTN), nervous system, or appendage development (**Fig. S6a, S6b, Supplemental Data 1**).

Total H2Bub1 during transition from iPSC to M increases, and stays stable from M to CMes (**Fig. S6c**), consistent with the ChIP-seq data. In contrast, total H2Bub1 increases from CMes to CP, while ChIP-seq shows a decrease in H2Bub1 around gene bodies during the equivalent stages of CM differentiation (**Fig. S5a, S6c)**. One explanation is that H2Bub1 is decreasing in occupancy around the gene-body, but increasing in occupancy in heterochromatic regions. To test this, we compared the amount of H2Bub1 in the B compartment (heterochromatic compartment) to expected values calculated by generating bootstrap replicates from previously published data at each cell stage (**Online Methods**)[36]. We found H2Bub1 first significantly decreases in the B compartment between iPSC and M, before significantly increasing as the cells progress through all the stages from M to CM (p-values < $1X10^{-5}$). By the CM stage, there is significantly more H2Bub1 in the B compartment than expected by chance (p-values < $1X10^{-5}$) (**Fig. S5c**). The gradual transition of H2Bub1 from gene bodies to heterochromatin is further supported by a significantly higher overlap of H2Bub1 and previously published H3K27me3 peaks (heterochromatic mark) on the same gene in CMs compared to iPSCs (p-values < $1X10^{-5}$), versus no change in overlap of H2Bub1 and previously published H3K4me3 peaks (active genes) (**Fig. S5d, S5e**)[37,38]. Further, in CMs, there is a significantly higher overlap of H2Bub1 and H3K27me3 peaks on the same gene than of H2Bub1 and H3K4me3 peaks (p-value = 0.04). In iPSCs, both overlaps are equivalent (**Fig. S5e)**. We conclude that there are more heterochromatic regions marked by H2Bub1 than active regions at the CM stage. Together these data indicate that H2Bub1 increases in heterochromatic regions,

which accounts for the discrepancy between the observed increase in total H2Bub1 from CMes to CP, while gene-specific H2Bub1 decreases at the same stages.

**H2Bub1 is selectively maintained on sarcomeric calcium genes**

The distribution of many epigenetic marks is dynamic during development[7,39]. To determine the temporal dynamics of H2Bub1 occupancy during cardiomyocyte differentiation, differential binding analysis on the H2Bub1 ChIP-seq data was performed during the progression from iPSCs to CMs. We found 316 regions that have increased H2Bub1 between iPSC and M, compared to 18 regions with decreased H2Bub1. Between M and CMes, there are 2 regions with decreased H2Bub1. The largest change in H2Bub1 happens upon transition from CMes to CP, with a decrease in 25,748 regions (corresponding to 8,909 ENSEMBL genes), while no gene-specific changes in H2Bub1 between CP and CM were observed (**Fig. S5f, Supplemental Data 1)**. This drop off in H2Bub1 occupancy between CMes and CP is further shown in a Venn diagram comparing the unique CMes Ensembl genes near H2Bub1 regions to the unique CP genes (**Fig. S5g).** This is consistent with the overview data (**Fig. S5a).** DAVID GO enrichment analyses on regions that change in H2Bub1 occupancy, based on the average of all three replicates, indicate that they are mostly cell maintenance genes **(Fig. S6d)**. We next compared the genes near regions that change in H2Bub1 occupancy to the genes that change in expression. While the changes in gene expression during the iPSC to M and M to CMes transitions appear to be H2Bub1 independent, there are a significant number of genes (p-value < $1X10^{-100}$) that both lose H2Bub1 and decrease in gene expression between CMes and CP (**Fig. S6e**). Since this mark is thought to be activating, we verified expression of the genes near H2Bub1 marks for all cell stages. We evaluated the overlap of H2Bub1 peaks and expressed genes in the terminal CM stage in more detail and identified 618 genes that are both occupied by H2Bub1 and expressed. Thus, we would predict that these genes are likely to be direct targets of H2Bub1 (**Fig. S5h**). Consistent with a role for H2Bub1 in CMs, DAVID GO analysis of these 618 genes identified a significant enrichment (p-value = 0.014) in cardiac conduction genes.

Given this connection between H2Bub1 and cardiac function genes, we hypothesized that the genes that maintain their mark during the large drop off in H2Bub1 occupancy between CMes and CP are

also cardiac genes. Genes that maintain H2Bub1 in CP are enriched for calcium signaling genes (**Fig. 2a)**. Interestingly, calcium signaling genes remain constant in H2Bub1 occupancy over time (**Fig. 2c**), unlike housekeeping genes which either are not occupied by H2Bub1 at all or increase in H2Bub1 mark between iPSC and M, and downregulate the mark between M and CM (**Fig. 2b**). Further, a significantly higher proportion of calcium signaling genes is identified within the genes that have maintained H2Bub1 compared to genes with downregulated H2Bub1 (Z score = 4.3996, p-value = $1.1 \times 10^{-5}$). Importantly, ten of the eleven calcium signaling genes that maintain H2Bub1 are associated with cardiomyopathy through patient variants and/or mouse models (**Fig. 2a**)[40-49]. These include *CACNA1C* and *RYR2* which have been linked to left ventricular non-compaction (LVNC)[48,49]. Together, these data indicate that H2Bub1 is selectively maintained on tissue-specific genes in CMs to promote their expression.

**Patterns in H2Bub1 occupancy correlate with gene expression**

The distribution of chromatin marks, such as H3K4me3, is known to affect gene expression, so we wanted to determine if the same is true for H2Bub1. We next asked if we could identify patterns of H2Bub1 distribution and amount on CMes marked genes that we could use to predict the subsequent loss of H2Bub1 and gene expression in CP. By using a previously established method, we identified multiple groups of genes (patterns) that have similar amount or distribution of H2Bub1 across the gene body (**Fig. S7a, Supplemental Data 1)**[50]. Interestingly, the calcium signaling genes were among the many genes that did not get placed into any pattern, hereafter referred to as 'unpatterned genes'.

We next calculated the ratio of H2Bub1 peaks that decrease in occupancy between CMes and CP to total H2Bub1 peaks for each of these patterns. Patterned genes have a higher ratio of peaks decreasing in occupancy than average **(Fig. S7b)**. On the contrary, some unpatterned genes, such as the calcium signaling genes, maintain their H2Bub1 occupancy. The decreased occupancy in patterned genes corresponds to their significantly decreased expression in CP compared to CMes (all p-values < 0.05). Thus, if H2Bub1 only occupies a particular region of a gene body (such as in housekeeping genes), it will likely decrease in H2Bub1 occupancy and gene expression between CMes and CP. In contrast, if H2Bub1 is relatively even throughout the gene body (such as in calcium signaling genes), it will instead maintain H2Bub1 occupancy and gene expression **(Fig. 2b, 2c, S7c).** This implies that patterned H2Bub1

marks precede loss of the mark and decreased expression, while evenly distributed H2Bub1 marks are retained and predict continued expression.

**Mutations in the RNF20-complex affect cardiomyocyte differentiation**

Since H2Bub1 is selectively maintained on sarcomeric calcium signaling genes linked to cardiomyopathy, along with the known link between ubiquitinase complex members (*RNF20* and *UBE2B*) and CHD, we next asked how defective H2Bub1 impacts CM development. To address this, we first established iPSC lines with mutations affecting *RNF20*, and evaluated how the mutations impact the capacity for iPSCs to differentiate into CMs. In each of two independent CRISPR experiments, we created one *RNF20*$^{+/-}$ iPSC line (*RNF20*$^{+/-}$1 and *RNF20*$^{+/-}$2). The two independent iPSC lines have different mutations, and both include one frameshift allele and one non-frameshift allele, and are predicted to be functionally hypomorphic. For simplicity, we will be referring to these iPSC lines collectively as *RNF20*$^{+/-}$ (**Fig. S8a, Online Methods**). *RNF20* knockdown was verified by western blotting for RNF20 (**Fig. S8b).**

*RNF20*$^{+/-}$ and wild-type iPSCs were then simultaneously differentiated into CMs[35]. At iPSC, M, and CM stages (when the cells are relatively homogenous),we performed immunofluorescence for markers of pluripotency (OCT4), M (Brachyury), CMes (NKX2.5), CP (ISL1), and CM (cardiac troponin T, TNNT)[35]. At the iPSC stage, both wild-type and *RNF20*$^{+/-}$ cells express OCT4, indicating that decrease in *RNF20* does not affect pluripotency (**Fig. S9a)**. At the M stage, again both wild-type and *RNF20*$^{+/-}$ cells express Brachyury, but only wild-type cells lose their OCT4 expression, suggesting the *RNF20*$^{+/-}$ cells have abnormal ability to exit pluripotency **(Fig. S9b).** Further, since these cells are able to exit pluripotency and it has previously been shown that *RNF20*$^{-/-}$ cells cannot, this provides evidence that the both hypomorphic lines have some residual functional RNF20 activity[19].

Wild-type CMs beat, no longer express Brachyury, and instead express NKX2.5, ISL1, and TNNT. However, most cells in both *RNF20*$^{+/-}$ lines continue to express Brachyury, and have no NKX2.5 or TNNT expression. Interestingly, the *RNF20*$^{+/-}$ cells do express ISL1, a pan-cardiac marker, despite not expressing NKX2.5 (**Fig. S9c**). Since ISL1 is ubiquitously expressed in all CP cells *in vivo*, this may indicate that *RNF20*$^{+/-}$ cells are in a less mature state[51,52]. Thus, even though they express ISL1, it is

unlikely that they are CP cells, since they do not express NKX2.5. Importantly, unlike wild-type cells, most *RNF20*$^{+/-}$ cells fail to beat (even by day 20) (**Fig. 3a, Movie S1, Movie S2, Movie S3, Movie S4**). Interestingly, mice with a mutation affecting H2Bub1 deubiquitination demonstrate a heterogeneity in phenotypes, in that most mice have gastrulation defects, but there are some mice that are able to develop beyond[28]. We are proposing that most *RNF20*$^{+/-}$ cells fail to differentiate into normal CMes, and therefore ultimately fail to make beating CMs, but there are rare cells that can escape and form beating CMs.

