Nascent transcript O-MAP reveals the molecular architecture of a single-locus subnuclear compartment built by RBM20 and the *TTN* RNA

Evan E. Kania[1,2,3,8], Aidan Fenix[2,4], Daphnée M. Marciniak[1,8], Qiaoyi Lin[1,2,3,6], Sara Bianchi[5], Borislav Hristov[6,8], Shuai Li[1,2,3], Conor K. Camplisson[6], Rose Fields[6], Brian J. Beliveau[2,3,6], Devin K. Schweppe[6], William S. Noble[6,7], Shao-En Ong[1], Alessandro Bertero[5,‡], Charles E. Murry[2,4,8‡], David M. Shechner[1,2,3,‡]

[1]Department of Pharmacology, University of Washington, Seattle, WA, USA

[2]Institute of Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, USA

[3]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA

[4]Departments of Laboratory Medicine and Pathology, Bioengineering, and Medicine/Cardiology, University of Washington, Seattle, WA, USA

[5]Molecular Biotechnology Center “Guido Tarone", Department of Molecular Biotechnology and Health Sciences, University of Turin, Torino, Italy

[6]Department of Genome Sciences, University of Washington, Seattle, WA, USA

[7]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, USA

[8]Present affiliations: Shape Therapeutics, Seattle, WA (E.E.K); Department of Biology, Massachusetts Institute of Technology (D.M. M.); Department of Computer Science and Software Engineering, California Polytechnic State University at San Luis Obispo (B.H); Department of Stem Cell and Regenerative Medicine, and the Eli and Edyth Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California (C.E.M.)

[‡]Corresponding authors: alessandro.bertero@unito.it charles.murry@med.usc.edu; shechner@uw.edu

ABSTRACT

Eukaryotic nuclei adopt a highly compartmentalized architecture that influences nearly all genomic processes. Understanding how this architecture impacts gene expression has been hindered by a lack of tools for elucidating the molecular interactions at individual genomic loci. Here, we adapt oligonucleotide-mediated proximity-interactome mapping (O-MAP) to biochemically characterize discrete, micron-scale nuclear neighborhoods. By targeting O-MAP to introns within the *TTN* pre-mRNA, we systematically map the chromatin loci, RNAs, and proteins within a muscle-specific RNA factory organized around the *TTN* locus. This reveals an unanticipated compartmental architecture that organizes *cis*- and *trans*-interacting chromosomal domains, including a hub of transcriptionally silenced chromatin. The factory also recruits dozens of unique RNA-binding and chromatin-scaffolding factors, including QKI and SAFB, along with their target transcripts. Loss of the cardiac-specific splicing factor RBM20—a master regulator of *TTN* splicing that is mutated in dilated cardiomyopathy—remodels nearly every facet of this architecture. This establishes O-MAP as a pioneering method for probing single-locus, microcompartment-level interactions that are opaque to conventional tools. Our findings suggest new mechanisms by which coding genes can "moonlight" in nuclear-architectural roles.

MAIN

In eukaryotes, the tasks of reading, modulating, copying, and repairing genomic information must be executed within the dense confines of the nucleus, where meters of chromatin are compacted into a space that is only microns wide. To achieve this, cells partition their nuclei into a highly compartmentalized architecture that enables spatial control over these core genomic functions[1-3]. For example, processes like DNA repair, ribosome biogenesis, ribonucleoprotein (RNP) assembly, RNA polymerase II (Pol II) transcription, and pre-mRNA splicing each appear to be coordinated in dedicated, micrometer-scale compartments that concentrate regulatory proteins and RNAs[4-7]. These compartments also spatially cluster co-regulated loci, including both long-range *cis*-contacts along the same chromosome, and *trans*-contacts between chromosomes[8-10]. Notably, a single nucleus may contain dozens to thousands of a given compartment type, each of which has a characteristic, core molecular architecture[10-12]. It is presumed that individual compartments can modify or functionalize this core architecture to support the unique needs of the loci they regulate. However, exploring this hypothesis has proven challenging without robust technologies for biochemically characterizing the distinct microenvironments surrounding individual target loci. As a result, the functional diversity of individual nuclear compartments, and the mechanisms by which this diversity is programmed, remain largely unexplored.

We recently reported the discovery of one such specialized compartment—an "RNA factory" that plays a key regulatory role during cardiomyogenesis[13]. In cardiomyocytes (CMs), this factory assembles around nascent pre-mRNAs encoding the giant protein Titin (*TTN*), and nucleate prominent foci of the muscle-specific splicing factor RBM20, a key regulator of *TTN* mRNA biogenesis[14, 15]. Genomic loci encoding other RBM20 targets (located on several chromosomes) also visit these RBM20 foci, which appears to enhance their alternate splicing. These observations suggest that *TTN* RNA factories might serve as "crucibles" that coordinate the biogenesis of cardiomyogenic genes[13]. Previous studies indicated that RBM20 foci are a class of nuclear speckles[16]—RNA-rich nuclear bodies that coordinate pre-mRNA splicing[5, 12]. But, while cardiomyocytes contain dozens of nuclear speckles, only those located near the *TTN* locus compartmentalize RBM20 and its target genes, suggesting their functional specialization[13]. A similar model of RNA-mediated compartmentalization is thought to underlie the assembly of many other prominent subnuclear structures—including nucleoli, Barr bodies, paraspeckles, and nuclear stress bodies (nSBs)[17-19]. Yet, while these compartments are all assembled on long

noncoding RNAs, this kind of architectural function has rarely been ascribed to a pre-mRNA[20, 21]. Hence, the *TTN* RNA factory may represent an under-explored mechanism of nuclear compartmentalization in which coding loci "moonlight" as sources of chromatin-regulatory RNAs.

Proper assembly of the *TTN* RNA factory may also be critical for human health. Mutations in RBM20, including those that sequester it outside of the nucleus[22], induce a malignant form of dilated cardiomyopathy (DCM) characterized by conduction system disorders, ventricular arrhythmia, and rapid progression to heart failure[23, 24]. Current therapies can only manage these symptoms; the sole effective treatment is heart transplantation[23, 24]. Much of the cardiac dysfunction observed in RBM20–DCM can be ascribed to defects in *TTN* splicing, which induces an elastic, fetal-like protein isoform that severely weakens left ventricular function[14]. However, other facets of disease etiology, including calcium current aberrations, cannot be directly ascribed to the missplicing of RBM20 target transcripts[25]. We thus hypothesized that nuclear depletion of RBM20 may also disrupt the architecture of the *TTN* RNA factory, with potentially deleterious effects on chromatin organization and gene expression.

To explore this hypothesis, we sought to develop a new experimental strategy that would enable targeted biochemical characterization of discrete subnuclear compartments like the *TTN* RNA factory—an approach for which few robust tools are currently available. To date, most efforts at characterizing nuclear architecture have focused on the looping and folding of chromatin at the nanometer-scale[26, 27]. Although powerful, these methods have difficulty probing the higher-order, micron-scale compartmental interactions that are pervasive in nuclear organization[28], including *trans*-chromosomal contacts[29]. Newer genomics approaches that explicitly probe such higher-order structures (*e.g.*, GAM[30]; SPRITE[28]) are technically demanding and have lower dynamic range[31]. Moreover, because these methods probe all nuclear compartments *en masse,* they might lack the sampling depth to discover unique interactions within any individual compartment. In principle, such compartment-specific interactions might be probed using pulldown-based biochemical strategies—for example, immunoprecipitation of RBM20, or antisense oligonucleotide-based capture of the *TTN* pre-mRNA or genomic locus[32-34]. However, these methods would be challenging to apply to compartments like the *TTN* RNA factory, which are lowly abundant and presumably too fragile to survive biochemical purification.

We reasoned that we could overcome this barrier using *in situ* proximity biotinylation. This approach uses promiscuous biotinylating enzymes (*e.g.* Horseradish Peroxidase (HRP); APEX), to affinity-tag biomolecules

within a target compartment, *in situ,* enabling unbiased discovery of compartment-level interactions without complex fractionation or pulldown[35]. Although proximity-labeling has been previously adapted to characterize nuclear structure—for example, by using CRISPR/Cas9 to localize APEX to target genomic loci[36, 37], or using antibody-HRP conjugates to label all compartments of a given type[38]—these established approaches lack the targeting specificity needed to probe single-locus compartments with high accuracy.

To overcome this challenge, we here adapt **O**ligonucleotide-mediated proximity-interactome **MAP**ping (**O–MAP**[39])—a novel RNA-targeted proximity-biotinylation tool—to achieve first-of-its-kind biochemical characterization of discrete subnuclear compartments, with unrivaled spatial precision. By targeting O-MAP to introns on the *TTN* pre-mRNA, we elucidate the chromatin loci, transcripts, and proteins that co-compartmentalize within the *TTN* RNA factory, and map how this molecular architecture is disrupted upon loss of RBM20. Our novel approach reveals unanticipated new facets of factory architecture that had eluded prior detection by conventional tools. This includes a "hub" of *trans*-chromosomal interactions built from transcriptionally silenced chromatin, which recruits hundreds of RNAs post-transcriptionally. Loss of RBM20 appears to remodel every facet of this architecture, disrupting factory chromatin interactions and rewiring the networks of co-compartmentalized proteins and RNAs, potentially dysregulating RNA biogenesis. To our knowledge, this work represents the first biochemical characterization ever attained for a single-locus subnuclear compartment, and demonstrates O-MAP's ability to probe facets of nuclear architecture that are opaque to conventional tools. This work furthermore reveals new molecular players that enable *TTN* to coordinate gene expression during cardiac development, with potential implications for RBM20–DCM. We anticipate that these mechanisms may be broadly utilized by nascent transcripts to sculpt nuclear structure in other developmental and disease contexts.

