Massively parallel in vivo CRISPR screening identifies RNF20/40 as epigenetic regulators of cardiomyocyte maturation

Nathan J. VanDusen1, Julianna Y. Lee1, Weiliang Gu1,2, Catalina E. Butler1, Isha Sethi1, Yanjiang Zheng1,3, Justin S. King1, Pingzhu Zhou1, Shengbao Suo4, Yuxuan Guo1, Qing Ma1, Guo-Cheng Yuan5 & William T. Pu1,6

The forward genetic screen is a powerful, unbiased method to gain insights into biological processes, yet this approach has infrequently been used in vivo in mammals because of high resource demands. Here, we use in vivo somatic Cas9 mutagenesis to perform an in vivo forward genetic screen in mice to identify regulators of cardiomyocyte (CM) maturation, the coordinated changes in phenotype and gene expression that occur in neonatal CMs. We discover and validate a number of transcriptional regulators of this process. Among these are RNF20 and RNF40, which form a complex that monoubiquitinates H2B on lysine 120. Mechanistic studies indicate that this epigenetic mark controls dynamic changes in gene expression required for CM maturation. These insights into CM maturation will inform efforts in cardiac regenerative medicine. More broadly, our approach will enable unbiased forward genetics across mammalian organ systems.
At birth, mammalian CMs undergo maturation, a dramatic and coordinated set of structural, metabolic, and gene expression changes that enable them to sustain billions of cycles of forceful contraction during postnatal life1,2. Fetal CMs are glycolytic, mononuclear, mitotic, and contract against low resistance, whereas adult CMs rely on oxidative phosphorylation, are diploid and postmitotic, and support heart growth by increasing in size (maturational hypertrophy). Sarcomeric and ultrastructural adaptations, such as plasma membrane invaginations known as T-tubules, facilitate coordinated and forceful CM contraction against high resistance. Unfortunately, CMs induced from stem cells or other nonmyocyte sources resemble fetal CMs and lack the hallmark features of mature, adult CMs3,4. This “maturation bottleneck” remains a major barrier to using stem cell-derived CMs for disease modeling or therapeutic cardiac regeneration.

The regulatory mechanisms that govern the diverse facets of CM maturation are poorly understood, in large part due to the lack of a suitable in vitro model and challenges associated with in vivo approaches. Although mosaic gene manipulation strategies have allowed more precise interpretation of in vivo experiments with respect to the regulation of maturation5-8, the low throughput of standard in vivo approaches remains a major barrier. To overcome this obstacle, we sought to perform an in vivo forward genetic screen in mice. The resource intensity of such a screen and the lack of available barcoding or index pooling technologies have limited throughput of standard in vivo approaches. Although mosaic gene manipulation strategies have allowed more precise interpretation of in vivo experiments with respect to the regulation of maturation5-8, the low throughput of standard in vivo approaches remains a major barrier. To overcome this obstacle, we sought to perform an in vivo forward genetic screen in mice. The resource intensity of traditional forward genetics has precluded their widespread use in mammals, but Cas9 mutation directed by a library of guide RNAs (gRNAs) makes introduction and recovery of gene mutations highly efficient9-11, and can be expeditiously deployed in mammals in vivo5,12,13. This capability has been used for forward genetic screens in cultured cells11,14,15, but its ability to interrogate endogenous biological processes in mammals in vivo has yet to be fully realized.

Here we report development of an in vivo forward genetic screen to discover transcriptional and epigenetic regulators of CM maturation. Our screen identified several candidate regulators, including RNF20 and RNF40, genes implicated in congenital heart disease16,17 that form an enzyme complex that mono-ubiquinates histone H2B on lysine 12018.

Results

Design of in vivo CRISPR screen for CM maturation factors.

We developed an in vivo forward genetic screen to discover factors that regulate murine CM maturation (Fig. 1a). We employed CRISPR/Cas9 AAV9 (CASAAV)-based somatic mutagenesis5 and a gRNA library targeting murine transcriptional regulators to create thousands of distinct mutations within different CMs of a single mammalian heart. We screened these mutant CMs with a flow cytometry-based single-cell assay of CM maturation. Sequencing of gRNAs from immature CMs compared with the input library identified gRNAs enriched or depleted in immature CMs, i.e., gRNAs that target genes that cells autonomously promote or antagonize CM maturation, respectively.

To separate individual CMs with immature and mature phenotypes, we took advantage of developmentally regulated sarcomere gene isoform switching, a hallmark of CM maturation1,9. In mouse, sarcomere maturation, an essential driver of the overall CM maturation program7, involves silencing of myosin heavy chain 7 (Myh7) and upregulation of Myh6. In Myh7<sup>YFP</sup> mice, YFP is fused to endogenous Myh7, such that YFP fluorescence is controlled by endogenous Myh7 regulatory elements. We hypothesized that Myh7<sup>YFP</sup> could be used as a single-cell readout of CM-maturation state. Examination of Myh7<sup>YFP</sup> myocardium from neonatal and adult mice verified that neonatal CMs exhibited strong YFP fluorescence, whereas YFP was nearly completely silenced in adult CMs (Fig. 1b). Since mutation of both Gata4 and Gata6 impairs CM maturation (Suppl. Fig. 1, Suppl. Data 1, and ref. 19), we next tested the effect of Gata4/Gata6 mutation on Myh7<sup>YFP</sup> expression. We found that CASAAV-Gata4/6, which expresses CM specific Cre and previously validated gRNAs targeting Gata4 and Gata6, markedly increased the fraction of transduced CMs that expressed YFP from 11% to 86% (Fig. 1c). These data show that Gata4/6 mutation inhibits normal CM maturation, including maturational Myh7 silencing, and supported the use of Myh7<sup>YFP</sup> as a single-cell reporter of cardiomyocyte-maturation state.