To further investigate the possible role for the RNF20-complex on CM differentiation, we also established iPSC lines with mutations affecting *UBE2B*, and evaluated how the mutations impact the capacity for iPSCs to differentiate into CMs. In each of two independent CRISPR experiments, we created one loss-of-function *UBE2B*$^{-/-}$ iPSC lines (*UBE2B*$^{-/-}$*1* and *UBE2B*$^{-/-}$*2*). Since UBE2A and UBE2B are more than 95% identical at the protein level, these knockouts were validated by sequencing the cDNA **(Fig. S10a, Online Methods)**[53]. Mutant and wild-type iPSCs were then simultaneously differentiated into CMs[35]. As with the *RNF20*$^{+/-}$ iPSCs, *UBE2B*$^{-/-}$ iPSCs mirror wild-type, but by M stage, most *UBE2B*$^{-/-}$ cells gain Brachyury expression, while retaining OCT4 expression (**Fig. S11a, S11b)**.

In contrast to the *RNF20*$^{+/-}$ cell lines, about a third of *UBE2B*$^{-/-}$ cells (*UBE2B*$^{-/-}$*1* 25/72, *UBE2B*$^{-/-}$*2* 24/69) beat at the CM stage (**Fig. 3a, Movie S1, Movie S5, Movie S6, Movie S7, Movie S8)**. Given this heterogeneity, we evaluated both beating (lactate selection) and non-beating (no lactate selection) cells at a time corresponding to the wild-type CM stage (**Online Methods**). The beating *UBE2B*$^{-/-}$ cells are mostly Brachyury negative, and NKX2.5, ISL1, and TNNT positive, while the non-beating *UBE2B*$^{-/-}$ cells are mostly Brachyury and ISL1 positive, and NKX2.5 and TNNT negative (**Fig. S11c)**. These data indicate that *UBE2B*$^{-/-}$ iPSCs can differentiate into beating CMs, but do so at a reduced efficiency compared to wild-type iPSCs. Additionally, non-beating *UBE2B*$^{-/-}$ cells are phenotypically equivalent to *RNF20*$^{+/-}$ cells. Collectively, these data demonstrate a requirement for the RNF20-complex in normal CM differentiation.

**Decreased RNF20 increases H2Bub1 occupancy in iPSCs**

Since the *RNF20*$^{+/-}$ cells arrest differentiation into cardiac mesoderm, we asked how global H2Bub1 deposition changes in the iPSC and M stages comparing *RNF20*$^{+/-}$ and WT cells. We determined

that *RNF20*$^{+/-}$ cells have a paradoxical increase in total H2Bub1 (**Fig. S8b)**. We then evaluated the genome-wide H2Bub1 binding profile and gene expression in *RNF20*$^{+/-}$ iPSCs (**Fig. S8c, Supplemental Data 3, Supplemental Data 4)**. RNA-seq of *RNF20*$^{+/-}$ cells shows significant changes in expression of RNF20-complex members and corresponding deubiquitinases, and we predict that the observed increase in total H2Bub1 is the result of combined dysregulation of components of the ubiquitination complex and deubiquitinases (**Fig. S8d**). Consistent with the increased total H2Bub1 levels, H2Bub1 near gene bodies is also increased and correlates with transcriptional changes (**Fig. S12a, S12b, S12d).** 2,100 genes have significant differential H2Bub1 occupancy in *RNF20*$^{+/-}$ iPSCs compared to wild type (**Fig. S12e**). DAVID GO enrichment analysis on genes with increased H2Bub1 occupancy and mRNA expression indicate that many of these genes are involved with transcription, splicing, and/or DNA and protein modifications (**Fig. S12c).** Since these classes of genes broadly affect downstream transcription, decreased *RNF20* likely leads to pleiotropic effects. Even though *RNF20*$^{+/-}$ iPSC H2Bub1 levels deviate so far from normal levels, they are able to reach the M stage. Despite an overall decrease in global H2Bub1, there are few genes with significant differential H2Bub1 occupancy in *RNF20*$^{+/-}$ M cells compared to wild-type (**Fig. S12f, S12g, S12h, Supplemental Data 3),** and the genome-wide H2Bub1 binding profile and gene expression in *RNF20*$^{+/-}$ M cells indicate that decreased *RNF20* causes pleiotropic responses upon exiting pluripotency (**Supplemental Data 3, Supplemental Data 4)**. Together these data suggest that the catastrophic dysregulation of gene expression at the iPSC stage in *RNF20*$^{+/-}$ cells likely prevents the cells from continuing to differentiate past the M stage.

**Total H2Bub1 reduction decreases sarcomeric calcium signaling gene expression *in vitro* and *in vivo***

In contrast to *RNF20*$^{+/-}$ cells, the *UBE2B*$^{-/-}$ cells have decreased total H2Bub1 and are able to form beating CMs (**Fig. S10b**). To understand how reduced H2Bub1 could alter CM gene expression, we performed RNA-seq analysis of beating *UBE2B*$^{-/-}$ CMs, and compared them to time-matched wild-type iPSC-derived CMs (**Fig. 4a, S10c**). Differential expression analysis identified 1393 downregulated and 1555 upregulated transcripts that are shared between both independent *UBE2B*$^{-/-}$ cell lines compared to wild-type (**Supplemental Data 6)**. To obtain a biological context for the genes with decreased expression,

we did DAVID GO enrichment analysis. About half of the significant GO terms (37/76) are related to sarcomeric calcium genes, sarcomere genes, and/or cardiomyopathy genes. Most notably, two genes with decreased expression, *CACNA1C* and *RYR2*, are also amongst the genes with selectively maintained H2Bub1 upon wild-type transition from CMes to CP (**Fig. 4c)**. Thus, we identified tissue-specific genes that are downregulated when *UBE2B* is decreased in embryonic cardiomyocytes, which is distinct from the metabolic genes identified when *RNF20* and *RNF40* were downregulated in postnatal cardiomyocytes[29]. This indicates that the RNF20-complex has two different functions at these two distinct phases of heart development.

To determine if this *in vitro* mechanism also functions *in vivo*, we analyzed expression of the embryonic sarcomeric calcium signaling genes (*Cacna1c*, *Ryr2*, *Ncx*, and *Serca2a*) at e11.5 (the mouse stage that is approximately equivalent to the fully differentiated cardiomyocytes) in $Rnf20^{fl/+}$*::Nkx2.5*$Cre^{+}$ and $Rnf20^{fl/-}$*::Nkx2.5*$Cre^{+}$ embryonic hearts (10 hearts were pooled together to make each sample), prior to any visible cardiac defects (**Fig. 4a**). We observed that the $Rnf20^{fl/-}$*::Nkx2.5*$Cre^{+}$ hearts have significantly lower *Cacna1c* and *Serca2a* expression than $Rnf20^{fl/+}$*::Nkx2.5*$Cre^{+}$ siblings (*Cacna1c* p-value = 0.03, *Serca2a* p-value = 0.001) (**Fig. 4b**). Thus, RNF20-complex dependent H2Bub1 is necessary for normal sarcomeric calcium gene expression in iPSC-derived CMs and in mouse embryo hearts.

**Sarcomeres are abnormal when total H2Bub1 is reduced**

Our results showing a decrease in cardiac sarcomeric gene expression in $UBE2B^{-/-}$ CMs, along with their beating inefficiency, led us to evaluate the effect of reduced H2Bub1 on cardiac sarcomere structure *in* vivo. Transmission electron microscopy revealed a missing H zone in e12.25 $Rnf20^{fl/-}$*::Nkx2.5*$Cre^{+}$ mouse sarcomeres (**Fig. 3b).** The H zone is the region of the sarcomere that is devoid of actin filaments. At the center of the H zone is the M band, consisting mostly of myomesin at e12.25, which functions to anchor the filaments to titin. Due to an elastic domain in the middle of the embryonic splice variant of myomesin, this structure is not able to be seen on TEM at this stage. An abnormal M band will lead to abnormal sarcomere organization, as observed in the $Rnf20^{flx/-}$*::Nkx2.5*$Cre^{+}$ mice[54]. Interestingly, our RNA-seq data show that MURF1, which has been implicated in human hypertrophic cardiomyopathy and cause sarcomeres to lack an H-zone[31,55,56], has decreased expression in the $UBE2B^{-}$

$^{-/-}$ CMs compared to the wild-type (b value = -1.10). Additionally, this RNA-seq data identifies two components of the M-band (Myomesin and Titin) and two proteins required for regulating their alternative splicing (*RBM20* and *RBM24*) with decreased expression (median b value = -1.54)[57,58]. This indicates that H2Bub1 regulates M band proteins, which are required for normal sarcomeric structure and beating efficiency.

**Accumulation of H2Bub1 near the center of tissue-specific genes correlates with enhanced efficiency of transcriptional elongation**

Strikingly, while loss of *UBE2B* leads to decreased total H2Bub1 levels, ChIP-seq of CMs comparing H2Bub1 between wild-type and *UBE2B*$^{-/-}$ cells demonstrates that gene-specific H2Bub1 is only decreased in 8 Ensembl genes (**Fig. S10b, S10d, Supplemental Data 5)**. To identify differences in H2Bub1 patterns between *UBE2B*$^{-/-}$ and wild-type CMs, we created metagenes corresponding to calcium signaling and sarcomeric genes (calcium genes n = 28, sarcomere genes n = 70, **Supplemental Data 5)**. Wild-type cells have an accumulation of H2Bub1 near the center of the metagenes, but this accumulation is either reduced or completely absent in *UBE2B*$^{-/-}$ cells (**Fig. 5a, 5b)**. To validate that this accumulation is not a technical artifact, we repeated our analysis on sixty "random" sets of quantity and sized matched gene sets and do not find this accumulation. We provide one representative graph (**Fig. 5c**). Thus, this accumulation is specific to calcium signaling and sarcomeric genes.