RESULTS

Nascent transcript O-MAP enables precise in situ *proximity-biotinylation of the TTN RNA Factory*

O-MAP uses programmable DNA oligonucleotide probes (similar to those used in RNA-FISH)[40, 41] to recruit proximity-labeling enzymes to an RNA of interest within its native context[39] (**Fig. 1a**). In O-MAP, samples are chemically fixed and oligo probes are hybridized to the target RNA *in situ*. These probes are appended with a universal "landing pad" sequence that, in a subsequent hybridization step, recruits a secondary oligo that is chemically conjugated to Horseradish Peroxidase (HRP). Upon addition of biotin-tyramide and hydrogen

Kania *et. al.* **Figure 1.**

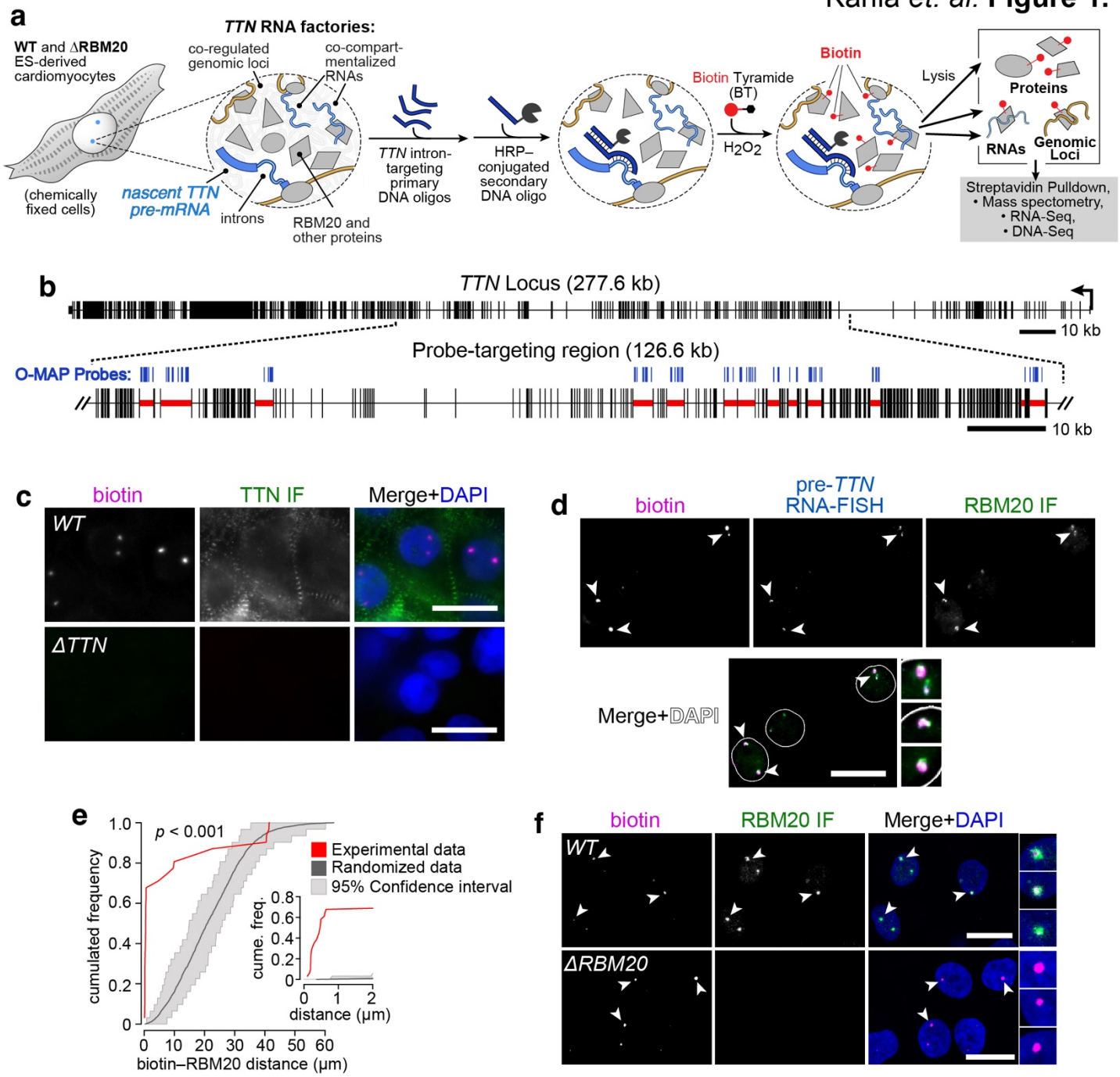

Figure 1. Microenvironment-mapping of the *TTN* RNA factory by nascent transcript O-MAP. a, Approach. WT and ΔRBM20 hESC-derived cardiomyocytes (CMs)[13] were chemically fixed, and pools of antisense DNA probes were hybridized to the introns on the *TTN* pre-mRNA. These probes recruit a common HRP-conjugated oligo that catalyzes *in situ* proximity-biotinylation. Biotinylated material was enriched by streptavidin pulldown, and interacting factors (proteins, RNAs, genomic loci) were isolated for downstream analyses[39]. **b,** O-MAP probes (*blue*) targeted *TTN* introns previously identified as RBM20 targets (*red highlight*)[16]. **c–f,** O-MAP enables highly precise *in situ* biotinylation in *TTN* RNA factories. **c,** In WT CMs (*top*) O–MAP induced two prominent biotinylation puncta per diploid nucleus; this signal is ablated in cells lacking the *TTN* promoter (ΔTTN, *bottom*)[13]. IF: immunofluorescence. **d,** Combined O-MAP, RNA-FISH, and RBM20 immunofluorescence demonstrates that O-MAP specifically labels *TTN* RNA factories. Arrowheads denote colocalization, highlighted in zoomed images (*bottom right*). **e**, Object-based quantification of the pre-*TTN* O-MAP—RBM20 colocalization distances, as observed in **d**. The distribution of the distances from experimental images (*red*) falls outside the confidence interval of the distribution obtained for 100 images in which object locations are randomly shuffled (*gray*). Inset highlights data below 2 μm. **f,** Loss of the RBM20 protein does not diminish the intensity or localization of pre-*TTN* O-MAP. All scale bars: 20 μm.

peroxide, HRP generates biotinyl radicals that diffuse outward (~50–75 nm)[42] and covalently tag nearby molecules, enabling their capture and analysis.

We sought to apply O-MAP to the *TTN* RNA factory by targeting the RNA "scaffold" around which this factory nucleates: unspliced *TTN* pre-mRNA[13]. To pinpoint these unspliced nuclear transcripts (and avoid mature cytoplasmic mRNA), we designed O-MAP probes against intronic sequences within the *TTN* gene, focusing on introns known to be bound and regulated by RBM20[16] (**Fig. 1b**; **Supplementary Table 1**). We applied these probes in an established cell culture model that generates highly pure (>90% cardiac troponin T-positive) cardiomyocytes (CMs) by *in vitro* differentiation from RUES2 human embryonic stem cells (hESCs)[13]. Parallel experiments in an isogenic, RBM20 knockout line[13] (ΔRBM20 CMs) enabled us to also map disruptions in pre-*TTN* factory architecture upon RBM20 depletion. Although such homozygous RBM20 deletions are not observed in human DCM patients[43], this line provides a biochemically clean model to study the effect of RBM20 nuclear depletion (which is observed with most disease-associated variants), while avoiding potentially confounding effects from mutant RBM20 aggregation[22].

We began by validating the targeting precision of our O-MAP probe set. In wild-type (WT) CMs, pre-*TTN* O-MAP induced two prominent (~1 μm diameter) biotinylation puncta within each diploid CM nucleus—corresponding to the two alleles of *TTN*—with undetectable background signal elsewhere (**Fig. 1c,** *top*). O-MAP puncta were not observed in CMs in which both *TTN* alleles had been CRISPR-edited to remove the promoter of full-length Titin[13], demonstrating the high specificity of our approach (**Fig. 1c**, *bottom*). In WT cells, biotinylation puncta exclusively overlapped with nascent *TTN* transcripts, as gauged by intron-targeting RNA-FISH[39], indicating that O-MAP had targeted pre-*TTN* with high precision and accuracy (**Fig. 1d**). These puncta also colocalized with foci of the RBM20 protein, as gauged by immunofluorescence, with an average O-MAP–RBM20 displacement near the resolution limit of conventional microscopy (overlap p-value < 0.01; n=100; **Fig. 1d, e**). Importantly, in ΔRBM20 CMs, the number, size and intensity of O-MAP biotinylation puncta were nearly indistinguishable from those of WT cells (**Fig. 1f**, *bottom*). This suggests that any differences observed between WT and ΔRBM20 cells are likely not artifacts of differential O-MAP labeling of the *TTN* microenvironment, but rather reflect *bona fide* remodeling of this microenvironment upon RBM20 loss.