Because the coordination of CM maturation suggested transcriptional regulation2, we focused our screen on transcriptional regulators. We developed a pooled CASAAV library comprising AAV-expressing CM-specific Cre and a gRNA designed to target a candidate gene. The candidate gene list (Suppl. Data 2) contained 1894 transcriptional regulators (transcription factors and epigenetic modifiers). Given that Gata4/6 regulate multiple aspects of CM maturation (Suppl. Fig. 1), we also included 259 genes differentially expressed in P6 Gata4/6 high-dose double-KO CMs, and 291 genes with strong adjacent G4 binding at P0 (Suppl. Data 1, 2). In total, 2444 genes were selected for targeting. Using a computational pipeline designed to optimize gRNA on-target activity and yield of frameshift mutations22,23 we designed six guides for each gene. Seven gRNAs that do not target the mouse genome were also included as negative controls. These 14675 gRNAs were synthesized as an oligonucleotide pool, cloned into the CM specific CASAAV vector, and packaged into AAV9 (see “Methods”). To introduce a positive control, a small amount of the CASAAV-Gata4/6 vector was spiked into the AAV pool.

Execution of in vivo CRISPR screen for CM-maturation factors.

We subcutaneously injected the AAV library into newborn R26<sup>Cas9-GFP/+;</sup>Myh7<sup>YFP</sup> pups at a dose sufficient to transduce approximately 50% of the myocardium. CMs were isolated from mice at four weeks of age. YFP<sup>+</sup> CMs were isolated from the input pool by flow cytometry (Suppl. Fig. 2a). gRNAs in YFP<sup>+</sup> and input pools were quantified by next generation sequencing. We analyzed 45 hearts in pools of three, resulting in 15 input and 15 YFP<sup>+</sup> samples. In total, four million transduced CMs were analyzed, of which 365,000 were YFP<sup>+</sup>. In total, 11 YFP<sup>+</sup> and 14 input samples passed quality control (Suppl. Fig. 2) and separated into distinct clusters (Fig. 1d). In total, 834 gRNAs were significantly enriched within YFP<sup>+</sup> samples (adj. < 0.001; Suppl. Data 3), with the Gata4 and Gata6 positive control gRNAs being among the most enriched (Fig. 1e). The seven human-targeting negative control gRNAs did not show enrichment (Fig. 1f). We used MaGeCK24 to consolidate the six gRNA enrichment scores per gene into a single score. gRNAs targeting 121 genes were significantly enriched within YFP<sup>+</sup> CMs, while gRNAs targeting 148 genes were depleted (P < 0.05; Suppl. Data 3). The top-ranked depleted gene was Myh7 (Fig. 1g), which was expected given that YFP fused to Myh7 was the screen readout. Among the enriched genes were thyroid hormone receptor alpha (Thra) and nucleolin (Ncl), established regulators of maturation3,32,36, as well as many novel candidates (Fig. 1h; Suppl. Data 3), including both Rnf20 and Rnf40, which encode components of an epigenetic complex that mono-ubiquitinitates histone 2B27,38.

Validation of candidate factors identified by genetic screen.

To validate the top 10 most enriched candidate genes (Fig. 2a), the most highly enriched gRNA for each factor was cloned into separate CASAAV vectors. These were used to deplete individually each factor at birth. To focus on cell-autonomous effects
and avoid secondary effects related to organ dysfunction\textsuperscript{5–7}, we used an AAV dose that transduced a small fraction of CMs. Among the transduced (GFP\textsuperscript{+}) cells, we observed robust upregulation of Myh7\textsuperscript{YFP} for seven of the ten candidates: Rnf40, Rnf20, Taf3, Taf2, Thra, Zbtb7a, and Ncl (Fig. 2a, b; Suppl. Fig. 3a). We focused subsequent analyses on these seven target genes. Of the three candidates that did not validate, one (Poc1b) was a false positive as it did not increase YFP expression. The remaining two (Setd6 and Eif3i) caused modest but significant persistence of YFP. Notably, all three of these candidates had less than three gRNAs with DESeq2 enrichment p-value <0.001, suggesting an additional criterion that could improve screen stringency.

We analyzed the effect of individual CASAAV vectors on additional hallmarks of maturation—CM nucleation, size, and T-tubulation. The seven CASAAV vectors, or AAV-Cre lacking gRNA (control), were delivered individually to newborn mice at a mosaic dose. At one month of age, CMs were dissociated, fixed, and stained to visualize T-tubules (CAV3) and nuclei (DAPI). The effects of CASAAV vectors on transduced (GFP\textsuperscript{+}) CMs were compared with AAV-Cre for three additional parameters of CM maturation: multinucleation, maturational hypertrophy, and T-tubulation. Three of the seven CASAAV vectors, Rnf40, Taf3, and Taf2, negatively impacted all maturation parameters (Fig. 2c–f; Suppl. Fig. 3a–e). CASAAV-Rnf20 impaired maturational growth and T-tubulation, but the increase in mononucleation did not reach statistical significance. Zbtb7a impaired T-tubulation, but did not influence maturational hypertrophy or mononucleation. These results validated the requirement of seven of the 10 candidates for multiple facets of CM maturation, and also demonstrate that the overall maturational program can be separated into independently regulated, dissociable subprograms\textsuperscript{7}.

RNF20/40 are required for regulation of CM maturation. Rnf20 and Rnf40 were two of the most enriched genes in the screen, and their depletion broadly impaired CM maturation. These genes encode subunits of an E3 ubiquitin ligase that monoubiquitinates histone 2B at lysine 120 (H2Bub1)\textsuperscript{18,29}, an epigenetic mark that regulates the expression of a subset of genes and that has been implicated in developmental transitions and cancer\textsuperscript{30}. Human genetic studies implicate de novo RNF20/RNF40 mutations in congenital heart disease\textsuperscript{16,17}. The postnatal cardiac functions of these genes have not been studied. For these reasons, we
investigated the mechanisms by which RNF20/40 regulate CM maturation.

We investigated the expression profile and transcriptional regulation of Rnf20/40. Mining of a single-cell RNA-seq data set from the adult mouse heart revealed that Rnf20 and Rnf40 are expressed in CMs and several other cell populations in the adult mouse heart (Suppl. Fig. 4a–c).31,32 In cardiomyocytes, Rnf20 and Rnf40 transcript levels were similar between neonatal and adult stages (Suppl. Fig. 4d). Examination of our previously reported core cardiac transcription factor ChIP-seq datasets33 demonstrated strong binding of multiple transcription factors, including GATA4, NKX2–5, and SRF, at the promoters of both genes (Suppl. Fig. 4e, f). Consistent with our observation of stable postnatal gene expression, we did not observe significant differences in transcription factor binding at either locus between fetal and adult stages. These results could suggest developmental post-transcriptional regulation of RNF20/40 function, developmental regulation of cofactors, or developmental regulation of RNF20/40 recruitment to regulate genes.