Since the RNF20-complex is known to be involved in transcriptional elongation and H2Bub1 is found on long tissue-specific genes during left-right patterning of the heart, we hypothesized that the accumulation of H2Bub1 near the center of long genes may support their transcriptional elongation (median length of all genes is 34 Kb, compared to 456 and 69 Kb for calcium signaling and sarcomeric genes, respectively). To test this hypothesis, we used an indirect assay to look for abbreviated transcripts by evaluating whether there was loss of the 3' end of transcripts in *UBE2B*$^{-/-}$ CMs compared to wild-type. When comparing the ratio of mutant to wild-type RNA in both the calcium and sarcomeric gene sets, transcripts are less abundant in *UBE2B*$^{-/-}$ cells in the most 3' 20% of the gene, indicating inefficient transcription elongation in the mutants; example transcript traces for *CACNA1C* and *RYR2* are shown (**Fig. 5d, 5e, 5g).** To validate this conclusion, we repeated the same analysis on our "random" gene sets,

379 which only had transcriptional efficiency drop-off in the last 5-10% of the 3' end. We provide one
380 representative graph (**Fig. 5f**). Thus, these data support the conclusion that accumulation of H2Bub1 near
381 the center of calcium and sarcomeric genes is UBE2B dependent and correlates with enhanced
382 transcription elongation efficiency.
383 To determine whether H2Bub1 accumulation near the center of genes is identified outside of
384 cardiomyocyte-specific genes, we generated metagenes corresponding to all long genes (defined as
385 genes in the 3$^{rd}$ quartile based on length) and found H2Bub1 accumulation near the center of genes in all
386 long genes. However, UBE2B-dependent accumulation near the center of the gene is unique to tissue-
387 specific long genes (**Fig. 5h**). We also created the metagene plots for genes in the other quartiles and
388 conclude that this accumulation is only present in genes in the 2$^{nd}$ and 3$^{rd}$ quartiles. Genes in the 1$^{st}$
389 quartile follow the previously published pattern of H2Bub1 occupancy and genes in the 4$^{th}$ quartile have
390 accumulation closer to the 3' end, which appears to be UBE2B-dependent (**Figs. S13a-d**). Therefore, the
391 UBE2B-dependent H2Bub1 accumulation near the center of the gene is limited to certain gene classes,
392 including calcium and sarcomeric genes.
393 We next asked if this H2Bub1 accumulation is unique to CMs. Given that previous literature
394 identified H2Bub1 on cilia genes (which have an average length of 73 Kb) in oviducts, we hypothesized
395 that this is likely more generalizable[27]. To address this, we used previously published H2Bub1-ChIP-seq
396 data of mESCs and MEFs as well as previously published H2Bub1-ChIP-seq data of undifferentiated
397 human fetal osteoblasts (hFOBs) and differentiated hFOBs[28,59]. We generated metagenes corresponding
398 to MEF-specific genes (extra-cellular matrix (ECM) (n = 130)) and hFOB-specific genes (epidermal
399 growth factor related genes (EGF) (n = 24)), with a median length within the 3$^{rd}$ quartile (median length is
400 48 Kb for ECM and 99 Kb for EGF) (**Supplemental Data 5)**. These metagenes show an accumulation of
401 H2Bub1 near the center of the gene, which is higher in the MEFs and differentiated hFOBs, as expected
402 (**Fig. 5i, 5j)**. As in CMs, all of the quantity and size-matched "random" genes sets generated show no
403 accumulation of H2Bub1 near the center of the gene, as shown in one representative graph (**Fig. S13e,**
404 **S13f**). The CM, MEF, and hFOB data collectively suggest that accumulation of H2Bub1 near the center of
405 tissue-specific long genes is a general mechanism for regulation of transcriptional efficiency, particularly
406 in specialized fully differentiated tissues. In CMs, central H2Bub1 accumulation and efficient transcription

of sarcomeric and calcium signaling genes depend on UBE2B, providing a mechanism to explain the cardiac phenotype in the *Rnf20*$^{fl/fl}$::*Nkx2.5*Cre mouse model.

## Discussion

Together, our data support a requirement for tight control of H2Bub1 levels in cardiogenesis. Increased total H2Bub1 in human cells leads to failure to form CMs, while decreased total H2Bub1 reduces the efficiency of CM differentiation *in vitro*, and leads to reduced expression of calcium signaling genes and structural abnormalities of the cardiac sarcomere, including a deficient H zone, *in vivo* **(Fig. 6a).** H2Bub1 is highly dynamic during human CM differentiation: early, the mark is increased and then decreased on housekeeping genes, and later the mark is selectively maintained on calcium signaling genes, while it shifts from euchromatic to heterochromatic regions (**Fig. 6b).** Finally, we show that the shape of the H2Bub1 mark changes between wild-type and UBE2B$^{-/-}$ CMs. The profile of H2Bub1 on short genes is as previously reported, with coverage over the entire gene-body, higher at the 5' end than the 3' end of the gene[15]. Notably, long genes have a vastly different profile: there is tissue-specific UBE2B-dependent H2Bub1 accumulation near the center of the gene, which correlates with efficient transcriptional elongation (**Fig. 6c**).

Our data, in combination with previous H2Bub1 literature, suggest that the RNF20-complex has tissue-specific and time-specific functions. Focusing on heart development, at the left-right organizer stage, H2Bub1 is predicted to be located on cilia genes, then at the cardiogenesis stage, H2Bub1 is located on calcium signaling genes, and finally at the cardiac maturation stage, H2Bub1 is located on metabolic genes[27,29]. These data support three separate functions for H2Bub1 in the heart: control of cardiac left-right asymmetry, establishing embryonic cardiac structure and function, and post-natal cardiomyocyte maturation; the first two of which have direct relevance to the mechanism underlying CHD.

The observed dynamic levels and distribution of H2Bub1 during mouse heart development and hiPSC to CM differentiation support a model where tissue-specific transcriptional effects are determined by the retention, instead of deposition, of H2Bub1 on tissue-specific genes. At the level of individual genes, we observed accumulation of H2Bub1 near the center of a subset of long genes correlating with previous findings that RNA Pol II exhibits a similar accumulation near the center of all expressed

genes[60,61]. We propose that H2Bub1 enhances transcriptional elongation of long tissue-specific genes. This conclusion is supported by previous work finding that long genes are more likely to depend on RNF20 to be induced upon differentiation into neuronal cells[19]. Further evidence is the presence of H2Bub1 on long cilia genes in the multi-ciliated cells of oviducts (which require expression of long motile cilia genes), but not in the liver cells, which do not[27]. Combined with our current data analyzing CMs, MEFs, and hFOBs, we propose that targeted, localized H2Bub1 accumulation is a more general mechanism regulating tissue-specific transcriptional elongation efficiency on long genes. Our data indicate that RNF20-complex-dependent H2Bub1 is necessary for normal cardiac development through regulating the transcriptional elongation efficiency of long cardiac calcium signaling and sarcomeric genes during CM differentiation and development.

The question remains how H2Bub1 affects development of cardiac structure, since both human patients with variants affecting H2Bub1 and mouse embryos with cardiac-specific deletion of *Rnf20* have structural heart defects. A possible link to the observed structural heart defects in mice is that altered expression of calcium signaling genes and abnormal sarcomeric structure observed in *Rnf20*$^{fl/-}$ ::*Nkx2.5*Cre$^{+}$ mouse embryos lead to defective cardiac function during embryonic development, and that the resulting hemodynamic derangement affects structural cardiac morphogenesis. Extensive evidence supports interdependence between embryonic hemodynamics and valve development, cardiac trabeculation, myocardial proliferation and formation of the epicardium (reviewed in [62]). Genomic studies of human CHD patients are beginning to provide further evidence of an overlap between genes classically linked to cardiomyopathy, and patients presenting with structural CHD. For example, dominant variants affecting myosin heavy chain 6 (*MYH6)* are associated with cardiomyopathy and atrial septal defects[63,64], while recessive variants in *MYH6* are found in 11% of patients with Shone syndrome, characterized by valve defects and multiple levels of left ventricular obstruction[2]. The shared role of H2Bub1 in CM differentiation and cardiogenesis in mouse and human provide further support for genetic overlap between cardiac structure development and myocardial function, and suggest that a subset of patients with structural heart defects caused by genetic defects affecting cardiomyocytes may be more vulnerable to myocardial dysfunction. Although there are likely variations in the absolute H2Bub1 levels required for normal iPSC-derived CM development, mouse embryo, and human heart development, our observations

in iPSC-derived CMs and mouse embryos indicate a shared requirement for precise control of H2Bub1 in the heart.

**Acknowledgements**

We thank the Center for Cellular and Molecular Imaging at Yale for transmission electron microscopy and the Developmental Histology Tissue Services at Yale for adult mouse heart histology.. We would also like to thank Dr. Yibing Qyang and Dr. William Pu for sharing mouse lines with us and Dr. Patrick Gallagher for a critical read of the manuscript. This work was supported by NSFGRFP to SB, and R35HL145249 and 2UO1HL098162 to MB.

**Author Contributions**

SB and MB conceived and designed the experiments. SB, JD, KB, IBB, LK, and MB performed the experiments in mouse. SB and LW performed the experiments in the iPSCs. SB performed the bioinformatics analysis of the ChIP-seq and RNA-seq experiments, and CES, JGS and SC provided conceptual advice on the analysis. SB and MB wrote the manuscript, which was read and approved by all of the authors.

**References**

1. Homsy, J. *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262-6 (2015).
2. Jin, S.C. *et al.* Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet* **49**, 1593-1601 (2017).
3. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220-3 (2013).
4. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet* **48**, 1060-5 (2016).
5. Ang, S.Y. *et al.* KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. *Development* **143**, 810-21 (2016).
6. Van Laarhoven, P.M. *et al.* Kabuki syndrome genes KMT2D and KDM6A: functional analyses demonstrate critical roles in craniofacial, heart and brain development. *Hum Mol Genet* **24**, 4443-53 (2015).
7. Wamstad, J.A. *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206-20 (2012).