Collectively, these data demonstrate intron-targeted O-MAP's ability to biotinylate single-locus subnuclear compartments at unprecedently high targeting precision. With these validated reagents in hand, we sought to elucidate the *TTN* RNA factory's chromatin architecture, transcriptome, and proteome.

The TTN RNA factory forms of hub of* cis *and* trans*-chromosomal interactions

The *TTN* RNA factory is thought to organize a network of higher-order chromatin interactions that recruits and compartmentalizes loci from multiple chromosomes, potentially facilitating their co-regulation[13]. We sought to elucidate this network using O-MAP-ChIP, which captures chromatin loci near a target RNA (here, the unspliced *TTN* pre-mRNA) by pulling down the biotinylated proteins to which these loci are bound[39] (**Fig. 2a**). In WT CMs, O-MAP-ChIP revealed the striking enrichment of megabase-scale domains spanning the entirety of chromosome 2 (on which the *TTN* gene is itself located), which we term *Cis*-Interacting Domains (CIDs; **Fig. 2b**, *top*). Notably, CID periodicity approximates the "A/B compartmentalization"[2] of chromosome 2—previously mapped in this cell line by DNase Hi-C[13]—with CIDs corresponding to the transcriptionally active "A compartment" (average Pearson's correlations: O-MAP-vs-Hi-C r = 0.41; Mock-vs-Hi-C r = –0.24, **Fig. 2b**, *bottom*, and **Supplementary Figure 1**). This suggests that, at the chromatin level, nascent *TTN* transcripts predominantly engage with other active genes on chromosome 2, consistent with the prevailing "chromosome territory" model of nuclear organization[44]. However, we also observed the marked enrichment of nineteen megabase-scale (average ~5.15 Mb) domains located on nine other chromosomes, which we term *trans*-Interacting Domains (TIDs; **Figs. 2c–d** and **Supplementary Table 2**). *Trans*-contacts of this sort are challenging to capture by conventional proximity-ligation-based tools (*e.g.*, Hi-C)[2, 29], which highlights O-MAP's unique ability to probe compartment-level chromatin architecture.

Remarkably, in the absence of RBM20, the *TTN* pre-mRNA loses all forms of higher-order chromatin associations. In ΔRBM20 CM's, pre-*TTN* O-MAP-ChIP revealed a complete absence of CIDs and TIDs, though a single notable peak centered on the *TTN* locus confirmed that its transcripts were still affiliated with the site of their transcription (**Fig. 2e**). This indicates that, while transcription of the *TTN* locus is required for factory assembly[13], it alone is insufficient to drive this assembly.

We next sought to validate these novel O-MAP TIDs using multicolor DNA SABER-FISH[45] to gauge *TTN*–CID and *TTN*–TID distances (one CID and eight TIDs probed; **Fig. 2f–i**, **Supplementary Fig. 2** and **Supplementary Table 3**). In WT cells, nearly all tested loci (one CID and seven TIDs) appeared to co-compartmentalize with *TTN*, with median locus–*TTN* distances of 0.65 μm (CID) and 1.2–1.8 μm (TIDs), comparable in scale to nuclear speckles and nuclear stress bodies[7] (**Fig. 2f–i**, *right,* and **Supplementary Fig. 2**). As predicted, these higher-order interactions were ablated in ΔRBM20 cells, with median locus–*TTN*

Kania *et. al.* **Figure 2.**

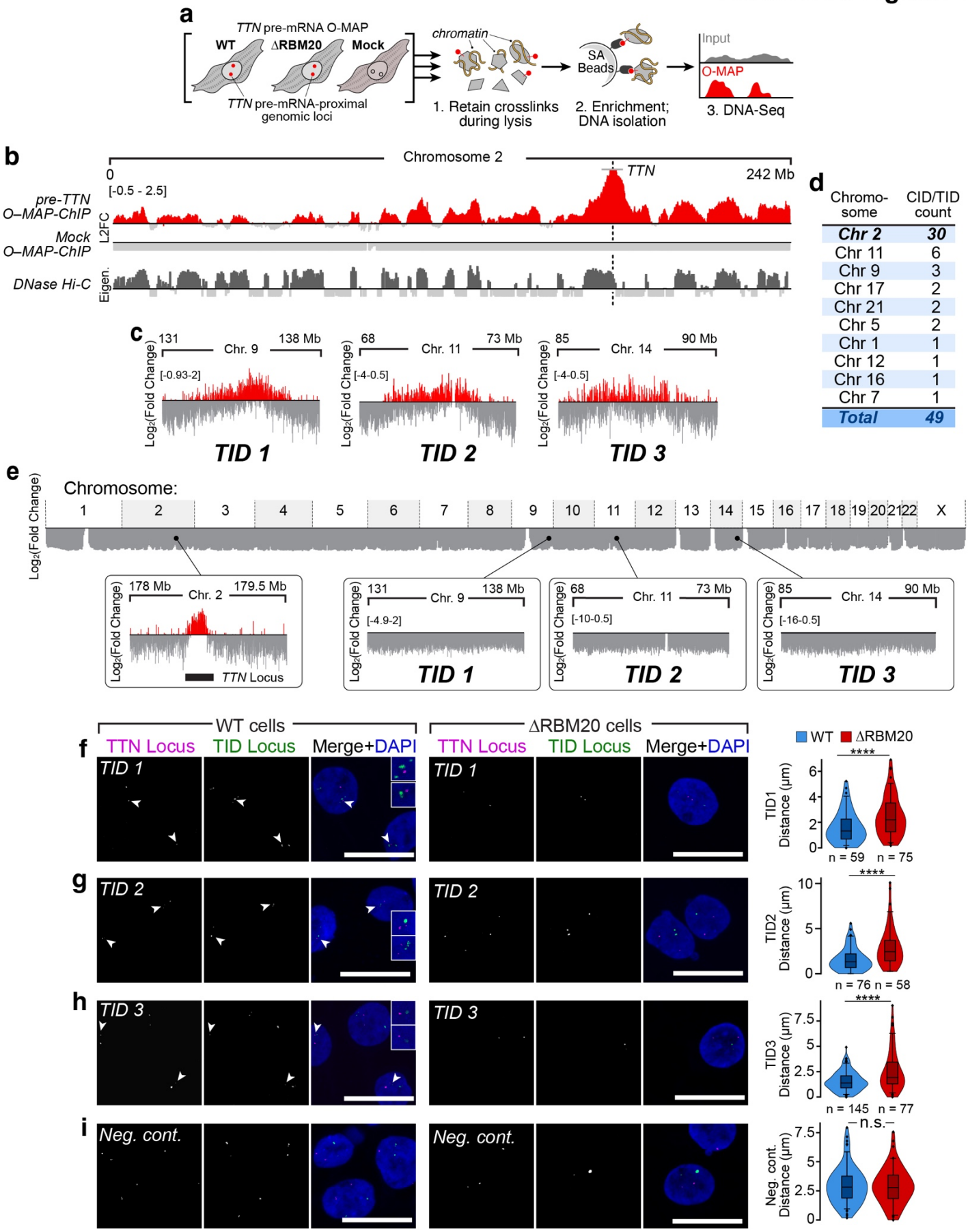

Figure 2. Higher-order chromatin interactions surrounding the *TTN* RNA factory. a, Approach. pre-*TTN* O-MAP-ChIP[39] was applied to WT and ΔRBM20 CMs, and in negative control ("mock") samples omitting primary probes. **b,** In WT cells, pre-*TTN* O-MAP ChIP revealed Mb-scale *cis* interacting domains (CIDs) along chromosome 2 (*top*), paralleling that chromosome's A/B compartmentalization[13] (*bottom*). O-MAP data are L2FC: $\log_2$(fold change, enriched/input). A/B compartmental data are the primary eigenvector derived from the DNase Hi-C correlation matrix[13] **c,** Examples of novel *trans*-interacting domains (TIDs). **d,** Table summarizing all observed pre-*TTN*-proximal chromatin domains, including CIDs (chromosome 2) and TIDs (nine other chromosomes). **e,** In ΔRBM20 cells, nearly all pre-*TTN*-chromatin interactions are ablated. *Top:* Global view over all chromosomes. *Bottom left:* A single statistically significant peak (FDR 0.05, EDD peak score > 40), was observed with the *TTN* locus itself (*black bar*). *Bottom right:* all other CID- and TID- interactions are lost. Shown are the same genomic windows as (**c**). All O-MAP data are L2FC (enriched/input); note differences in y-axis scaling. **f–i,** *Left and center:* two-color DNA SABER-FISH[45] targeting the *TTN* (*magenta*) and putative TID loci (*green*). Arrows denote co-compartmentalization, highlighted in zoomed images. All scale bars: 20 µm. *Right:* quantification of *TTN*-TID FISH distances; the number of cells in each condition are indicated below. Note that TIDs are significantly closer to *TTN* in WT CMs. Internal box and whisker plots indicate median, 25th, and 75th percentile, and the 5th-95th percentile range. p-values by Kruskal-Wallis test followed by Dunn's multiple comparisons vs WT. n.s. > 0.05, **** < 0.001. Negative control (n.c.) is a *trans* locus of chr. 4. See also, (**Supplementary Fig. 2**).

distances of 1.3 µm (CID) and 1.9–2.7 µm (TIDs)—roughly equivalent to that of a negative control locus on chromosome 4 (**Fig. 2i**). We also observed a similar increase in displacement between *TTN* and a control locus on chromosome 2 (**Supplementary Fig. 2**), suggesting that loss of RBM20 impairs *TTN's* ability to co-compartmentalize even with loci on its own chromosome (**Fig. 2b, e**).