We created a single CASAAV vector (CASAAV–RNF20/40) containing gRNAs that target both factors and validated that it depleted RNF20/40 (RNF Cas-KO) and H2Bub1 (Fig. 3a, b).

H2Bub1 was more depleted than RNF20/40, possibly because some CRISPR mutations in Rnf20 or Rnf40 might impair activity but not immunoreactivity, and because RNF20 and RNF40 are likely both required for enzymatic activity to write H2Bub1.18. When given at a dose that transduced most CMs (“full dose”), CASAAV-RNF20/40 caused cardiac dysfunction, dramatic elevation of Myh7YFP expression, and death (Fig. 3c–e). At a dose that transduced a small fraction of CMs (“mosaic dose”), RNF20/40 deletion cell autonomously impaired CM maturation (Fig. 3f–n).

As CMs mature, they become polyploid, making simultaneous mutagenesis of all alleles by CASAAV-based indel formation more difficult. To bypass this difficulty, and further validate our CASAAV-based findings, we acquired a conditional Rnf20 allele, Rnf20flox. Injection of Rnf20flox/R26lYFP newborn pups with a mosaic dose of AAV-TnT-Cre resulted in robust depletion of H2Bub1 within transduced CMs (Rnf20 KO; Suppl. Fig. 5a, b). Subsequent analysis of CM area, length, width, and T-tubule organization revealed an immature phenotype consistent with our CASAAV-based observations (Suppl. Fig. 5c–i). To determine if the dependence of mature CM phenotype on RNF20 is developmental stage specific, we treated Rnf20flox/R26lTomato

Fig. 2 Validation of top candidates from the forward genetic screen. a Top ten candidates ranked by MAGeCK enrichment score (lower rank indicates greater enrichment). The six guides for each candidate are ranked by individual enrichment. Enriched gRNAs with Benjamini–Hochberg corrected Wald test P < 0.001 are colored in red and summed in the first column. Candidates were tested by treating R26Cas9-cre, Myh7YFP neonatal pups with the most highly ranked individual gRNA. AAV-Cre without gRNA served as negative control. Presence (column 2, yellow box) or absence (gray box) of persistent YFP expression at four weeks was assessed by imaging fixed, dissociated CMs. b Quantification of Myh7YFP activation within transduced (GFP +) CMs by flow cytometry. AAV-Cre was used as negative control. n = 3. For all genes, except Poc1b and Setd6, was <0.0001. Setd6 P = 0.0111. c Quantification of mononuclearity among dissociated, GFP + control CMs or YFP + candidate-depleted CMs. n = 3. Significant P: Rnf40, 0.0128; Thra, 0.0102; Taf3, 0.0027; Taf2, 0.0001. Error bars denote standard error. d–f Normalized projected area (d) and length:width ratio (e) of control and candidate-depleted CMs. Measurements from dissociated YFP + CMs were normalized to YFP control CMs from the same heart. Significant P: area: Rnf40, Thra, Taf3, Rnf20, Taf2, and Ncl <0.0001. Significant P, length:width: Taf3, 0.007; Rnf20, 0.0042. f T-tubule transverse element density for control and candidate-depleted CMs as measured by AutoTT software-based quantification of CAV3 immunostaining. Significant P: Rnf40, Taf3, Rnf20, Taf2, Ncl <0.0001; Zbtb7a, 0.0003. Bar plots show mean ± SD. At least 69 CMs from three mice were measured per group. Violin plots: shape indicates data distribution; point, median; whiskers, start at quartile and extends 1.5 times the interquartile distance. Dunnett’s two-tailed t-test vs. Cre control: *P < 0.05, **P < 0.001. Source data are provided as a Source Data file.

ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-24743-z | www.nature.com/naturecommunications
mice with a mosaic dose of AAV-TnT-Cre at eight weeks of age and dissociated CMs for analysis at 12 weeks. Area, length, and width of CMs were unaffected by mosaic RNF20 KO (Suppl. Fig. 5j–l). In contrast, T-tubule organization was disrupted, with KO CMs demonstrating decreased transverse element density and increased longitudinal element density (Suppl. Fig. 5m–p). These data indicate that RNF20 is necessary in adult CMs for maintenance of some features of maturation (T-tubules) but not others (size and overall morphology).

**Transcriptional regulation of CM maturation by RNF20/40.** To measure the effect of RNF20/40 depletion on gene expression during maturation, we conducted transcriptional profiling. Newborn R26Cas9-2A-GFP/+;Myh7YFP/+ mouse pups were injected with a mosaic dose of CASAAV-RNF20/40 vector or a control vector containing Cre without gRNAs. At P28, hearts were dissociated and flow cytometry was used to recover YFP+ CMs transduced with CASAAV-RNF20/40 or control GFP+ CMs transduced with Cre. We analyzed the transcriptomes of each
immunoprecipitation followed by next-generation sequencing (ChIP-seq) to measure H2Bub1 chromatin occupancy at neonatal (first week of life) and mature (four-week-old) stages in mouse heart apex tissue. Biological duplicates at each stage correlated well with one another (Fig. 5a and Suppl. Fig. 7a). Consistent with prior reports\(^39,40\), H2Bub1 predominantly occupied gene bodies, with greater density toward promoters (Fig. 5a, b, Suppl. Fig. 7b). We quantified H2Bub1 density on gene bodies (Suppl. Data 6). The vast majority (>88%) of genes with detectable H2Bub1 signal were shared between time points (Fig. 5c), although the signal intensity of a dynamic subset varied between developmental stages (Fig. 5a, d). Consistent with H2Bub1 being an activating mark\(^41-43\), at both stages, genes with greater H2Bub1 generally were more highly expressed, although the correlation was poor (\(r = 0.07 - 0.09\); Suppl. Fig. 7c, d). Interestingly, genes with stronger H2Bub1 signal at P1 (or P28) were less likely to be differentially expressed between these stages (Fig. 5e, top plot; Suppl. Fig. 7e), consistent with observations in murine embryonic fibroblasts\(^44\). Gene sets enriched by greater H2Bub1 occupancy included oxidative phosphorylation and TCA cycle (Fig. 5g). We next analyzed the RNA-seq data to identify biological processes enriched within the RNF Cas-KO differentially expressed genes (Suppl. Table 2). Many metabolism-related gene sets were enriched within the downregulated genes (Fig. 4d), including fatty acid metabolism and PPAR signaling, two pathways strongly associated with CM maturation\(^35,38\). Upregulated genes were enriched for a greater diversity of biological processes, with spliceosome-associated genes being most enriched (Fig. 4f).