8. Robzyk, K., Recht, J. & Osley, M.A. Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**, 501-4 (2000).
9. Wood, A. *et al.* Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol Cell* **11**, 267-74 (2003).
10. Kim, J., An, Y.K., Park, S. & Lee, J.S. Bre1 mediates the ubiquitination of histone H2B by regulating Lge1 stability. *FEBS Lett* **592**, 1565-1574 (2018).
11. Kim, J., Hake, S.B. & Roeder, R.G. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol Cell* **20**, 759-70 (2005).
12. Kim, J. *et al.* RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**, 459-71 (2009).
13. Zhang, F. & Yu, X. WAC, a functional partner of RNF20/40, regulates histone H2B ubiquitination and gene transcription. *Mol Cell* **41**, 384-97 (2011).
14. Gallego, L.D. *et al.* Phase separation directs ubiquitination of gene-body nucleosomes. *Nature* **579**, 592-597 (2020).
15. Jung, I. *et al.* H2B monoubiquitylation is a 5'-enriched active transcription mark and correlates with exon-intron structure in human cells. *Genome Res* **22**, 1026-35 (2012).
16. Lee, J.S. *et al.* Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* **131**, 1084-96 (2007).
17. Xie, W. *et al.* RNF40 regulates gene expression in an epigenetic context-dependent manner. *Genome Biol* **18**, 32 (2017).
18. Chen, S., Li, J., Wang, D.L. & Sun, F.L. Histone H2B lysine 120 monoubiquitination is required for embryonic stem cell differentiation. *Cell Res* **22**, 1402-5 (2012).
19. Fuchs, G. *et al.* RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. *Mol Cell* **46**, 662-73 (2012).
20. Karpiuk, O. *et al.* The histone H2B monoubiquitination regulatory pathway is required for differentiation of multipotent stem cells. *Mol Cell* **46**, 705-13 (2012).
21. Shema, E., Kim, J., Roeder, R.G. & Oren, M. RNF20 inhibits TFIIS-facilitated transcriptional elongation to suppress pro-oncogenic gene expression. *Mol Cell* **42**, 477-88 (2011).
22. Shema, E. *et al.* The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev* **22**, 2664-76 (2008).
23. Tarcic, O. *et al.* RNF20 Links Histone H2B Ubiquitylation with Inflammation and Inflammation-Associated Cancer. *Cell Rep* **14**, 1462-76 (2016).
24. Wang, E. *et al.* Histone H2B ubiquitin ligase RNF20 is required for MLL-rearranged leukemia. *Proc Natl Acad Sci U S A* **110**, 3901-6 (2013).
25. Ooga, M., Suzuki, M.G. & Aoki, F. Involvement of histone H2B monoubiquitination in the regulation of mouse preimplantation development. *J Reprod Dev* **61**, 179-84 (2015).
26. Xu, Z. *et al.* H2B ubiquitination regulates meiotic recombination by promoting chromatin relaxation. *Nucleic Acids Res* **44**, 9681-9697 (2016).
27. Robson, A. *et al.* Histone H2B monoubiquitination regulates heart development via epigenetic control of cilia motility. *Proc Natl Acad Sci U S A* **116**, 14049-14054 (2019).

28. Wang, F. *et al.* Histone H2Bub1 deubiquitylation is essential for mouse development, but does not regulate global RNA polymerase II transcription. *Cell Death Differ* (2021).
29. VanDusen, N.J. *et al.* Massively parallel in vivo CRISPR screening identifies RNF20/40 as epigenetic regulators of cardiomyocyte maturation. *Nat Commun* **12**, 4442 (2021).
30. Brade, T., Pane, L.S., Moretti, A., Chien, K.R. & Laugwitz, K.L. Embryonic heart progenitors and cardiogenesis. *Cold Spring Harb Perspect Med* **3**, a013847 (2013).
31. McElhinny, A.S., Kakinuma, K., Sorimachi, H., Labeit, S. & Gregorio, C.C. Muscle-specific RING finger-1 interacts with titin to regulate sarcomeric M-line and thick filament structure and may have nuclear functions via its interaction with glucocorticoid modulatory element binding protein-1. *J Cell Biol* **157**, 125-36 (2002).
32. Valenzuela, D.M. *et al.* High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* **21**, 652-9 (2003).
33. Moses, K.A., DeMayo, F., Braun, R.M., Reecy, J.L. & Schwartz, R.J. Embryonic expression of an Nkx2-5/Cre gene using ROSA26 reporter mice. *Genesis* **31**, 176-80 (2001).
34. Pfaff, S.L., Mendelsohn, M., Stewart, C.L., Edlund, T. & Jessell, T.M. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309-20 (1996).
35. Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. *Nat Protoc* **8**, 162-75 (2013).
36. Bertero, A. *et al.* Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. *Nat Commun* **10**, 1538 (2019).
37. Inoue, F. *et al.* A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* **27**, 38-52 (2017).
38. Kazachenka, A. *et al.* Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* **175**, 1259-1271 e13 (2018).
39. Mei, H. *et al.* H2AK119ub1 guides maternal inheritance and zygotic deposition of H3K27me3 in mouse embryos. *Nat Genet* **53**, 539-550 (2021).
40. Tang, Y., Tian, X., Wang, R., Fill, M. & Chen, S.R. Abnormal termination of Ca2+ release is a common defect of RyR2 mutations associated with cardiomyopathies. *Circ Res* **110**, 968-77 (2012).
41. Knight, W.E. *et al.* PDE1C deficiency antagonizes pathological cardiac remodeling and dysfunction. *Proc Natl Acad Sci U S A* **113**, E7116-E7125 (2016).
42. Wang, X. *et al.* Generation and phenotypic characterization of Pde1a mutant mice. *PLoS One* **12**, e0181087 (2017).
43. Wang, L. *et al.* Dual LQT1 and HCM phenotypes associated with tetrad heterozygous mutations in KCNQ1, MYH7, MYLK2, and TMEM70 genes in a three-generation Chinese family. *Europace* **18**, 602-9 (2016).
44. Vasti, C. & Hertig, C.M. Neuregulin-1/erbB activities with focus on the susceptibility of the heart to anthracyclines. *World J Cardiol* **6**, 653-62 (2014).
45. Xu, H. *et al.* A Genome-Wide Association Study of Idiopathic Dilated Cardiomyopathy in African Americans. *J Pers Med* **8**(2018).

46. Boczek, N.J. *et al.* Identification and Functional Characterization of a Novel CACNA1C-Mediated Cardiac Disorder Characterized by Prolonged QT Intervals With Hypertrophic Cardiomyopathy, Congenital Heart Defects, and Sudden Cardiac Death. *Circ Arrhythm Electrophysiol* **8**, 1122-32 (2015).
47. Kepenek, E.S. *et al.* Differential expression of genes participating in cardiomyocyte electrophysiological remodeling via membrane ionic mechanisms and Ca(2+)-handling in human heart failure. *Mol Cell Biochem* **463**, 33-44 (2020).
48. Mazzarotto, F. *et al.* Systematic large-scale assessment of the genetic architecture of left ventricular noncompaction reveals diverse etiologies. *Genet Med* **23**, 856-864 (2021).
49. Vasilescu, C. *et al.* Genetic Basis of Severe Childhood-Onset Cardiomyopathies. *J Am Coll Cardiol* **72**, 2324-2338 (2018).
50. Zhang, L. & Zhang, S. Learning common and specific patterns from data of multiple interrelated biological scenarios with matrix factorization. *Nucleic Acids Res* **47**, 6606-6617 (2019).
51. Prall, O.W. *et al.* An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* **128**, 947-59 (2007).
52. Dorn, T. *et al.* Direct nkx2-5 transcriptional repression of isl1 controls cardiomyocyte subtype identity. *Stem Cells* **33**, 1113-29 (2015).
53. Sarcevic, B., Mawson, A., Baker, R.T. & Sutherland, R.L. Regulation of the ubiquitin-conjugating enzyme hHR6A by CDK-mediated phosphorylation. *EMBO J* **21**, 2009-18 (2002).
54. Lange, S., Pinotsis, N., Agarkova, I. & Ehler, E. The M-band: The underestimated part of the sarcomere. *Biochim Biophys Acta Mol Cell Res* **1867**, 118440 (2020).
55. Higashikuse, Y. *et al.* Perturbation of the titin/MURF1 signaling complex is associated with hypertrophic cardiomyopathy in a fish model and in human patients. *Dis Model Mech* **12**(2019).
56. Chen, S.N. *et al.* Human molecular genetic and functional studies identify TRIM63, encoding Muscle RING Finger Protein 1, as a novel gene for human hypertrophic cardiomyopathy. *Circ Res* **111**, 907-19 (2012).
57. Yang, J. *et al.* RBM24 is a major regulator of muscle-specific alternative splicing. *Dev Cell* **31**, 87-99 (2014).
58. Schneider, J.W. *et al.* Dysregulated ribonucleoprotein granules promote cardiomyopathy in RBM20 gene-edited pigs. *Nat Med* **26**, 1788-1800 (2020).
59. Najafova, Z. *et al.* BRD4 localization to lineage-specific enhancers is associated with a distinct transcription factor repertoire. *Nucleic Acids Res* **45**, 127-141 (2017).
60. Young, M.D. *et al.* ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res* **39**, 7415-27 (2011).
61. Fuchs, G., Hollander, D., Voichek, Y., Ast, G. & Oren, M. Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. *Genome Res* **24**, 1572-83 (2014).
62. Andres-Delgado, L. & Mercader, N. Interplay between cardiac function and heart development. *Biochim Biophys Acta* **1863**, 1707-16 (2016).
63. Ching, Y.H. *et al.* Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet* **37**, 423-8 (2005).

64. Hershberger, R.E. *et al.* Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ Cardiovasc Genet* **3**, 155-61 (2010).