TTN Trans-Interacting Domains are transcriptionally repressed in an RBM20-dependent manner

Next, we sought to identify epigenomic signatures that might control chromatin localization to the *TTN* RNA factory. ChromHMM analysis[46] using established epigenome data for RUES2 CMs[47, 48] revealed that, in aggregate, *TTN*-interacting loci were markedly enriched in genic enhancer domains (EnhG1 and EnhG2; **Fig. 3a,** *top*). However, most of this signal stems from CIDs, consistent with their observed correlation to the activating "A compartment[49]" (**Figs. 2b,** and **3a,** *middle*). To our surprise, TID loci were predominantly enriched in repressive signatures, including strong and weak polycomb sites (ReprPC; ReprPCWk) and bivalent transcription start sites and enhancers (TssBiv; EnhBiv; **Fig. 3a,** *bottom*). We thus hypothesized that, contrary to our initial model that *TTN* factories are exclusively crucibles of cardiomyogenic gene activation[13], they may also coordinate a hub of transcriptional silencing. To investigate this, we examined the expression of TID-localized genes (277 total), using input RNA used in O-MAP-seq analyses, generated from WT and ΔRBM20 CMs (*see below)*. Consistent with our new hypothesis, in WT CMs, TID-enclosed genes are lowly expressed as a class (median TPM = 3.355), and a substantial proportion are altogether transcriptionally silent (52 genes; 19% **Fig. 3b**, *left*). This transcriptional repression is partially RBM20 dependent. In ΔRBM20 CMs, where TID architecture appears to be lost (**Fig. 2e–h; Supplementary Fig. 2**), the expression of TID-enclosed genes increases to an average TPM of 3.763, with concomitant activation of 40 TID genes that are silent in WT cells (**Fig. 3b,** *right*). These data suggest that the *TTN* RNA factory contributes to maintaining some TID loci in a transcriptionally repressive environment.

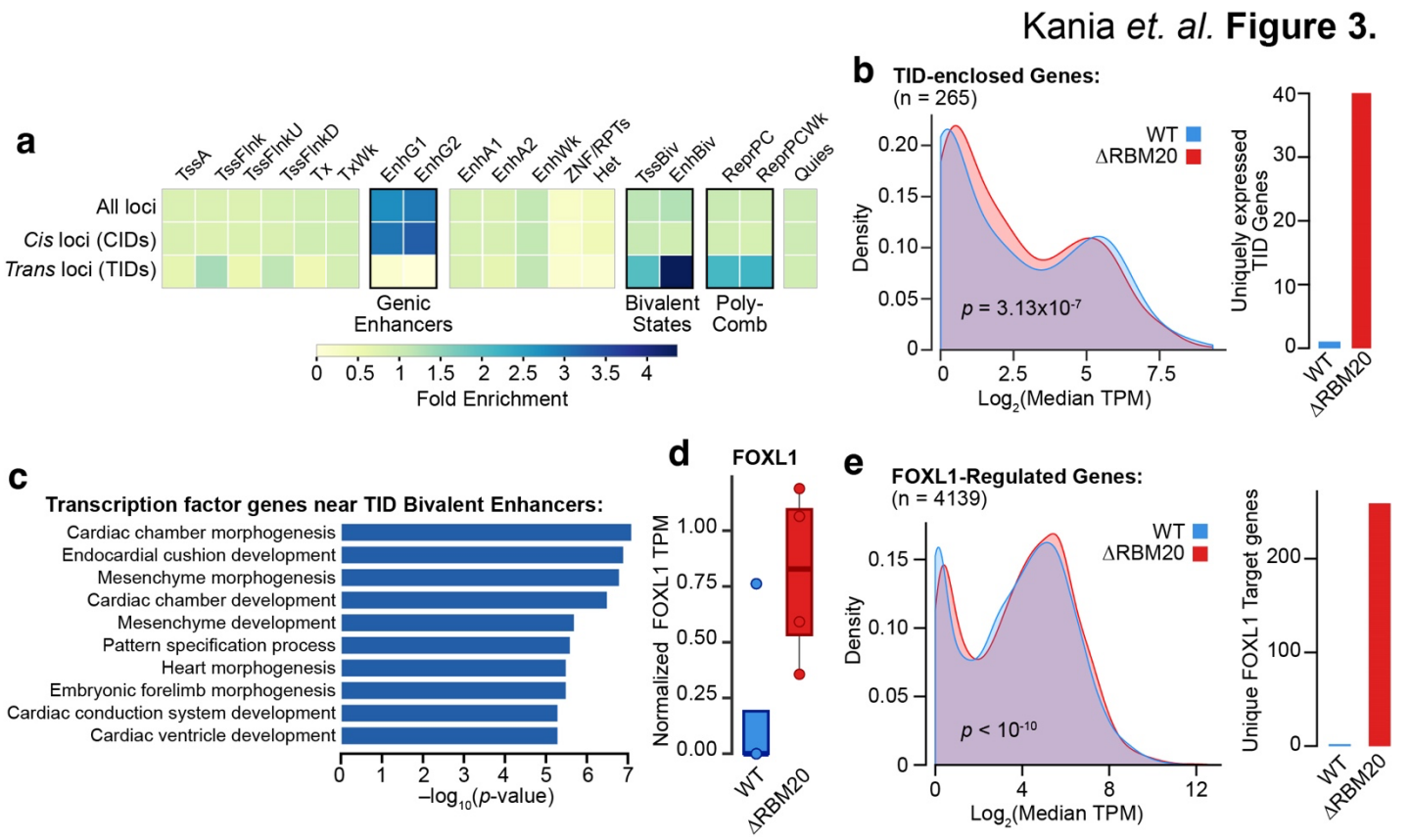

Figure 3. The *TTN* RNA factory includes an RBM20-dependent hub of bivalent chromatin. a, ChromHMM analysis revealed differential chromatin signatures in pre-*TTN*-associated chromatin domains. Note enrichment of genic enhancers in *cis*-loci (CIDs), and of Bivalent- and Polycomb-repressed domains in *trans*-loci (TIDs). **b**, Loss of RBM20 induces transcriptional activation of TID-localized genes. *Left:* histogram of TID-gene expression in WT and ΔRBM20 CMs (*p*-value: Kolmogorov-Smirnov test). *Right:* activation of silenced genes in ΔRBM20 cells. **c,** Gene Ontology (GO) analysis for transcription factor genes near TID-localized bivalent enhancers. The top ten enriched terms are shown. **d–e,** FOXL1 and its target genes are activated in ΔRBM20 cells. **d,** FOXL1 is activated upon loss of RBM20, as measured by sequencing O-MAP-Seq input samples (n = 4 biological replicates)**. e,** *Left:* density plot of FOXL1 target gene expression in WT and ΔRBM20 CMs (*p*-value: 1 Kolmogorov-Smirnov test); *Right:* activation of silenced FOXL1 targets in ΔRBM20 cells.

We were particularly intrigued by the prominent enrichment of bivalent domains—which harbor both activating (H3K4me3) and repressive (H3K27me3) histone modifications—in TID loci (**Supplementary Fig. 4**). Bivalent enhancers are frequently associated with transcription factor (TF) genes that control stem cell differentiation and cell-fate determination events, and are thought to hold these genes in a silent, "poised" state[50]. To explore this, we performed GO analysis on the 27 TF genes harbored in pre-*TTN* TIDs[51]. This analysis revealed a marked enrichment for terms involving cell differentiation and cardiomyogenesis (q-value < 0.05; **Fig. 3c**) and included processes like cardiac conduction system development and ventricle development, which are frequently dysregulated in RBM20-DCM[43, 52] (**Supplementary Fig. 5a**). We thus hypothesized that loss of RBM20 (and the concomitant ablation of TID architecture) may aberrantly reactivate certain TID-resident TF genes along with their downstream targets. Indeed, in ΔRBM20 cells we observed significant upregulation of several key TFs, with members of the Forkhead class (*e.g., FOXC2; FOXL1*) increasing between three and 60-

fold (**Fig. 3d** and **Supplementary Fig. 5b**). These genes also were upregulated in engineered CM lines bearing disease causal RBM20 mutations[22], in a dose-dependent manner (**Supplementary Fig. 5c**). To examine the downstream effects of this spurious TF activation, we inspected the expression of known Forkhead TF target genes[53] in WT and ΔRBM20 cells. In WT cells, targets of *FOXL1* exhibited a bimodal distribution, with prominent populations of lowly-expressed and silent genes. Loss of RBM20 induced a modest yet significant upregulation of these loci, including induction of more than 259 *FOXL1* targets (for 4139 total *FOXL1* targets; median TPM = 15.01 in WT; 17.25 in ΔRBM20; **Fig. 3e**). This suggests that the disruption of TID architecture upon RBM20 loss leads to abnormal expression of key developmental transcription factors, resulting in broader remodeling of the transcriptome.