**RNF20/40 regulation of the maturational epigenetic landscape.** The RNF20/40 complex is an E3 ubiquitin ligase that mono-ubiquitines H2B on lysine 120 (H2Bub1). To determine the relationship between H2Bub1 written by RNF20/40 and normal maturation of CM gene expression, we used chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) to measure H2Bub1 chromatin occupancy at neonatal (first week of life) and mature (four-week-old) stages in mouse heart apex tissue. Biological duplicates at each stage correlated well with one another (Fig. 5a and Suppl. Fig. 7a). Consistent with prior reports\(^39,40\), H2Bub1 predominantly occupied gene bodies, with greater density toward promoters (Fig. 5a, b, Suppl. Fig. 7b). We quantified H2Bub1 density on gene bodies (Suppl. Data 6). The vast majority (>88%) of genes with detectable H2Bub1 signal were shared between time points (Fig. 5c), although the signal intensity of a dynamic subset varied between developmental stages (Fig. 5a, d). Consistent with H2Bub1 being an activating mark\(^41-43\), at both stages, genes with greater H2Bub1 generally were more highly expressed, although the correlation was poor (\(r = 0.07 - 0.09\); Suppl. Fig. 7c, d). Interestingly, genes with stronger H2Bub1 signal at P1 (or P28) were less likely to be differentially expressed between these stages (Fig. 5e, top plot; Suppl. Fig. 7e), consistent with observations in murine embryonic fibroblasts\(^44\). Gene sets enriched by greater H2Bub1 occupancy included oxidative phosphorylation and TCA cycle (Suppl. Fig. 7f-g).

Although gene expression and H2Bub1 occupancy at each stage were not well correlated, there was a significant correlation between the change in H2Bub1 and the change in gene expression between stages (\(r = 0.39; P < 0.0001\); Fig. 5e, bottom plot). Genes upregulated in RNF Cas-KO (red, Fig. 5f and Suppl. Fig. 7e) overall were more highly expressed in immature cardiomyocytes, and conversely those downregulated in RNF Cas-KO (blue, Fig. 5f and Suppl. Fig. 7e) were overall more highly expressed in mature cardiomyocytes. Among genes that were up- or downregulated by more than 2-fold in RNF20/40 depletion, 295 genes gained H2Bub1 and increased expression during maturation (upper right quadrant, Fig. 5f). Of these, 218 were downregulated by inactivation of RNF20/40 (blue color versus 398 of 985 genes in the other three quadrants: Fisher exact test \(P < 0.00001\)). Among these 218 genes were Myh6 and Tnni3, whose well-established isoform switches are cardinal features of maturation\(^35\). These 218 genes were highly enriched for functional terms related to metabolism (Fig. 5g, right). Conversely, genes that lost H2Bub1

---

**Fig. 4 Transcriptional profiling of mosaics RNF20/40-depleted CMs.** a Experiment overview. b Principal component analysis of RNA-seq replicates. c Differentially expressed genes. Red indicates Benjamini–Hochberg-corrected Wald test \(P < 0.05\). d, e Gene set enrichment analysis with custom gene sets for “Adult Specific Genes” (d) or “Neonatal Specific Genes” (e). RNF20/40-depleted cells were depleted for adult-specific genes and enriched for adult-specific genes. f Gene set enrichment analysis of indicated KEGG pathway gene sets, using genes ranked by their expression in RNF20/40-depleted CMs compared with control CMs. GSEA nominal \(P\)-val is not corrected for set size or multiple-hypothesis testing, the Normalized Enrichment Score. Sourcedata are provided as a Source Data file.
and decreased expression during maturation were highly enriched for upregulation by RNF20/40 inactivation (167 of 261 in lower left quadrant of Fig. 5f, compared with 497 of 1019 in the other three quadrants; Fisher exact test \( P < 0.00001 \)). Gene ontology analysis of these 167 genes showed that they were enriched for diverse cell-signaling functional terms, such as protein kinase A signaling and cardiac hypertrophy signaling (Fig. 5g, left).

H3K4me3 and H3K36me3 are epigenetic marks that correlate with H2Bub1. Although H2Bub1 was required for H3K4me3 modification in yeast, this effect was less pronounced and confined to specific genes in mammalian cells. To study the effect of RNF20 inactivation on H3K4me3 and H3K36me3 chromatin occupancy in maturing CMs, we treated Rnf20^ex/^-neonates with a full dose of AAV-TnT-Cre to inactivate Rnf20 in the large majority of CMs (Rnf20 KO). Wild-type pups injected with a high dose of AAV-TnT-Cre were used as controls (WT). H3K4me3 and H3K36me3 ChIP-seq was performed on heart apices at P7. As expected, H3K4me3 and H3K36me3 consistently decreased in Rnf20 KO compared to WT, consistent with a reduction in H2Bub1 that is thought to promote chromatin decondensation in yeast.
marked promoters and gene bodies, respectively, of actively transcribed genes in all replicates (Fig. 6a, b). H3K4me3 and H3K36me3 signal at each TSS and gene body, respectively, was quantified in WT and Rnf20 KO (Fig. 6c, d upper panels). Genes with reduced H3K4me3 or H3K36me3 signal in Rnf20 KO were strongly enriched for the set of genes downregulated in RNF Cas-KO (Fig. 6c, d lower panels). Among these were the mature sarcomere isoform genes Myh6 and Tnni3, which were downregulated in RNF Cas-KO and had reduced H3K4me3 and H3K36me3 occupancy in Rnf20 KO (Fig. 6e, f). Additional gene sets, such as the maturation genes identified in Fig. 5f–g, were also significantly enriched among genes with reduced H3K4me3 and H3K36me3 occupancy (Suppl. Fig. 8a, b; gene set “Intersect. mature”). The correlation between change in gene expression in RNF Cas-KO and change in H3K4me3 or H3K36me3 occupancy was poor (Suppl. Fig. 8c, d), suggesting that changes in these epigenetic marks are not simply the downstream effects of altered gene expression. Together, these analyses identify RNF20/40 as the first epigenetic regulators of CM maturation.