Figure 1

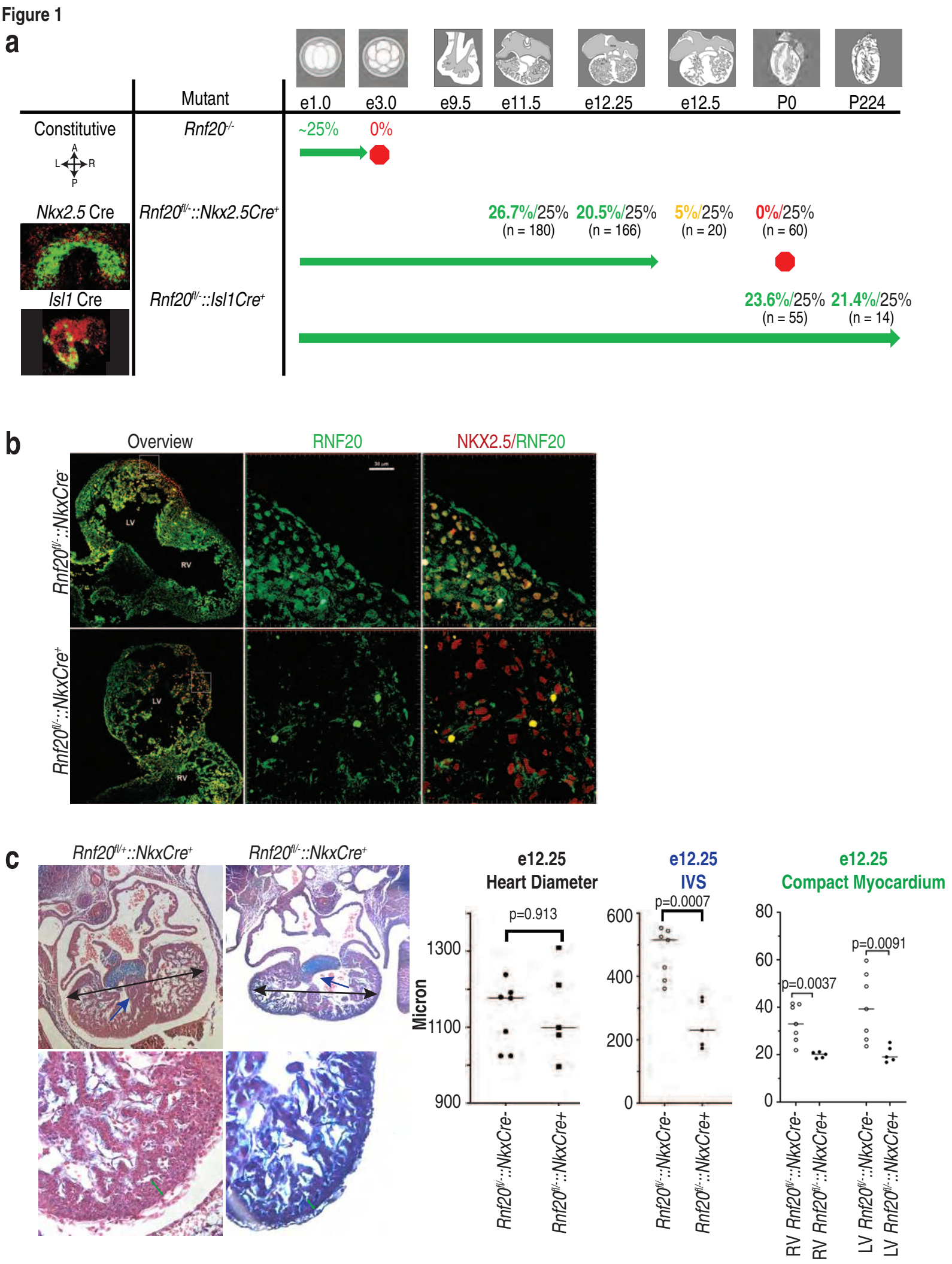

a
Mutant
e1.0
e3.0
e9.5
e11.5
e12.25
e12.5
P0
P224
Constitutive
Rnf20-/-
~25%
0%
A
L
R
P
Nkx2.5 Cre
Rnf20fl/-::Nkx2.5Cre+
26.7%/25%
(n = 180)
20.5%/25%
(n = 166)
5%/25%
(n = 20)
0%/25%
(n = 60)
Isl1 Cre
Rnf20fl/-::Isl1Cre+
23.6%/25%
(n = 55)
21.4%/25%
(n = 14)
b
Overview
RNF20
NKX2.5/RNF20
Rnf20fl/-::NkxCre-
Rnf20fl/-::NkxCre+
LV
RV
c
Rnf20fl/+::NkxCre+
Rnf20fl/-::NkxCre+
e12.25
Heart Diameter
p=0.913
Micron
1300
1100
900
e12.25
IVS
p=0.0007
600
400
200
0
e12.25
Compact Myocardium
p=0.0037
p=0.0091
80
60
40
20
0
Rnf20fl/-::NkxCre-
Rnf20fl/-::NkxCre+
RV Rnf20fl/-::NkxCre-
RV Rnf20fl/-::NkxCre+
LV Rnf20fl/-::NkxCre-
LV Rnf20fl/-::NkxCre+

**Figure 1: *Rnf20* is required for heart development**

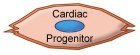
**a)** *Rnf20* mutant mouse survival chart for constitutive nulls, *Rnf20*$^{flx/-}$::*Nkx2.5*Cre$^{+}$, and *Rnf20*$^{flx/-}$::*Isl1*Cre$^{+}$. Colored percentages indicate the observed percentage of null or conditionally null mice (green indicates mendelian ratios, yellow indicates deviation from mendelian ratios, and red indicates no mice). Black percentages indicate the predicted percentages. The sample size is listed below each percentage. See Figure S2 and S3 for more details on these crosses. Images were taken of the cross between a *Nkx2.5*-cre positive (e8.5 (cardiac crescent), signal in cardiac crescent) or *Isl1*-cre positive mouse (e9.0 (heart tube), signal in outflow tract and atria) and a ROSA$^{mt/mg}$ mouse to illustrate distribution of Cre-positive cells. Drawings indicate the expected morphology at each stage mouse heart development. OFT – outflow tract, RV – right ventricle, LV – left ventricle, AVC – atrioventricular canal, ECC – endocardial cushions, RA – right atrium, LA – left atrium, A – anterior, P – posterior, L – left, R - right.

**b)** Immunofluorescent staining for NKX2.5 and RNF20 in e11.5 wild-type (*Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{-}$) and mutant (*Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{+}$) mouse hearts. RV – right ventricle, LV – left ventricle.

**c)** Example hematoxylin and eosin stained e12.25 wild-type (*Rnf20*$^{fl/+}$::*Nkx2.5*Cre$^{+}$) and mutant (*Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{+}$) mouse. Quantifications of interventricular septum length, heart diameter, thickness of right ventricle compact myocardium, and thickness of left ventricle compact myocardium are displayed as individual data points with a line representing the median (n = 7 wild-type and 5 mutant hearts). Unpaired 2-tailed, heteroscedastic t-test.

Figure 2

a

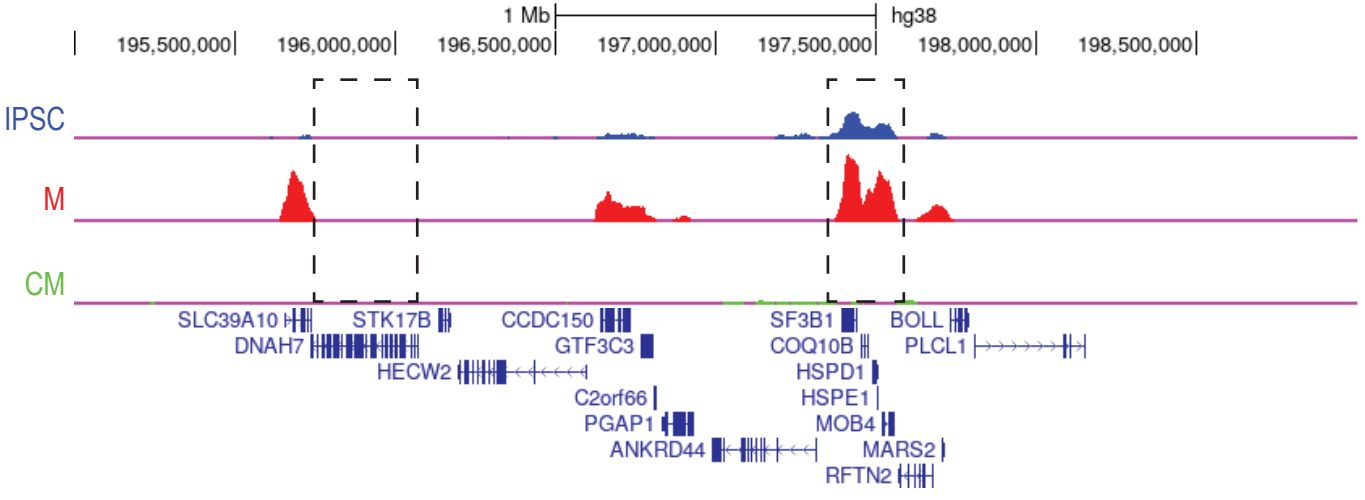



| GO Term | Benjamini |
|---|---|
| TBC | 2.10E-03 |
| Rab GTPase binding | 7.20E-03 |
| Calcium signaling pathway | 8.90E-03 |
| GTPase activation | 1.40E-02 |
| Rab-GTPase TBCdomain | 1.50E-02 |
| GTPase activator activity | 3.00E-02 |