The TTN RNA factory is a node of post-transcriptional RNA compartmentalization

Many classes of nuclear bodies (*e.g.,* Cajal bodies; paraspeckles; nuclear stress bodies) function in part by compartmentalizing and co-regulating specific populations of target transcripts[7]. Inspired by this, and pursuing our initial hypothesis that the *TTN* RNA factory serves as a hub of RBM20-splicing, we sought to characterize the factory's transcriptome and how this transcriptome is remodeled upon RBM20 loss. To achieve this, we applied pre-*TTN* O-MAP-seq[39] in both WT and ΔRBM20 cardiomyocytes (**Fig. 4a**). This analysis revealed hundreds of RNAs that co-compartmentalize with nascent *TTN* in each cell type, with significant though modest enrichment (average fold–enrichment, O-MAP/input: 2.21 (WT), 2.32 (ΔRBM20)l; **Supplementary Table 4**). That *TTN* associates with a specific cohort of transcripts in ΔRBM20 cells contrasts sharply with our O-MAP-ChIP results, wherein loss of RBM20 appeared to abolish *TTN* pre-mRNA′s affiliation with most chromatin (**Fig. 2**). This suggests that *TTN* doesn't solely compartmentalize with transcripts encoded by nearby genomic loci, but rather, that its chromosomal and RNA interactions are, to some degree, programmed independently. Indeed, relatively few *TTN*-proximal transcripts were encoded by either CID or TID loci (~7.2% and 0.7%, respectively, in WT cells; 4.7% and 0.6% in ΔRBM20 cells, **Fig. 4b**) and transcripts encoded on chromosome 2 did not appear enriched in either cell type (**Supplementary Fig. 6a**). Likewise, RBM20-targets were not enriched as a class in either cell type (**Supplementary Fig. 6b**). Collectively, these data suggest that the majority of transcripts captured by our O-MAP analyses were recruited to the *TTN* RNA factory post-transcriptionally, and perhaps represent a distinct set of interactions from those discovered previously[13].

Kania *et. al.* **Figure 4.**

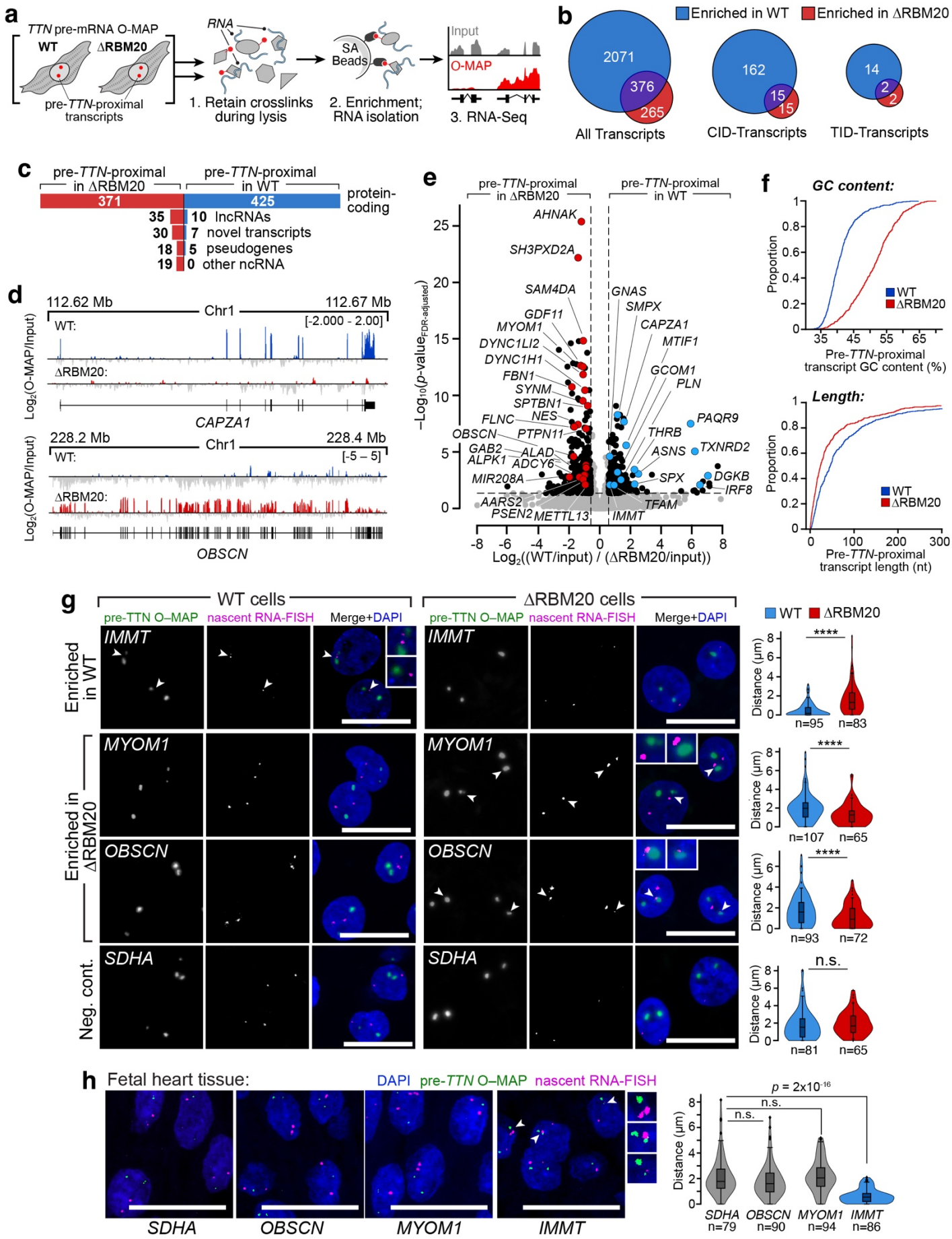

Figure 4. The *TTN* RNA factory compartmentalizes hundreds of mRNAs post-transcriptionally. a, Experimental scheme. Transcripts within *TTN* factories were mapped by O-MAP-Seq[39], applied in parallel to WT and ΔRBM20 CMs. As in O-MAP-ChIP (**Fig. 2**), transcripts were enriched by pulling down the biotinylated proteins to which they are bound and quantified by high-throughput sequencing. **b,** pre-*TTN*-proximal transcripts arise from loci outside the *TTN* RNA factory. Venn diagrams summarizing all (*left*), CID-enclosed (*middle*), and TID-enclosed (*right*) enriched transcripts, in both WT and ΔRBM20 cells. Enrichment was measured relative to matched samples of whole-cell RNA. **c**, Overview of the pre-*TTN*-proximal transcriptome. Most transcripts are coding, irrespective of RBM20 status. **d**, Examples of pre-*TTN*-proximal transcripts specifically enriched in WT (*top*) and ΔRBM20 (*bottom*) CMs. Both *CAPZA1* and *OBSCN* are RBM20 targets[43]. **e,** Volcano plot demonstrating differential enrichment of pre-*TTN*-proximal transcripts in WT- and ΔRBM20 cells (n=4 biological replicates). Notable WT-specific transcripts encoding Ca^{2+}–handling and metabolic proteins are highlighted in blue; ΔRBM20-specific transcripts encoding proteins implicated in DCM etiology are highlighted in red. **f,** Cumulative Distribution Function (CDF)-plots of pre-*TTN*-proximal transcripts' nucleotide content (*top*) and length (*bottom*). ΔRBM20-specific transcripts have markedly higher GC content and are modestly shorter than their WT-specific counterparts. **g,** Proximity of pre-*TTN* and novel interacting transcripts. *IMMT* is a WT-specific pre-*TTN* interactor; *MYOM1* and *OBSCN* are ΔRBM20-specific; *SDHA* is a negative control. O-MAP Imaging[39] was used to visualize pre-*TTN* (*green*); intron-targeting RNA-FISH visualizes putative interacting RNAs (*magenta*). Arrows denote co-compartmentalization, highlighted in zoomed images. All scale bars: 20 μm. Minimal distances were quantified and plotted on the right; the number of cells in each condition are indicated below. **h**, *IMMT*, but not *MYOM1* or *OBSCN*, co-compartmentalize with pre-*TTN* in fetal heart tissue. pre-*TTN* and other transcripts were visualized as in **g**; quantified data are plotted on the right. Internal box and whisker plots indicate median, 25th, and 75th percentile, and the 5th–95th percentile range. *p-values* by Kruskal-Wallis test followed by Dunn's multiple comparisons vs WT unless otherwise noted. n.s.: $p > 0.05$; ****: $p < 0.001$.