Discussion

In this study, we developed a resource-efficient, in vivo, forward genetic screen for transcriptional regulators of CM maturation. Our strategy uses a massively parallel approach in which the screening unit is the individual cell, such that thousands of individual genetic mutations can be screened using a small number of animals. This strategy will be broadly applicable for...
cells efficiently transduced by AAV and for phenotypes amenable to high-throughput, single-cell measurement. For example, the forward genetic screen described here could be readily extended to study pathological cardiomyocyte hypertrophy using the same Myh7-YFP reporter, which is reactivated by pathological cardiomyocyte stress. The scope of accessible phenotypes could be dramatically expanded by developing methods to sort individual cells by morphology.48

Among the validated transcriptional and epigenetic regulators of CM maturation uncovered by our screen are Rnf20, Rnf40, Taf2, and Taf3. Notably, these genes are components of protein complexes whose perturbation is linked to congenital heart disease. Rnf20 and Rnf40 encode subunits of a H2B ubiquitin ligase, and CHD patients were enriched for de novo mutations in genes in this H2B monoubiquitination pathway, including RNF20 and UBE2B, an E2 ubiquitin-conjugating enzyme involved in writing H2Bub116,17. Taf2 and Taf3 are TATA-binding protein-associated factors, components of the TFIIID protein complex that positions RNA Polymerase II at gene promoters. Another member of this complex, Taf1, was found to have a significant burden of predicted damaging mutations in CHD compared with controls.19 This overlap between genes that regulate CM maturation and genes that cause CHD likely reflects the important function of these genes in both heart morphogenesis and cardiomyocyte maturation. One implication of this overlap is that patients with CHD caused by mutations in these genes may be at risk for later complications, such as myocardial dysfunction, due to impaired CM maturation.

Rnf20 and Rnf40 interact to form a ubiquitin ligase complex that writes the H2Bub1 epigenetic mark. Depletion of Rnf20/R40 broadly impaired CM maturation by abrogating writing of H2Bub1 and preventing essential transcriptional changes. H2Bub1 has been implicated in regulating cell-state transitions during development and cancer. Consistent with our results, a previous study in murine embryonic fibroblasts found that genes differentially expressed following Rnf40 depletion were those with low or moderate H2Bub144. Strongly activated genes with high H2Bub1 may not be dependent on H2Bub1 because of redundant activating mechanisms, whereas a subset of genes with lower H2Bub1 may be more reliant upon H2Bub1-dependent mechanisms.44 Further experiments are needed to understand what makes the expression of specific genes, such as genes activated during cardiomyocyte maturation and other developmental transitions, dependent upon RNF20/40 and H2Bub1. In addition, nonhistone RNF20/40 substrates have been identified50–53, and it is possible that they also contribute to gene regulation of CM maturation.

**Methods**

**Mice.** All relevant ethical regulations for animal testing and research were followed. All procedures were performed following protocols approved by the Boston Children’s Hospital Institutional Animal Care and Use Committee. Mice were maintained at 65–75°F with 40–60% humidity on a normal 12-hour light, 12-hour dark cycle. Animals were genotyped by PCR and the genotypes were confirmed by DNA sequencing using primers that do not amplify endogenous DNA. Cells were transfected using Lipofectamine 2000 (Invitrogen, 12259-028) according to the manufacturer’s protocol.

**Library design and construction.** A list of 3000+ potential transcription factors and epigenetic regulators was compiled using publicly available data from the RNKEN Transcription Factor Database, and AnimalTFDB. RNAseq data from isolated p6 wild-type CMs was used to prioritize 1894 factors. In addition, 259 genes differentially expressed at p6 within GATA/6 double-KO CMs (adjusted p-value < 0.05), and 291 genes with strong adjacent GATA4 binding at p0 were also included, for a total of 2444 genes targeted. Six gRNAs were selected for each gene using CRISPR RGEN Tools Cas-Database23. To select the best six guides per gene, the following five rules were applied, in order, by sequentially applying each rule if the candidate set contained less than six guides: (1) gRNAs target constitutive exons, CDS range 5–50%, number of mismatches 0,1,2,0,0 (i.e., only one target-exact match, zero off-target sites with one mismatch, and zero off-target sites with two mismatches); and with two mismatches from the H2Bub1 epigenetic mark. Depletion of previous study in murine embryonic H2Bub1 and preventing essential transcriptional changes.

**Injection, sample collection, and flow cytometry.** One-day-old R26Cas9-2A-GFP/+, Myh7-YFP/+ pups were subcutaneously injected with 50 µl of gRNA library virus at a concentration of 2 × 1011 vg (viral genomes) per ml and spiked with a single CASAAV virus targeting both GATA4 and GATA6 at a final concentration of 1 × 108 vg/ml. At 18 days of age, hearts were dissected and retrograde perfused in ice-cold 200-bp amplicon. Eight reactions were pooled, cleaned up via the DNA Clean and Concentrator column and electroporated into Agilent Sure Electroporant columns with a Bio-Rad Gene Pulser Xcell Electroporation System (1700V, 200-ohms resistance, 25 μF capacitance, and 1-mm cuvette gap). In all, 40 ng of purified ligation product in 2 µl volume was electro- transduced into 40 µl cells. About 900 µl of SOC media was added immediately after electroporation, and bacteria were incubated at 37 °C for 1 h. Bacteria were plated on LB agar containing ampicillin and allowed to grow for 18 h. Four electro- porations were cotransfected in each (250 µg DNA/collection). DNA was scraped into SOC media, cultured for an additional 1.5 h, and plasmid DNA harvested (Invitrogen PureLink HiPure Maxiprep, 210017), yielding 240 µg of DNA. The gRNA plasmid library has been deposited at Addgene (Cat# 138015). AAV library was packaged into AAV using PEI transfection of ten 15-cm plates of HEK293T cells (Agilent, 240073). AAV was titrated by qPCR targeting the Tnt promoter, using standard qPCR parameters.57