Cardiomyopathy Genes
Non-cardiomyopathy Genes

*CACNA1C*
*CACNA1E*
*CACNA1H*
*CHRM3*
*ERBB4*
*MYLK*
*PDE1A*
*PDE1C*
*PLCB1*
*RYR2*
*TACR1*

b




c

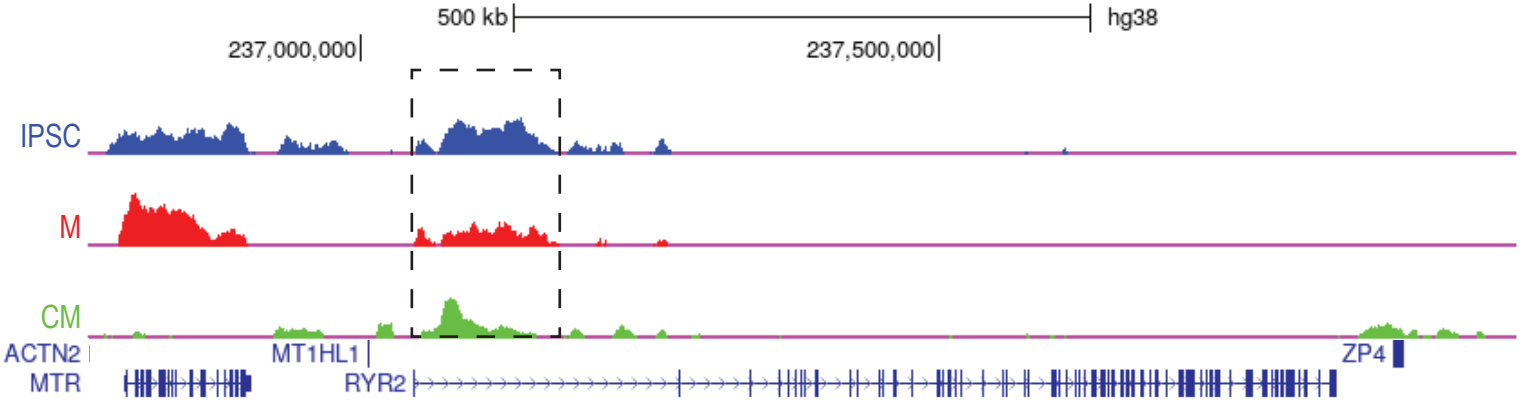


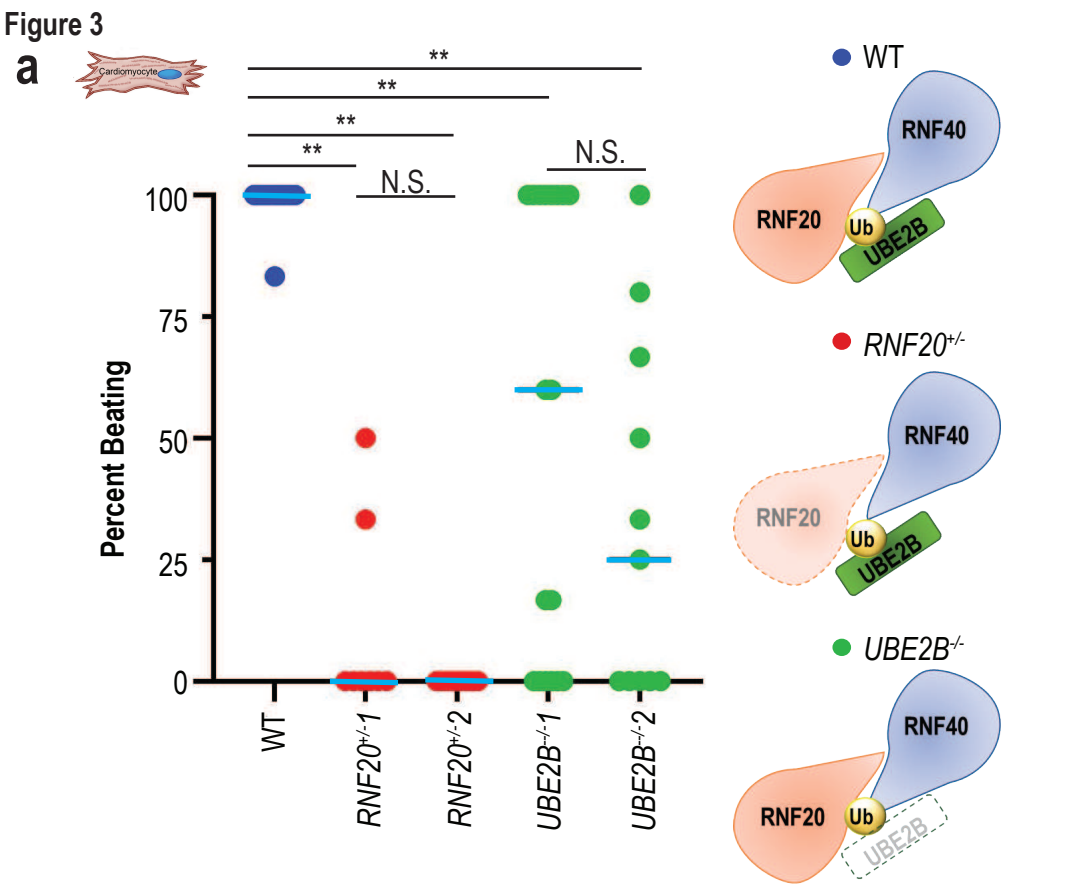
**Figure 2: H2Bub1 in iPSC-derived cardiomyocyte development shows selective maintenance of sarcomeric calcium genes**

a) The significant gene ontology terms from the genes near regions that maintain H2Bub1 between CMes and CP. Genes associated with the calcium signaling pathway are listed and are colored blue if they are associated with cardiomyopathy from patient variants and/or mouse models and are colored black if they are not.

b) Example H2Bub1 occupancy, depicted using fold enrichment against random distribution (values range from 0.95 to 3.5) across non-selectively maintained genes at three stages in CM differentiation (iPSC (blue), M (red), and CM (green)). Gene structure is indicated below the gene. The first box highlights a gene that has no H2Bub1 signal at any stage. The second box highlights a gene that has dynamic H2Bub1 signal across the stages.

c) Example H2Bub1 occupancy, depicted using fold enrichment against random distribution (values range from 0.95 to 2) across a selectively maintained genes at three stages in CM differentiation (iPSC (blue), M (red), and CM (green)). Gene structure is indicated below the gene. The selectively maintained region is indicated in the box.

Figure 3


a
Cardiomyocyte
**
**
**
**
N.S.
N.S.
100
75
50
25
0
Percent Beating
WT
RNF20+/-1
RNF20+/-2
UBE2B-/-1
UBE2B-/-2
WT
RNF40
RNF20
Ub
UBE2B
RNF20+/-
RNF40
RNF20
Ub
UBE2B
UBE2B-/-
RNF40
RNF20
Ub
UBE2B

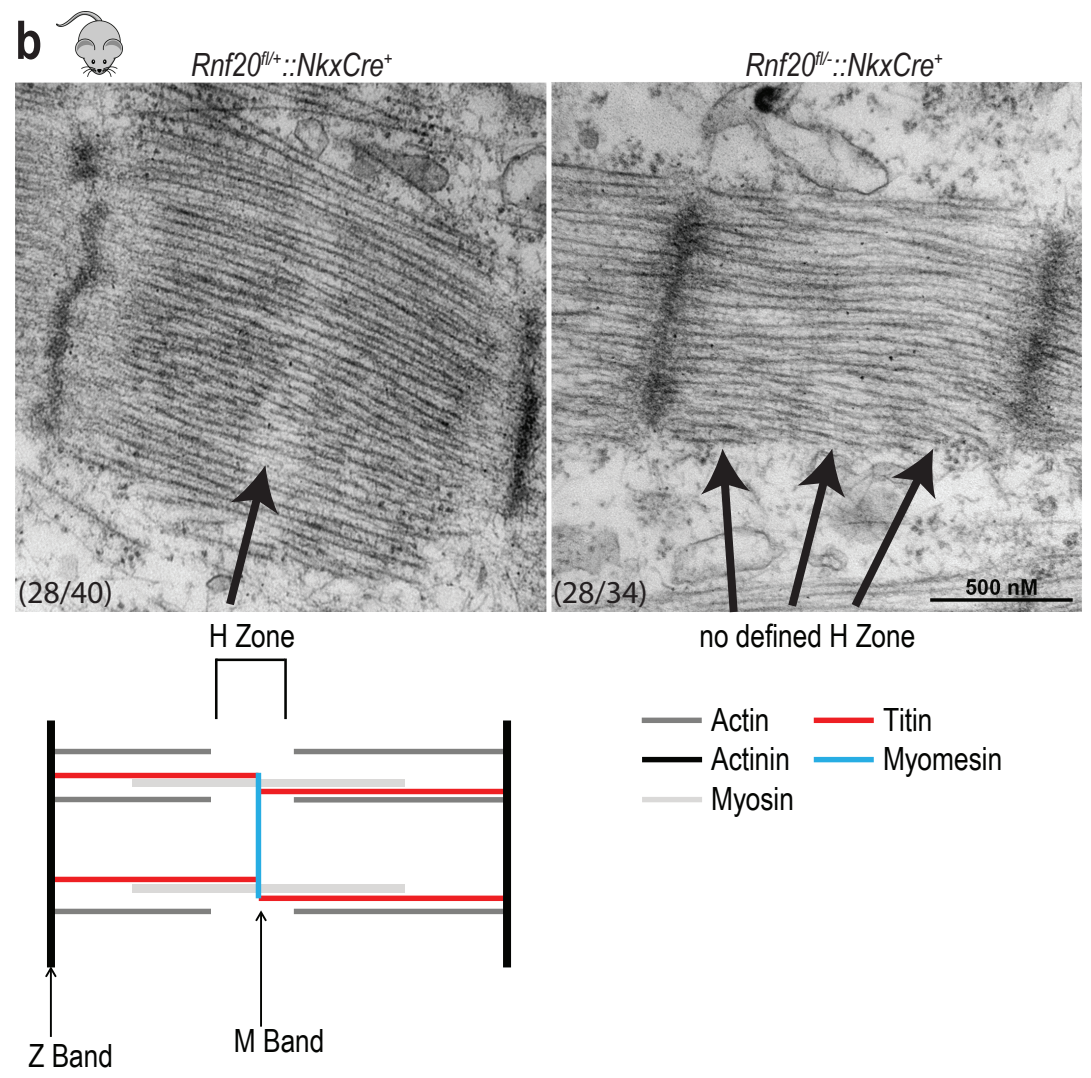

b
Rnf20fl/+::NkxCre+
Rnf20fl/-::NkxCre+
(28/40)
(28/34)
500 nM
H Zone
no defined H Zone
Actin
Titin
Actinin
Myomesin
Myosin
Z Band
M Band

**Figure 3: Reduced total H2Bub1 levels lead to abnormal cardiomyocytes**

**a)** The percent of each 6 well plate (differentiation started on the same day and same strain) of iPSC-derived cardiomyocytes that beats by day 20 is shown with colored dots (WT (n = 12) is blue, *RNF20*$^{+/-}$ (mutant 1: n= 8, mutant 2: n = 11) is red, and *UBE2B*$^{-/-}$ (mutant 1: n = 18, mutant 2: n = 11) is green). Data are shown as individual data points and a blue line representing the median. Unpaired 2-tailed, heteroscedastic t-test, * $p < 0.05$, ** $p < 0.01$, N.S. is not significant.

**b)** Transmission electron microscopy of e12.25 wild-type (*Rnf20*$^{fl/+}$::*Nkx2.5*Cre$^{+}$, n = 2) and mutant (*Rnf20*$^{flx/-}$::*Nkx2.5*Cre$^{+}$, n = 2) mouse heart sarcomeres.