We next examined how the *TTN*-proximal transcriptome is altered upon loss of RBM20. Comparing the O-MAP-enriched transcripts from WT and ΔRBM20 CMs—each normalized to the corresponding input RNA—yielded catalogs of 447 and 473 RNAs that preferentially localize near nascent *TTN* in each cell type (**Fig. 4c**). These transcripts were predominantly coding, though in ΔRBM20 cells we observed increased enrichment of several classes of noncoding RNA (collectively ~22% of ΔRBM20-specific transcripts, vs. ~5% in WT, **Fig. 4c**). Notably, only a minority of RBM20 target transcripts (identified previously using eCLIP)[22] were differentially enriched in either cell type (**Supplementary Fig. 6b**), with rare examples found in both WT and ΔRBM20 cells (*e.g.,* **Fig. 4c**). Likewise, TID-encoded genes were not enriched in either cell type (**Supplementary Fig. 6c**). Collectively, these data suggest that cell type-specificity of the O-MAP pre-*TTN*-proximal transcriptome is not determined principally by transcripts' direct engagement with RBM20 or by changes in local chromatin architecture, but instead appears to be controlled through other facets of RNA structure. Indeed, transcripts' differential localization exhibited notable correlations with both length and sequence composition, as pre-*TTN*-proximal RNAs were markedly longer (by ~34 kb) and AU-rich (~10% lower GC-content) in WT CM's than in ΔRBM20 cells (**Fig. 4f**). We hypothesize that these structural preferences may be driven by the binding affinities of co-compartmentalized RNA-binding proteins, as described below.

We sought to experimentally validate that our novel O-MAP-enriched transcripts compartmentalize within *TTN* RNA factories by using intron-targeted SABER-FISH[45] to visualize the nuclear pools of these RNAs (**Supplementary Table 1**) and O-MAP-Imaging[39, 54] to visualize nascent *TTN.* These experiments largely confirmed the predictions made by our O-MAP-Seq data. For example, the WT-specific transcript *IMMT* exhibited preferential co-compartmentalization with nascent *TTN* in WT CMs (**Fig. 4g**, *top*), while the mutant-specific

transcripts *MYOM1* and *OBSCN* each preferentially co-compartmentalized with nascent *TTN* in ΔRBM20 CMs (**Fig. 4g,** *middle*). Of note, these findings were recapitulated in fetal human heart tissue, suggesting that the interactions discovered in our O-MAP model system are also formed *in vivo* (**Fig. 4h**). In all cases, transcripts compartmentalized near pre-*TTN* at mean displacements of ~0.5–1.4 μm, similar to the distances observed in our O-MAP *trans*-interacting chromatin domains (**Fig. 2g–j**; **Supplementary Fig. 2**), and comparable to structures like nuclear speckles[12]. This again highlights O-MAP's unique ability to capture compartment-level interactions that are challenging to probe by conventional tools.

The TTN RNA factory transcriptome is alternatively spliced upon RBM20 loss

Emerging data suggest a potential link between subnuclear RNA localization and alternative splicing, as several classes of nuclear bodies are proposed to influence isoform choice or to compartmentalize differentially spliced transcripts[39, 55-57]. We thus sought to characterize how loss of RBM20 might influence alternative splicing in the *TTN* RNA factory transcriptome. As expected, global changes in pre-mRNA splicing between WT and ΔRBM20 CMs appear largely to be driven by RBM20 loss, as principal component analysis clustered samples tightly by genotype (**Fig. 5a**). These global analyses identified 3,477 differential splicing events between WT and ΔRBM20 cells (FDR<0.05; ΔPercent Spliced-in (ΔPSI) = 0.1), with skipped-exon (SE) events exhibiting the largest change between cell types (3.15% difference; **Fig. 5b**). While this result is consistent with RBM20's established role in promoting exon exclusion[58], we noted that eCLIP-confirmed RBM20-target genes[22] represented only a small fraction (5.43 %) of observed SE events (**Fig. 5c**). This suggests that many such splicing changes might be attributed to pleiotropic or downstream effects induced by RBM20 loss, and we hypothesized that these effects might be more pronounced in the *TTN* RNA factory transcriptome. Indeed, we observed an enrichment of alternative splicing events in the WT- and ΔRBM20-specific O-MAP-enriched RNAs (**Fig. 4e**, and **Supplementary table 4**), relative to bulk RNA (~20% of *TTN*-proximal transcripts, compared to ~15% of all RNAs, $p = 2.6 \times 10^{-5}$; **Fig. 5d**, *middle*). These differences were largely due to skipped exon (SE), mutually exclusive exon (MXE), and retained intron (RI) events, which were significantly more pronounced in pre-*TTN*-proximal transcripts than in the broader transcriptome (**Fig. 5d**, *right*). Furthermore, the majority (~60%) of these alternative splicing events were specific to ΔRBM20-cells (**Fig. 5e**, *left*), with nearly 85% of RI events being unique to the knockout line (**Fig. 5e** and **Supplementary Fig. 7a**). Hence, in the absence of RBM20, O-MAP