**NGS library preparation.** Reverse transcription of gRNAs and adapter addition was achieved using a custom protocol with the Clontech SMART-Seq v4 Ultra Low Input RNA sequencing kit (634894). Approximately 30 ng of RNA from YFP+ CMs or 500 ng RNA from input CMs was reverse transcribed as directed in the manufacturer’s protocol, except that SMART-Seq 2DS Primer IIA was replaced with a gRNA scaffold-specific primer (Suppl. Table 3). This reverse-transcription step utilizes a template-switching oligo to add an adapter of known sequence to the variable 5′ end of the gRNA.58 cDNA was then amplified in two sequential rounds of PCR to add NGS sequencing adapters. In the first round of amplification, the full-length forward-read adapter was added to the 5′ end of the gRNA and a half-adapter was added to the 3′ end (Suppl. Table 3). NEB Phusion (M0530L) was used for five cycles of amplification, according to the manufacturer’s protocol. First-round PCR product was purified via Zymo DNA Clean and Concentrator column and eluted in 25 µl of H2O. About 5 µl of the purified product was used as input in a second round of 20 cycles of amplification, which completed the reverse adapter and added a sample-specific Illumina TruSeq multiplexing index. The resulting cDNA was used for library preparation, which included, for a total of 2444 genes targeted of each sample was assessed using the KAPA Library Quantitation Kit (KR0405), allowing samples to be evenly pooled and submitted for single-end 75 nt
sequencing on a NextSeq500. An average sequencing depth of 4.8 million reads per sample was achieved.

**Screen NGS analysis.** gRNA NGS libraries were trimmed to remove adapters and the gRNA scaffold, leaving only the 20-bp variable region. Bowtie2 was used to align trimmed sequence files to mm10. Counts for each gRNA were acquired by quantifying sequencing coverage of genomic regions corresponding to the gRNA library via BedTools Coverage. Differential expression of individual gRNAs in YFP⁺ versus input samples was calculated by using gRNA counts as input for DESeq2. Differential gene representation in YFP⁺ versus input samples was calculated using median-normalized gRNA counts as input for the MAGeCK software package, which consolidates scores for multiple individual gRNAs targeting the same gene into a single gene-level enrichment score. Five samples were removed prior to MAGeCK analysis due to insufficient enrichment of control gRNAs, improper clustering, or poor library coverage (Suppl. Fig. 2).

**In situ T-tubule imaging.** CASAAV virus was injected subcutaneously at P1 into R26Cam-LoxP; Myh7F7;RNF40⁺/⁻ pups. Animals were sacrificed at P28 and hearts were perfused with the T-tubule-binding dye FM 4-64 (Thermo, T3166) at 5 µm at 20 °C, followed by direct imaging on an Olympus FV3000RS confocal microscope. Organization and abundance of transverse and longitudinal T-tubule elements was quantified using AutoTT software.

**Immunostaining and phenotyping.** Freshly dissociated adult CMs were cultured on laminin-coated (2 µg/cm², Life Technologies, P36965) and imaged on an Olympus FV3000RS confocal microscope. Coverslips were mounted on slides using Diamond Antifade mountant (Thermo, P36965) and imaged on an Olympus FV3000RS confocal microscope. For H2Bub1 staining, CMs were incubated in 1% SDS/BSA for 5 min, rinsed with BSA three times, and then blocked and stained as described above. MyHC protein expression was quantified without staining for the fraction of transfected cells and percent YFP⁺ (YFP fluorescent) was manually counted. Nuclearization of each CM was also manually counted, while area, length, and width were measured using ImageJ. T-tubule organization was quantified using AutoTT software. Cell size, dimensions, nucleation, and T-tubule organization were all measured on the same cells. In Fig. 2, to control for variation in dissociation quality, measurements were made on GFP⁺ (transduced) CMs and GFP⁻ CMs (untransduced), and data from transduced CMs were normalized to that of untransduced CMs from the same mouse (Fig. 2). For histology, hearts were collected at P7 and fixed overnight in 4% PFA. Hearts were then transferred to 15% sucrose/BSA for 3 hr, and then to 30% sucrose overnight. Hearts were then cryoprotected in 70% ethylene glycol with 30% glycerol at −80 °C. Hearts were then sectioned on a cryostat at 10 µm, and sections stored at −20 °C. Staining was performed as described above for dissociated CMs, except sections that were incubated in BSA + 0.1% Triton X-100 for 20 min to remove Tissue Freezing Medium before starting the staining protocol.

**Western blotting.** Ventricles were homogenized in 1 ml of RIPA buffer and agitated at 4 °C for 30 min. Lysates were spun at 10,000 g for 10 min at 4 °C, and the supernatant transferred to a new tube. About 40 µg of protein was boiled in SDS loading buffer for 5 min, loaded onto an InvitroGen Bolt 4–12% gradient precast mini gel (NV0412BOXX), and run at 165 volts for 45 min. Protein was transferred to a pre-equilibrated PVDF membrane in transfer buffer (Boston BioProductsorpB-190) via Biorad Trans-Blot SD Semi-Dry Transfer Cell, at 20 volts for 40 min. Blots were cut and blocked in 2% milk/TBST for 1 h at 4 °C. Blots were then incubated overnight at 4 °C with primary antibodies (Suppl. Table 4) for H2Bub1, NPPA, and GAPDH, in block solution. The next day, blots were rinsed in TBST and incubated with HRP-conjugated donkey anti-rabbit secondary antibody (Rockland 611-703-127) diluted 1:20,000 in blocking solution for 1 h at room temperature. Blots were then rinsed in TBST, incubated in Millipore Immobilon Western Chemiluminescent HRP Substrate (WBKLSL5000) for 1 minute, and imaged on a GE Healthcare ImageQuant LAS4000. In the case of RN20/40, the blot was first probed for RN20, then stripped, and reprobed for RN40.