Figure 4

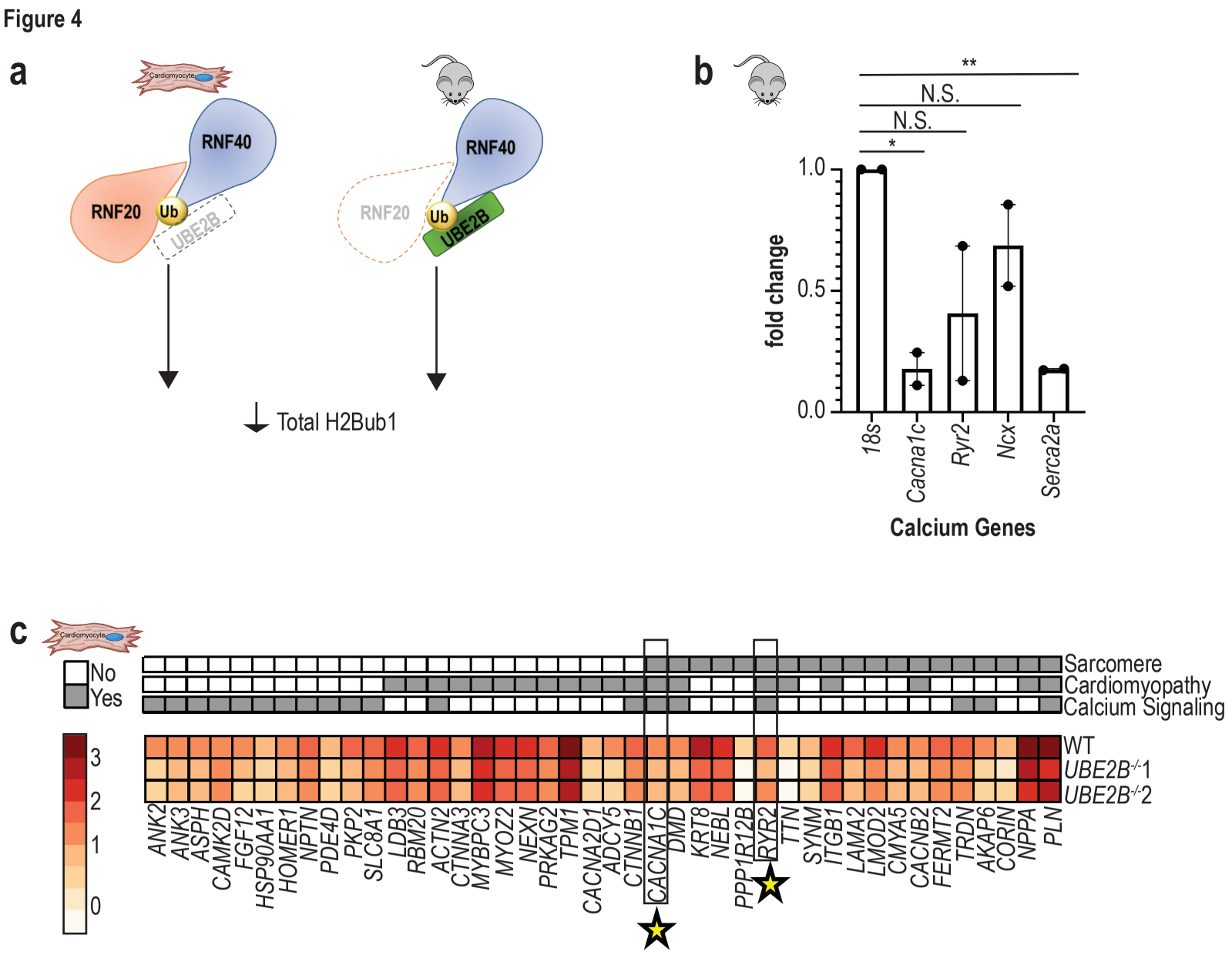

a
Cardiomyocyte
RNF40
RNF20
Ub
UBE2B
RNF40
RNF20
Ub
UBE2B
Total H2Bub1
b
**
N.S.
N.S.
*
1.0
0.5
0.0
fold change
18s
Cacna1c
Ryr2
Ncx
Serca2a
Calcium Genes
c
Cardiomyocyte
No
Yes
Sarcomere
Cardiomyopathy
Calcium Signaling
3
2
1
0
WT
UBE2B-/-1
UBE2B-/-2
ANK2
ANK3
ASPH
CAMK2D
FGF12
HSP90AA1
HOMER1
NPTN
PDE4D
PKP2
SLC8A1
LDB3
RBM20
ACTN2
CTNNA3
MYBPC3
MYOZ2
NEXN
PRKAG2
TPM1
CACNA2D1
ADCY5
CTNNB1
CACNA1C
DMD
KRT8
NEBL
PPP1R12B
RYR2
TTN
SYNM
ITGB1
LAMA2
LMOD2
CMYA5
CACNB2
FERMT2
TRDN
AKAP6
CORIN
NPPA
PLN

**Figure 4: Sarcomeric calcium signaling gene expression is reduced in cells with decreased total H2Bub1 levels**

**a)** RNF20-complex schematic illustrating the *UBE2B*$^{-/-}$ iPSC mutant and the *Rnf20*$^{fl/-}$ mouse mutant. Both of these mutations lead to decreased total H2Bub1 levels. The mouse or cell icons in (b) and (c) indicate whether the analysis was done on mouse or cells.

**b)** Quantitative RT-PCR of calcium signaling genes in e11.5 wild-type (*Rnf20*$^{fl/+}$::*Nkx2.5*Cre$^{+}$) and mutant (*Rnf20*$^{fl/-}$::*Nkx2.5*Cre$^{+}$) mouse hearts for *Cacna1c,Ryr2, Ncx*, and *Serca2a.* Levels of expression are normalized to *18s* rRNA. Individual data points are shown in black dots. Data are shown as mean ± SEM (n = 2). Unpaired 1-tailed, heteroscedastic t-test, * $p < 0.05$, ** $P < 0.01$, N.S. is non-significant.

**c)** Gene ontology analysis on genes with differing gene expression levels comparing wild-type to both independent *UBE2B*$^{-/-}$ cell lines at the cardiomyocyte stage. Many of these genes are related to sarcomere, cardiomyopathy, and/or calcium signaling. Expression levels for example genes in each category are shown. Stars indicate genes shared with maintained H2Bub1 marks upon transition to CPs shown in Fig. 2a. Data represent three RNA-seq replicates of each of the 2 cell lines. Gene ontology (upper panel) is in gray scale; mRNA expression data (lower panel) is shown as a heat map.

Figure 5

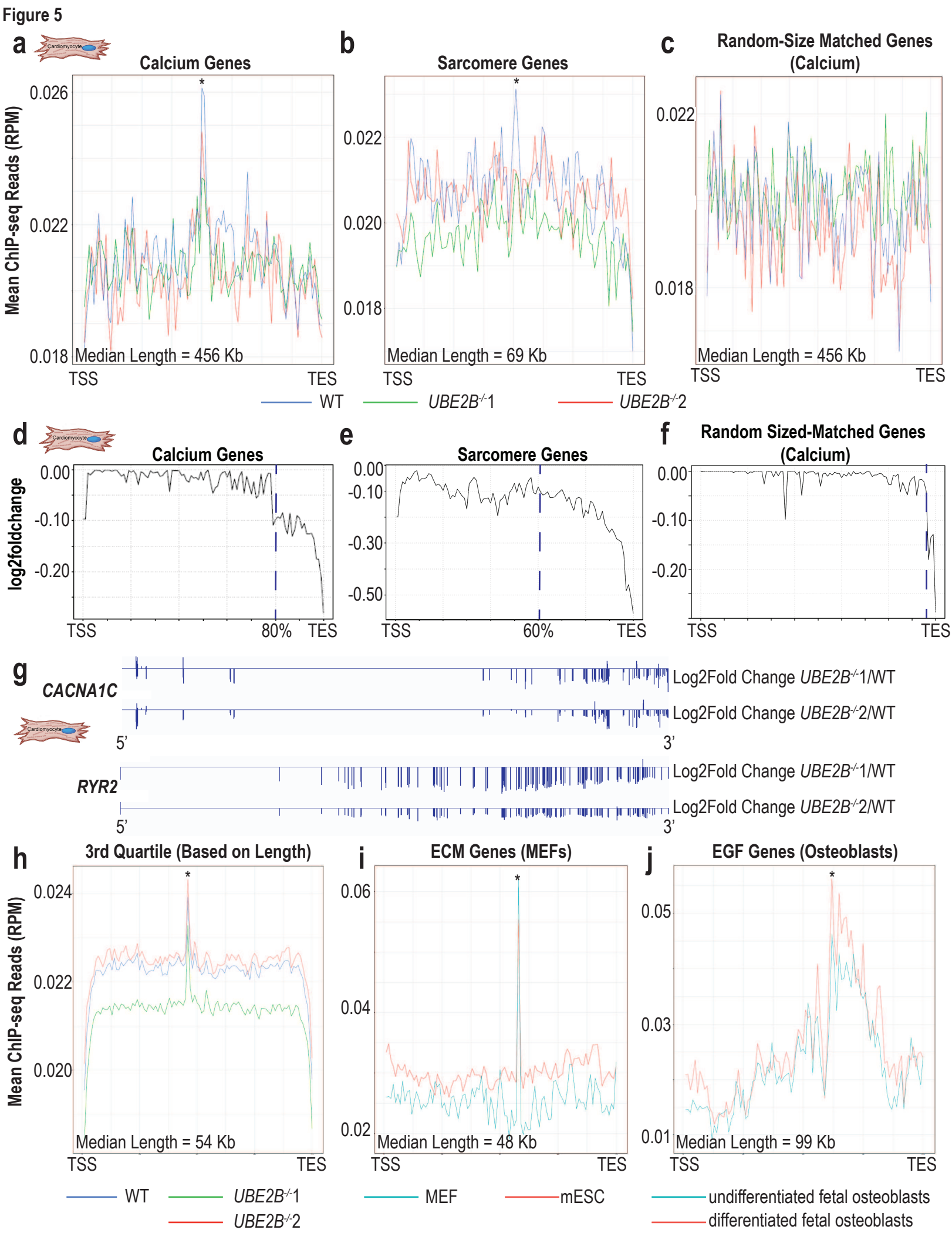

a
Cardiomyocyte
Calcium Genes
Mean ChIP-seq Reads (RPM)
0.026
0.022
0.018
Median Length = 456 Kb
TSS
TES
b
Sarcomere Genes
0.022
0.020
0.018
Median Length = 69 Kb
TSS
TES
c
Random-Size Matched Genes (Calcium)
0.022
0.018
Median Length = 456 Kb
TSS
TES
WT
UBE2B-/-1
UBE2B-/-2
d
Cardiomyocyte
Calcium Genes
log2foldchange
0.00
-0.10
-0.20
TSS
80%
TES
e
Sarcomere Genes
0.00
-0.10
-0.30
-0.50
TSS
60%
TES
f
Random Sized-Matched Genes (Calcium)
0.00
-0.10
-0.20
TSS
TES
g
CACNA1C
Cardiomyocyte
Log2Fold Change UBE2B-/-1/WT
Log2Fold Change UBE2B-/-2/WT
5'
3'
RYR2
Log2Fold Change UBE2B-/-1/WT
Log2Fold Change UBE2B-/-2/WT
5'
3'
h
3rd Quartile (Based on Length)
Mean ChIP-seq Reads (RPM)
0.024
0.022
0.020
Median Length = 54 Kb
TSS
TES
i
ECM Genes (MEFs)
0.06
0.04
0.02
Median Length = 48 Kb
TSS
TES
j
EGF Genes (Osteoblasts)
0.05
0.03
0.01
Median Length = 99 Kb
TSS
TES
WT
UBE2B-/-1
UBE2B-/-2
MEF
mESC
undifferentiated fetal osteoblasts
differentiated fetal osteoblasts

**Figure 5: Accumulation of H2Bub1 near the center of tissue specific genes is correlated with enhanced efficiency of transcriptional elongation**

**a)** Metagene plot for H2Bub1 levels in wild-type (blue) and *UBE2B*$^{-/-}$ mutants (green and red) across the calcium signaling genes that are differentially expressed between wild-type and *UBE2B*$^{-/-}$ mutants and/or selectively maintained between cardiac mesoderm and cardiac progenitor stages (n = 3, for each of 2 cell lines). The * indicates the accumulation near the center of the gene.