Kania *et. al.* **Figure 5.**

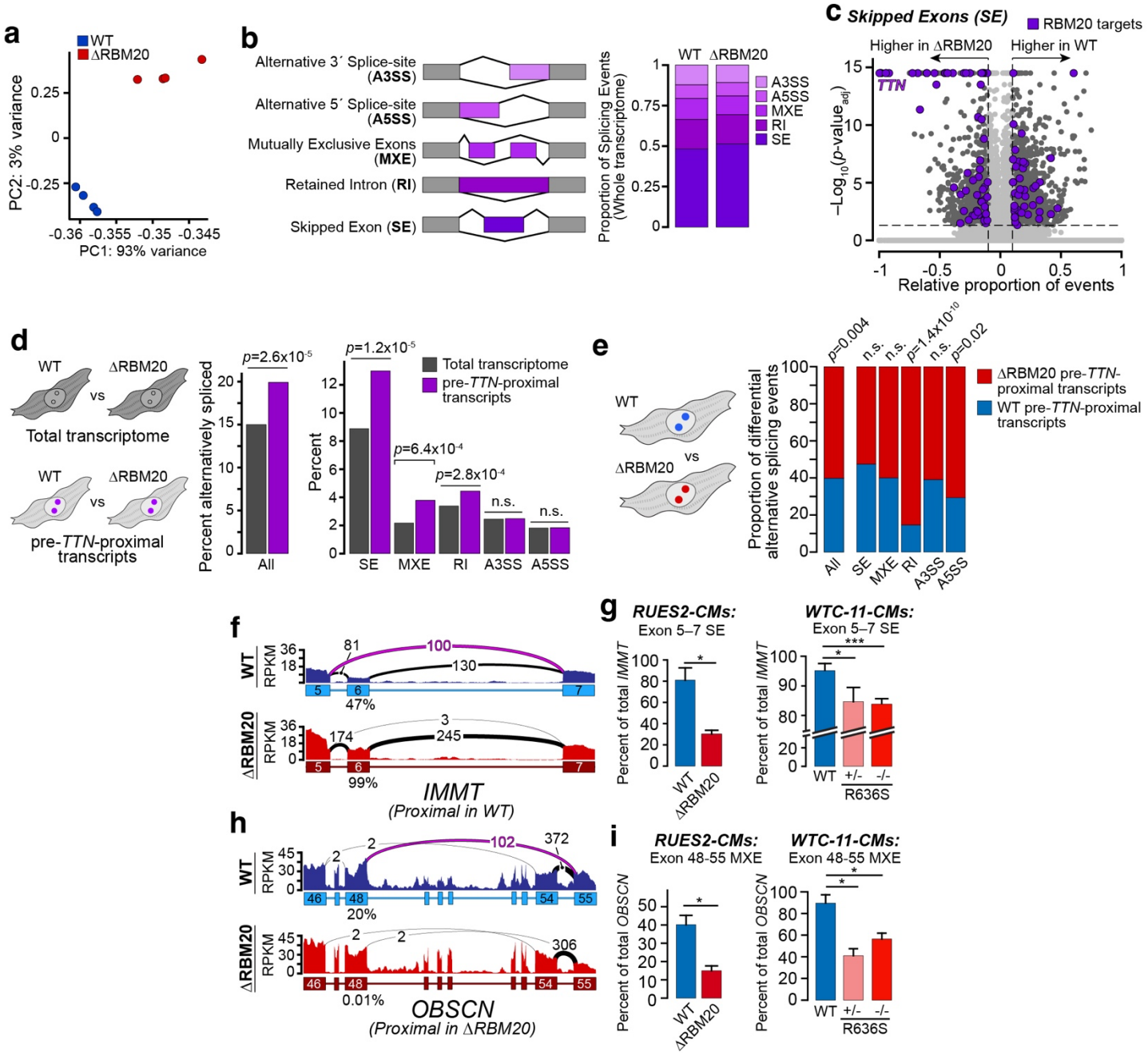

Figure 5. The *TTN RNA* factory enriches alternatively spliced transcripts. a, Loss of RBM20 induces global changes in pre-mRNA splicing. Principal component analysis (PCA) performed using Percent Spliced in (PSI) values for Skipped Exon events that differed significantly between WT and ΔRBM20 cells (FDR < 0.05). **b**, *Left:* schematic summarizing the major class of alternate splicing. *Right:* percentage of differential splicing events between WT and ΔRBM20 CMs, observed transcriptome-wide (FDR< 0.05, ΔPSI > 0.1). **c,** Dysregulated SE events in ΔRBM20 cells cannot be explained by RBM20 loss alone. Overview of transcripts undergoing differential SE alternative splicing between WT and ΔRBM20, measured transcriptome-wide. Confirmed RBM20 target transcripts observed by eCLIP[22] shown in purple. **d,** Alternative splicing is more prevalent in pre-*TTN*-proximal transcripts. *Left:* analytical approach. Differential splicing between WT and ΔRBM20 CMs was quantified for both the total transcriptome (*blue*) and pre-*TTN*-proximal transcripts (*purple*). Compared to the global transcriptome, pre-*TTN*-proximal transcripts exhibited more alternative splicing overall (*middle*), driven largely by increases in SE, MXE, and RI events (*right*) (*two-sided z-test).* **e,** Dysregulated splicing is more prevalent in ΔRBM20-specific pre-*TTN*-transcripts. Proportion of total differential alternative splice events are shown; *p*–values compare WT- and ΔRBM20-specific transcripts from each splicing event class (*two-sided z-test*). **f–i,** Examples of altered splicing in WT-specific SE event (*IMMT*) and ΔRBM20-specific MXE event (*OBSCN*) in pre-*TTN*-proximal transcripts. **f, h,** Sashimi plots visualizing splicing patterns in each cell type. Numbers below exons 6 (*IMMT*) and 48 (*OBSCN*) indicate the percent retention of the that exon. Splicing events highlighted in magenta were targeted for qRT–PCR quantification. **g, i**, qRT–PCR analysis of **g,** an SE event between *IMMT* exons 5–7, and **i,** an *MXE* event between *OBSCN* exons 48–55. *Left:* female ESC-derived ΔRBM20 CMs; *Right:* male iPSC-derived CMs bearing DCM-associated point mutations (*n*=4 biological replicates, run in technical quadruplicate; *p*-value: paired two-sided students t-test). Unless otherwise noted, in **b–e,** differential alternative splicing events were defined by (FDR<0.05; ΔPSI > 0.1).

pre-*TTN*-proximal transcripts appear to suffer disproportionately from defects in splicing. Notably, Gene Ontology analysis[51] of these aberrantly spliced, *TTN*-proximal transcripts revealed the enrichment of genes involved in actin cytoskeleton organization, muscle cell differentiation, and striated muscle cell development (**Supplementary Fig. 7b**). This suggests that preferential mis-splicing of these transcripts may contribute to some of the cryptic cellular dysfunction observed in RBM20-DCM[59-61]. To explore this further, we used isoform-specific quantitative RT-PCR to examine whether the same *TTN*-proximal transcripts are mis-spliced in other cellular models of DCM, using a series of WTC-11 iPSC-derived cardiomyocyte lines bearing the DCM-associated RBM20 R636S mutation[22] (**Supplementary Table 5**). A SE event we observed in *IMMT* (pre-*TTN*-proximal in WT cells, **Fig. 4g**) and a MXE event we observed in *OBSCN* (pre-*TTN*-proximal in ∆RBM20 cells, **Fig. 4d,g**) were each recapitulated in both heterozygous and homozygous R636S lines (**Fig. 5f–i**). Neither of these splicing events is thought to be directly regulated by RBM20[22]. This suggests that our observed remodeling of the pre-*TTN*-proximal transcriptome is not unique to our RUES2-derived cardiomyocyte model, nor does it require complete loss of RBM20.

Collectively, these findings demonstrate that RBM20 loss not only induces mis-splicing of RBM20-target transcripts but also remodels the composition and splicing of the *TTN* factory transcriptome. In this state, nascent *TTN* preferentially co-compartmentalizes with other mis-spliced RNAs, whose dysregulation may contribute to DCM etiology.

The TTN RNA factory proteome couples chromatin organization and RNA compartmentalization

Having mapped new facets of the *TTN* RNA factory's chromatin architecture and compartmentalized transcriptome, we next sought to characterize its proteome. However, our mass spectrometry pipeline, O-MAP-MS, was initially optimized for more abundant target RNAs (e.g., 47S pre-rRNA; *Xist*)[39], which yield significantly more biotinylated material than the trace signal from pre-*TTN* O-MAP (**Fig. 1c**). To improve experimental sensitivity, we adopted a modified protocol that enriches nuclei prior to pulldown (reducing background noise) and which minimizes protein loss during streptavidin enrichment and peptide workup (**Fig. 6a**).

Applying this modified O-MAP-MS workflow, and manually filtering for proteins with established nuclear localization, we identified lists of 254 and 180 candidate factors enriched in the *TTN* RNA Factories of WT and ∆RBM20 cardiomyocytes, respectively (**Supplementary Table 6**). RBM20 was enriched in WT cells, as

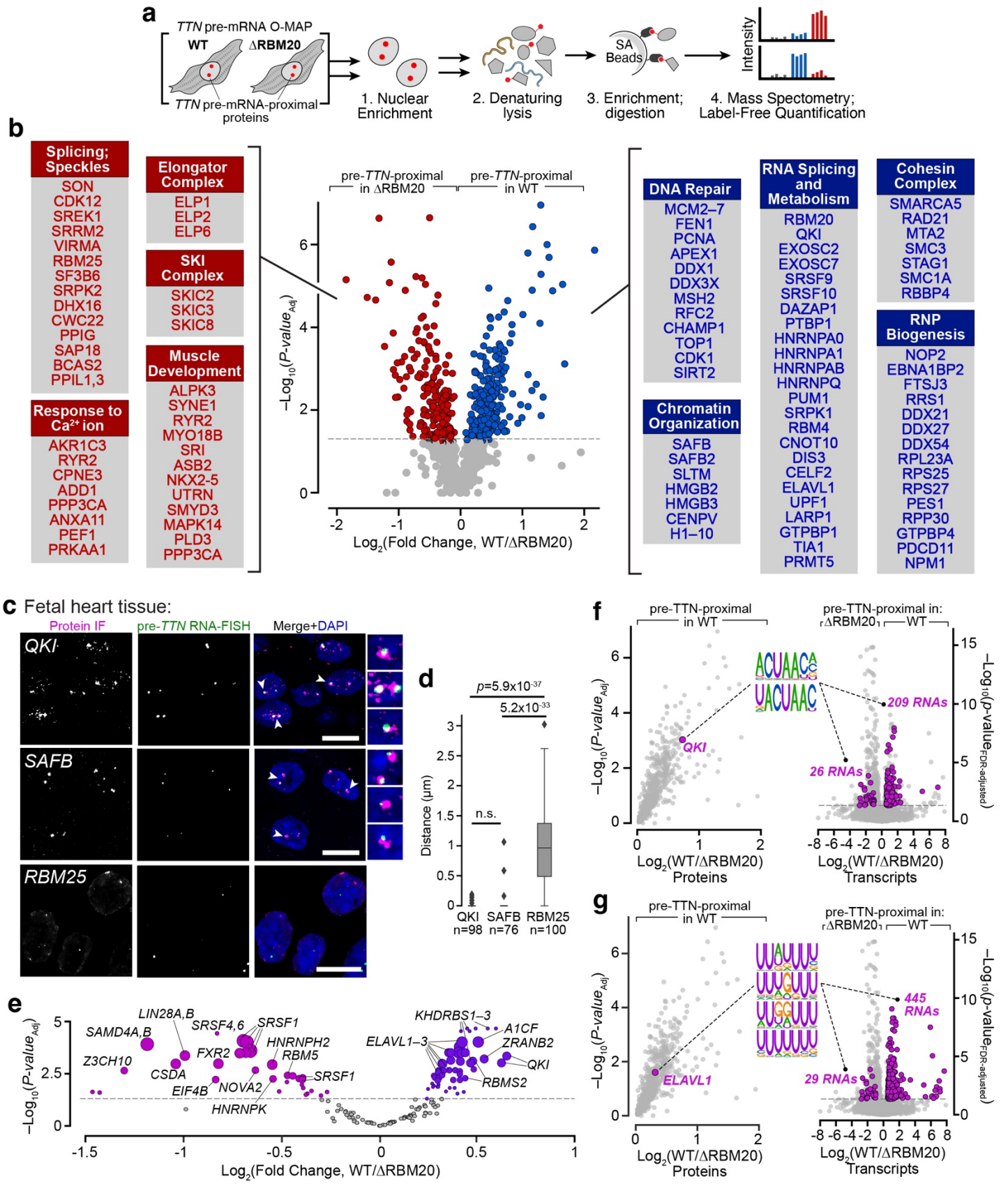

Figure 6. The TTN RNA factory proteome couples chromatin architecture to RNA compartmentalization. a, Experimental scheme. pre-*TTN* O-MAP was performed in parallel in WT and ΔRBM20 CMs. Following nuclear isolation, biotinylated proteins were enriched under denaturing conditions and analyzed by mass spectrometry with label-free quantification (LFQ). **b,** Volcano plot summarizing pre-*TTN* O-MAP-MS hits specific for WT (*blue; right*) and ΔRBM20 (*red; left*) CMs. Prominent functional classes and representative protein

hits are shown for each. **c,** Novel proteomic hits compartmentalize within *TTN* RNA factories. Immunofluorescence for two WT-specific hits, QKI and SAFB, and negative control, RBM25, in WT human fetal heart tissue sections. Nascent *TTN* was visualized by RNA-FISH. Arrowheads denote conspicuous colocalization, highlighted in zoomed images (*right*). All scale bars, 20 μm. **d,** quantification of 3D-distances between pre-*TTN* and each protein hit. Box and whisker plots indicate median, 25th, and 75th percentile, and the 5th–95th percentile range. *p*-values by Kruskal-Wallis test followed by Dunn's multiple comparisons, vs RBM25. **e,** Differential enrichment of RNA-binding protein recognition motifs in WT- and ΔRBM20-specific pre-TTN-proximal transcripts. The size of each motif's circle is proportional to the total number of transcripts harboring that motif. Significant differential motifs are highlighted in purple (Log_2(Fold Change)>0.58; FDR<0.05). **f–g**, *TTN* RNA factories compartmentalize RNA-binding proteins (RBPs) along with their target transcripts. *Left*: significant enrichment of each protein (*highlighted in magenta*) in O-MAP-MS. Data are the same as in **b**. *Middle*: Logos for the major RNA recognition motifs for each RBP. *Right*: Incidence of these motifs within the WT- and ΔRBM20-pre-TTN-proximal transcriptomes, observed by O-MAP-Seq (*highlighted in magenta*). Data are the same as in (**Fig. 4e**). The number of transcripts bearing each binding motif is indicated. **f,** QKI; **g,** ELAVL1.