**RNA sequencing.** One day old R26Cam-LoxP; Myh7F7;RNF40⁺/⁻ pups were subcutaneously injected with 50 µl of CASAAV containing the most enriched Rnf20 targeting gRNA and the most enriched Rnf40 targeting gRNA in a single vector (CASAAV-Rad20/40) or a control vector containing Cre but no gRNAs. Vessels were injected at a concentration of 1 x 10¹³ vg/ml, which was sufficient for ~20% CM transduction. At four weeks of age, mice were sacrificed by cervical dislocation and treated with DNase prior to elution in 20 µl of H₂O, as described above. Reverse transcription of 40 ng of RNA for each sample, and library amplification, was performed using the SMARTseq v4 Ultra Low Input RNA-seq kit (Clontech 634989). The standard manufacturer’s protocol was followed, with eight cycles of amplification. To prepare amplified libraries for sequencing, 300 pg was fragmentated and indexed using the Nextera XT DNA Library Preparation Kit (Illumina FC-131-1024) and Index Kit (Illumina FC-131-1001), according to the standard manufacturer’s protocol. Single-end 75-bp sequencing of pooled libraries was performed using the NextSeq500. Following the firstumi sequencing to remove adapter sequences, reads were aligned to the mm10 transcriptome (ftp://ftp.ensembl.org/pub/release-93/fasta/mus_musculus/cdna/) using Kallisto. An average sequencing depth of ~32 M pseudoreads per sample was achieved. Kallisto counts for individual transcripts were consolidated into gene-level counts using TXImport, and analyzed for differential expression between control and RN20+/−9-depleted sample groups using DESeq2. Gene ontology analysis of differentially expressed genes was conducted with Gene Set Enrichment Analysis, with 10,000 permutations and reported gene sets being limited to those containing at least 30 genes (Suppl. Table 2).

To confirm gene expression between wild-type neonatal and mature CMs, we dissociated postnatal P1 or P28 heart ventricles. Neonatal CMs were isolated by Neonatal CM Isolation Kit (Cellutron, Cat. nc-6031) and purified using the Miltenyi Neonatal CM Isolation Kit (Miltenyi Biotec, 130-100-825). P28 CMs were dissociated by collagenase perfusion and purified by differential sedimentation. Total RNA was isolated from purified cells using the RNeasy mini kit (Qiagen), and samples were then fragmented and reverse transcribed into sequencing libraries using Script-seq v2 (Illumina). Single-end 75-bp sequencing was performed on a NextSeq500. We analyzed biological triplicate samples for each developmental stage.

Chip sequencing. Each final Chip-seq sample consisted of 20 apexes from bisected P1 hearts, eight apexes from P7 hearts, or four apexes from P28 hearts. Each sample consisted of male and female tissue in equal proportions. Tissue, either two sets of four P7 apexes or a single set of four P1 apexes, was harvested into ice cold PBS, transferred into 1.2 ml of 1% formaldehyde (Sigma F8775), and homogenized with a T10 Ultra Turrax homogenizer at setting six for 30 s. The homogenate was cross-linked at room temperature on a rotator for 25 min. Formaldehyde was quenched by adding glycine to a final concentration of 300 mM and incubating for 5 min. CMs were pelleted at 3000 g for 3 min at 4 °C. The pellet was resuspended in 1 ml of cold PBS. This PBS wash was repeated a total of three times. After the third wash, the pellet was resuspended in hypotonic buffer (20 mM HEPES pH 7.5, 10 mM KC1, 1 mM EDTA, 0.1 mM activated Na3VO4, 0.5% NP-40, 10% glycerol, 1 mM DTT, and 1X Complete Protease Inhibitor cocktail). Transduced samples were transferred to a glass dounce homogenizer, and dounced with pestle “B” (20x for P1, 100x for P7, or 200x for adult). Samples were transferred into siliconized Eppendorf tubes and incubated on ice for 15 min, then centrifuged at 13,000 g for 2 min at 4 °C. Pellets were stored in liquid nitrogen until multiple samples were ready for sonication.

When sufficient samples were available for Chip, pellets were thawed, resuspended in 1 ml of hypotonic buffer per 20 neonatal apexes, four P7 apexes, or two adult apexes, and incubated on ice for 5 min. Lysate was transferred into glass dounce homogenizer and again dounced with pestle “B” as before. Lysate was centrifuged at 13,000 g for 2 min at 4 °C, and the pellet was resuspended in 20 mM Tris-HCl (pH 8.0) 2 mM EDTA, 150 mM NaCl, and 1% Triton X-100) supplemented with SDS 1%, and 1X protease inhibitor. Samples were then equally split into two 0.65-ml tubes (BrandTech 781310 or Corning 3208) for sonication. Samples were sonicated using a Qsonica 800 R, at 65% amplitude, 10 s on, 30 s off, for 25 min of on-time (100 min in total). After sonication, the bisected samples were re-suspended in 1.5 ml siliconized tubes and centrifuged at 18,500 g for 5 min at 4 °C. The chromatin/supernatant was transferred to a new tube and a 20-µl aliquot was removed as an input sample for each replicate. In all, 120 µl of Protein A Dynabeads (Life Technologies 10020D) were prepared for each Chip sample by washing beads in 1 ml of a solution (3%BSA) at 4 °C for 20 min, washed in CAC buffer, and treated with 1% Protease inhibitor. After the third wash, beads were resuspended in 1 ml of BSA solution, 250 µl of which was set aside to preclar the chromatin. In all, 5 µl of H2Bub1 antibody (Cell Signaling 5546), 3 µl of H3K4me3 antibody (Abcam 448), 1 µl of a specified anti-actin antibody, or 0.5 µl of a specified anti-H2Bub antibody was added to the remaining 750 µl of Dynabeads, along with 250 µl of fresh BSA solution, and incubated overnight at 4 °C. Sheared chromatin samples were...
precleared by incubating with the 250-µl aliquot of Dynabeads resuspended in 2 ml of Dilution Buffer (20 mM Tris pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, and 1:50 Roche eComplete protease inhibitor) for 1 h at 4 °C. Next, the antibody–Dynabead conjugate was rinsed three times in BSA solution, resuspended in 150 µl of Dilution Buffer, and added to the precleared chromatin. After 36 h of incubation at 4 °C, beads were collected on a magnet and washed three times with 1 ml of LiCl wash buffer (1% w/v Na-Deoxycholate, 500 mM LiCl, 1% NP40 Substitute, and 100 mM LiCl, pH 7.5) and twice with TE buffer (10 mM Tris-–HCl, 1 mM EDTA). After removing TE buffer, beads were suspended in 75 µl of SDS Elution Buffer (1% w/v SDS, 0.1 M NaHCO3). Samples were then incubated at 37 °C for 15 min with 1000-RPM mixing. This step was repeated once and the eluates were combined. To reverse crosslinking, input and immunoprecipitated samples were diluted to 200 µl with SDS Elution Buffer, NaCl was added to a final concentration of 200 mM, and Proteinase-K to a final concentration of 0.2 mg/ml. Samples were then incubated at 65 °C overnight. The next day, RNase-A was added to a final concentration of 1 mg/ml and incubated at 37 °C for 15 min. DNA was then purified using a Zymo DNA Clean and up purification. DNA was used as input. In all, 75-nt single-end sequencing was performed on a NextSeq500, or 150-bp paired-end on an Illumina HiSeq. In cases where paired-end sequencing was performed, the reverse reads were discarded, and forward reads clipped up to 75 bp, so that all sequencing could subsequently be processed in a uniform manner.