**b)** Metagene plot for H2Bub1 levels in wild-type (blue) and *UBE2B*$^{-/-}$ mutants (green and red) across the sarcomeric genes that are differentially expressed between wild-type and *UBE2B*$^{-/-}$ mutants (n = 3, for each of 2 cell lines). The * indicates the accumulation near the center of the gene.

**c)** Example metagene plot for H2Bub1 levels in wild-type (blue) and *UBE2B*$^{-/-}$ mutants (green and red) across "randomly" selected quantity and size-matched genes to the calcium gene set (n = 3, for each of 2 cell lines). 30 "random" metagene plots were created from the calcium gene set and 30 "random" metagene plots were created from the sarcomere gene set: 10 from genes that are upregulated between wild-type and both *UBE2B*$^{-/-}$ cell lines, 10 that are non-regulated between wild-type and both *UBE2B*$^{-/-}$ cell lines, and 10 that are down-regulated between wild-type and both *UBE2B*$^{-/-}$ cell lines.

**d)** Log2 of fold change in transcript abundance between wild-type and *UBE2B*$^{-/-}$ mutants is shown at each position along the genes. The genes being shown are the calcium signaling genes describe in (a).

**e)** Log2 of fold change in transcript abundance between wild-type and *UBE2B*$^{-/-}$ mutants is shown at each position along the genes. The genes being shown are the sarcomeric genes described in (b).

**f)** Log2 of fold change in transcript abundance between wild-type and *UBE2B*$^{-/-}$ mutants is shown at each position along the genes. The genes being shown are "randomly" selected quantity and size-matched genes to the calcium signaling genes described in (c).

**g)** Example log2 of fold change gene traces for calcium genes (*CACNA1C* and *RYR2*) that have accumulation of H2Bub1 near the center of the gene and decreased transcriptional elongation efficiency. 5' on the left of the diagram.

**h)** Metagene plot for H2Bub1 levels in wild-type (blue) and *UBE2B*$^{-/-}$ mutants (green and red) across all of the 3$^{rd}$ Quartile genes (greater than 33.940 Kb and less than 93.323 Kb) (n = 3). The * indicates the accumulation near the center of the gene.

**i)** Metagene plot for H2Bub1 levels in MEFs (teal) and mESCs (red) across the ECM genes that are differentially expressed between MEFs and mESCs. The * indicates the accumulation near the center of the gene.

**j)** Metagene plot for H2Bub1 levels in undifferentiated hFOBs (teal) and differentiated hFOBs (red) across the EGF related genes that are differentially expressed between undifferentiated hFOBs and differentiated hFOBs. The * indicates the accumulation near the center of the gene.

Figure 6

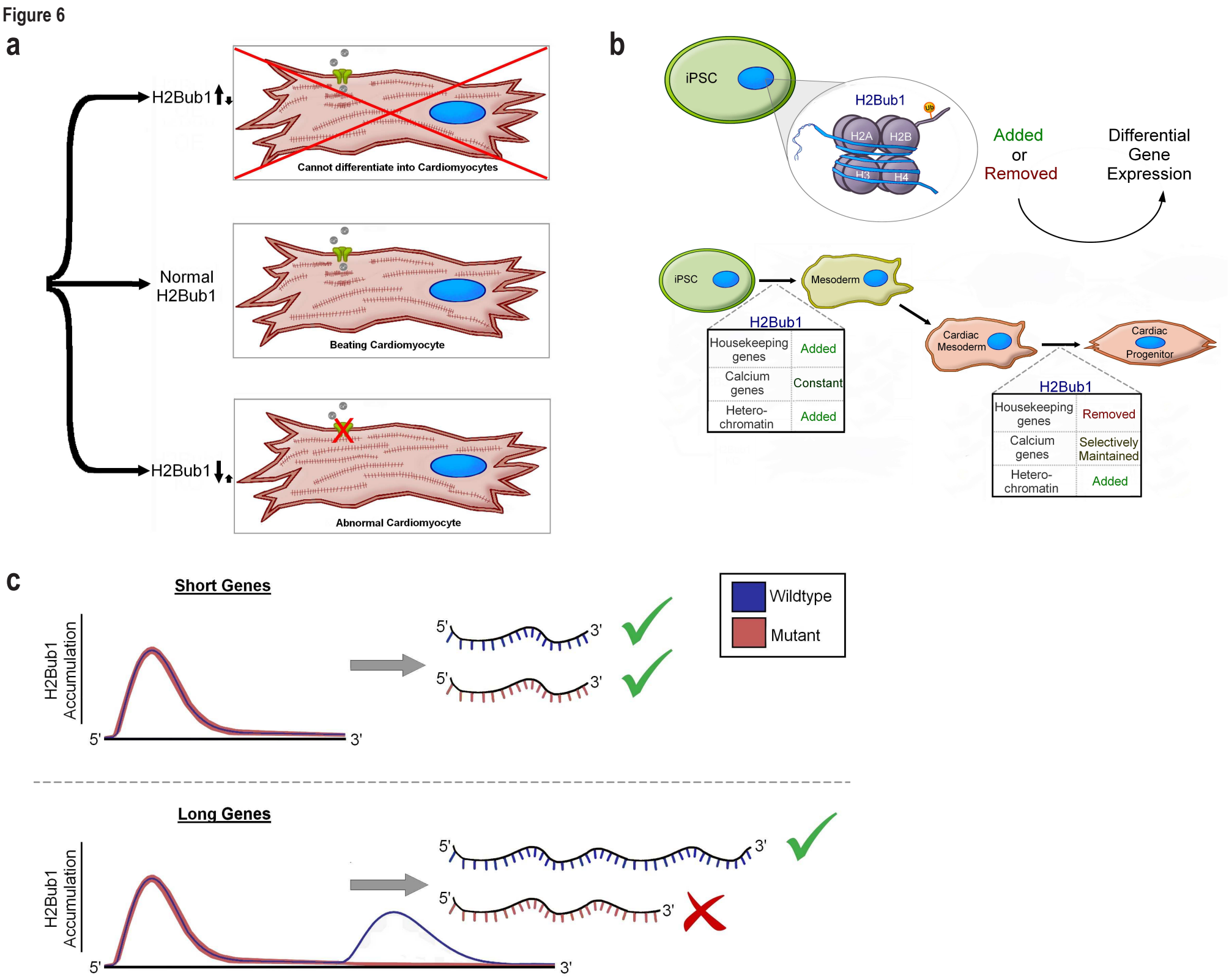

a
H2Bub1
Cannot differentiate into Cardiomyocytes
Normal H2Bub1
Beating Cardiomyocyte
H2Bub1
Abnormal Cardiomyocyte
b
iPSC
H2Bub1
Ub
H2A
H2B
H3
H4
Added or Removed
Differential Gene Expression
iPSC
Mesoderm
H2Bub1
Housekeeping genes
Added
Calcium genes
Constant
Hetero-chromatin
Added
Cardiac Mesoderm
Cardiac Progenitor
H2Bub1
Housekeeping genes
Removed
Calcium genes
Selectively Maintained
Hetero-chromatin
Added
c
Short Genes
H2Bub1 Accumulation
5'
3'
5'
3'
5'
3'
Wildtype
Mutant
Long Genes
H2Bub1 Accumulation
5'
3'
5'
3'
5'
3'

**Figure 6: Working Model**

**a)** When H2Bub1 levels are misregulated by altering levels of complex components, cardiomyocytes do not form normally. When total H2Bub1 levels are increased by decreased RNF20, cardiomyocytes do not form. When total H2Bub1 levels are decreased by creating *UBE2B*$^{-/-}$ or *Rnf20*$^{fl/-}$, sarcomeres do not form normally and calcium signaling genes have reduced expression.

**b)** During wild-type CM differentiation, housekeeping genes and heterochromatic regions have dynamic H2Bub1 levels. H2Bub1 is sparsely maintained during the transition from cardiac mesoderm to cardiac progenitor. Notably, selectively maintained genes are enriched for sarcomeric calcium genes.

**c)** Genes that are short in length have an increase in H2Bub1 at the 5' end of the gene that decreases towards the 3' end. However, longer tissue-specific genes have an accumulation in H2Bub1 near the center of the gene in wild-type cells and not in *UBE2B*$^{-/-}$. In the mutants when this accumulation is absent, the transcriptional efficiency is reduced.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementwithFigures.pdf
- SupplementalData1WildTypeChIPseq.xlsx
- SupplementalData2WildTypeRNAseq.xlsx
- SupplementalData3RNF20ChIPseq.xlsx
- SupplementalData4RNF20RNAseq.xlsx
- SupplementalData5UBE2BChIPseq.xlsx
- SupplementalData6UBE2BRNAseq.xlsx
- Movie1wildtype.mov
- Movie2RNF201Beating.mov
- Movie3RNF201NotBeating.mov
- Movie4RNF202NotBeating.mov
- Movie5UBE2B1Beating.mov
- Movie6UBE2B1NotBeating.mov
- Movie7UBE2B2Beating.mov
- Movie8UBE2B2NotBeating.mov