ChiP and input reads were aligned to mm10 using Bowtie26, and reads for each gene were extracted using Bedtools66. Profile plots were created using DeepTools67. For H2Bub1, library-size normalization was performed with Deseq2. Reads per gene were then normalized to gene length, from NCBI RefSeq first transcription start site to the last transcription stop site, to give a H2Bub1 binding score (Suppl. Data 6). Input values were subtracted from ChiP values for the final ChiP signal. Genes with ChiP signal were considered to be occupied by H2Bub1. For H3K4me3 libraries were RPMK-normalized and reads at the TSS +/- 1 kb were scored using Bedtools (max signal per interval). Replicates were averaged and input signal was subtracted from ChiP signal. Only the highest-scoring TSS was used for each gene. To exclude genes with a low signal, a change in H3K4me3 was analyzed for genes with TSS scoring greater than 500 RPMK (~7000 genes). For H3K36me3 each gene. To exclude genes with a low signal, a change in H3K36me3 was analyzed for the top quintile (~7000 genes).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The data in this publication have been deposited in the GEO database under accession code GSE139975. Both raw and processed data are available. The raw data are also available on the Cardiovascular Development Consortium Server at https://bib2.bch.tuhh. edu/gencode/, experiment number 483R (login as guest). The source data underlying bar, violin, and scatter plots in all figures and supplemental figures are provided in the Supplementary Information/Source Data file. The Riken Transcription Factor Database (TFDB) can be accessed at http://genome.gsc.riken.jp/TFDB/hidoc/.. The Animal Transcription Factor Database (AnimalTFDB) can be accessed at http://bioinfo. life.bust.edu/AnimalTFDB/. Single cell cardiac RNA-seq data were mined from Panglao database (PanglaoDB), accession code SRA653146 (https://panglaoadb.org/ view_data.php?src=SRAS653146&ses=SR582874273). Source data are provided with this paper.

Received: 9 January 2020; Accepted: 2 July 2021; Published online: 21 July 2021

**References**

1. Galdos, F. X. et al. Cardiac regeneration: lessons from development. *Circ. Res.* **120**, 941–959 (2017).
2. Guo, Y. & Pu, W. T. Cardiomyocyte maturation: new phase in development. *Circ. Res.* **126**, 1086–1106 (2020).
3. Scuderi, G. J. & Butcher, J. Naturally engineered maturation of cardiomyocytes. *Front Cell Dev. Biol.* **5**, 50 (2017).
4. Zhang, D. & Pu, W. T. Exercising engineered heart muscle to maturity. *Nat. Rev. Cardiol.* **15**, 383–388 (2018).
5. Guo, Y. et al. Analysis of Cardiac Myocyte Maturation Using CASAAV, a Platform for Rapid Dissection of Cardiac Myocyte Gene Function In Vivo. *Circ. Res.* **120**, 1874–1888 (2017).
6. Guo, Y. & Pu, W. T. Genetic Mosaics for Greater Precision in Cardiovascular Research. *Circ. Res.* **123**, 27–29 (2018).
7. Guo, Y. et al. H2B monoubiquitination regulates meiotic recombination by promoting chromatin relaxation. *Nucleic Acids Res.* **44**, 9681–9697 (2016).
8. Bedada, F. B. et al. Acquisition of a quantitative, stoichiometrically conserved ratiometric marker of maturation status in stem cell-derived cardiac myocytes. *Stem Cell Rep.* **3**, 594–605 (2014).
9. Huang, E. et al. Genomic-wide enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Nat. Protoc.* **12**, 1054–1065 (2017).
Acknowledgements

The authors thank the Dana-Farber Flow Cytometry Core for assistance with cell sorting and Ewa Prasad for guidance related to gRNA reverse transcription. N.J.V was supported by NIH (T32HL007572, F32HL1343501, K99HL143194) and the Boston Children’s Hospital Kaplan Fellowship. W.W.P was supported by NIH (2UM1 HL098166, R01 HL146634) and the American Heart Association (17TRG33410894). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. Portions of this research were conducted on the O2 High Performance Compute Cluster, supported by the Research Computing Group, at Harvard Medical School. See http://rc.hms.harvard.edu for more information.

Author contributions
N.J.V and W.W.P conceived and designed the study. N.J.V executed most experiments. J.Y.L., W.G., C.E.R., Y.Z. and J.S.K. generated plasmids, viruses, and other necessary reagents, and assisted with processing tissues and cells. Y.G assisted with experimental design. P.Z. conducted wild-type C.M.-specific P0 and 4-wk RNA sequencing. Q.M. conducted echocardiography. N.J.V, I.S., S.S., G.C.Y and W.W.P. analyzed the data. N.J.V and W.W.P. wrote the paper.

Competing interests
The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-24743-z.

Correspondence and requests for materials should be addressed to W.T.P.

Peer review information Nature Communications thanks Gianluigi Condorelli and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information are available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.