---

anticipated. We also observed the marked enrichment of factors involved in chromosome organization and compartmentalization (**Fig. 6b**, *right*), consistent with the factory's potential role in coordinating higher-order chromatin architecture (**Fig. 2**). These included nearly all members of the core cohesin complex—the master regulator of chromatin looping and folding[62]—and factors like HMGB2 and 3, which play pervasive roles in chromatin organization[63]. This group also included SAFB and its homologues, SAFB2 and SLTM, which were originally identified as components of the putative "nuclear matrix," and have recently been implicated in mediating RNA-directed chromatin compartmentalization[1, 64]. We furthermore observed the prominent enrichment of DNA repair factors—including FEN1, PCNA, and APEX1—and all members of the MCM complex, which plays a "moonlighting" role in chromatin folding by antagonizing cohesin-mediated loop-extrusion[65]. These chromatin-regulatory hits were joined by a cluster of RNA-metabolic and RNP-biogenesis factors. These included proteins that play broad roles in general RNA metabolism (*e.g.*, hnRNPAB; EXOSC2), and those that regulate specific cohorts of target transcripts (*e.g.*, SRSF9; QKI; PUM1). We also observed marked enrichment for several nucleolar proteins (*e.g.*, PES1, NPM1), consistent with the factory's frequent localization to the nucleolar periphery. Collectively, these data provide insight into the mechanisms by which the factory couples its dual roles in chromatin-organization (**Figs. 2,3**) and RNA-compartmentalization (**Fig. 4b**).

Critically, loss of RBM20 appears to grossly reorganize the factory proteome (**Fig. 6b**, *left*). In ΔRBM20 cells, pre-*TTN* O-MAP-MS yielded no specific enrichment of chromatin-architectural factors, consistent with our observation that nascent *TTN* requires RBM20 to affiliate with chromatin (**Fig. 2e**). Instead, we observed a varied assortment of core splicing factors (*e.g.*, the promiscuous splicing regulator RBM25)[66] and nuclear speckle proteins (*e.g.*, SON, SRRM2). This suggests that, in the absence of RBM20, nascent *TTN* localizes to (or nucleates[17]) a generic nuclear speckle.

We next sought to validate our novel hits by combined immunofluorescence and pre-*TTN* RNA-FISH in human fetal heart tissue. Strikingly, both SAFB and QKI (each enriched in WT O-MAP) formed prominent nuclear

puncta that overlapped almost exclusively with nascent *TTN* transcripts, whereas RBM25 (enriched in ΔRBM20 O-MAP) showed no significant colocalization (**Fig. 6c,d**). SAFB is an understudied scaffolding factor with both chromatin- and RNA-binding functions[67], and has recently been implicated in playing a role in organizing other RNA-scaffolded subnuclear compartments[1, 64]. That this factor is almost exclusively compartmentalized in the *TTN* RNA factory suggests that, in cardiomyocytes, it may play a similar role in coupling *TTN* mRNA biogenesis to chromatin organization. Likewise, QKI is a developmentally regulated splicing factor with putative roles in mRNA nuclear export[68] and which has been recently implicated in contributing to DCM etiology[69], potentially through its interactions with *TTN*-derived noncoding transcripts[70]. Its pronounced co-compartmentalization with nascent *TTN* suggests that the *TTN* RNA factory may coordinate and regulate QKI target transcripts—as well as those recognized by the other factory-resident RNA-binding proteins (RBPs).

To explore this hypothesis, we examined whether known RBP recognition motifs were enriched in WT and ΔRBM20-specific pre-TTN-proximal transcripts (**Fig. 4e**; **Supplementary Table 4**), using the Transite pipeline[71]. In the WT transcriptome, we observed a marked enrichment of binding motifs recognized by several of the RBPs captured by O-MAP-MS (**Fig. 6e**). Among the most prominent of these was QKI, for which consensus binding motifs were observed in nearly half of WT-specific *TTN* factory transcripts (~47%), but which were largely depleted in the ΔRBM20 factory transcriptome (~5.5% of transcripts; **Fig. 6f**). Enrichment was even more striking for ELAVL1 (HuR), for which cognate motifs were found in nearly all WT-specific transcripts (99.6%), and which were likewise depleted from the ΔRBM20 factory (~6.1% of transcripts; **Fig. 6g**). We note that RBM20 targets could not be identified by these analysis, as the splicing factor's short cognate motif (5´–UCUU–3´)[72] fell below Transite's minimal length requirement[71]. The prevalence of ELAVL1's U-rich binding motif may also explain the pronounced AU bias observed in the WT factory transcriptome (**Fig. 4f**, *top*). Remarkably, in ΔRBM20 cells we did not observe a similar concordance between the *TTN*-proximal proteome and transcriptome, as none of the RBP motifs enriched in ΔRBM20-specific transcripts corresponded to O-MAP-MS hits. Given the sparseness of our proteomic data, it is possible that these RBPs were enriched below the limit of detection. Alternatively, given the disproportionate mis-splicing of ΔRBM20-specific transcripts (**Fig. 5e–i**), it is possible that the compartmentalization of these RNAs is controlled by nonspecific RNA quality-control pathways, and not by programmatic interactions with a defined set of RBPs. Regardless, even these sparse data underscore nascent transcript O-MAP's ability to probe RNA compartmental proteomes in contexts where input

material is limiting.

DISCUSSION

In this study, we adapted the versatile RNA interaction-discovery tool O-MAP into a first-of-its-kind method for biochemically dissecting discrete subnuclear compartments, by targeting the technique to nascent transcripts and maximizing the sensitivity of our target-recovery workflows. Based on the multiomic O-MAP analysis presented here, we propose a working model for the molecular architecture of the *TTN* RNA factory—to our knowledge, the first such model to be attained for a single-locus subnuclear compartment (**Fig. 7**, *left*). In WT cells, the factory compartmentalizes *TTN* mRNA biogenesis within a specialized chromatin neighborhood, which comprises an array of *cis*-interacting domains (CIDs) on chromosome 2, and multiple *trans*-interacting domains (TIDs) interspersed across other chromosomes (**Fig. 2b**–**d**). These domains assemble around a micron-wide compartment surrounding unspliced *TTN* transcripts (**Fig. 2f–i**), which appears to hold many genes within these TIDs in a transcriptionally repressed state (**Fig. 3**). We hypothesize that this higher-order compartmentalization is driven by a cohort of chromatin-architectural proteins, including the cohesin complex[62] and scaffolding factors like SAFB[64] and HMGB2[63]—each of which is highly enriched within the factory (**Fig. 6b,c**). These are joined by an array of DNA-repair factors, potentially suggesting that synthesis of the ~278 kb *TTN* transcript may induce DNA damage, as has been observed with other long, repeat-laden transcripts[73]. In principle, these factors might also play a role in chromatin compartmentalization, by organizing foci near the site of *TTN* transcription[65, 74], or by imposing topological constraints on cohesin-mediated chromatin looping[65]. Concomitantly, the factory compartmentalizes dozens of RNA-binding proteins, including factors like QKI and ELAVL1 (**Fig. 6b**), which in turn recruit hundreds of their target RNAs (**Figs. 4e** and **6f**), presumably post-transcriptionally (**Fig. 4b** and **Supplementary Fig. 6**). This suggests that, in addition to its roles in mediating *TTN* biogenesis and TID-silencing, the factory may also provide a post-transcriptional hub that sequesters a defined ensemble of target transcripts. This compartmentalization phenomenon has been observed in other nuclear bodies—including nuclear speckles[55], paraspeckles[75], nuclear stress bodies[76], and nucleoli[77, 78]—though the functional significance of RNA sequestration in the *TTN* Factory remains unclear.

In the absence of RBM20, nearly every facet of this molecular architecture is remodeled or altogether lost (**Fig. 7**, *right*). We hypothesize that this gross remodeling may underlie some of the cryptic facets of RBM20-DCM etiology[79], and hope that the extensive molecular characterization presented here will facilitate future

Kania *et. al.* **Figure 7.**

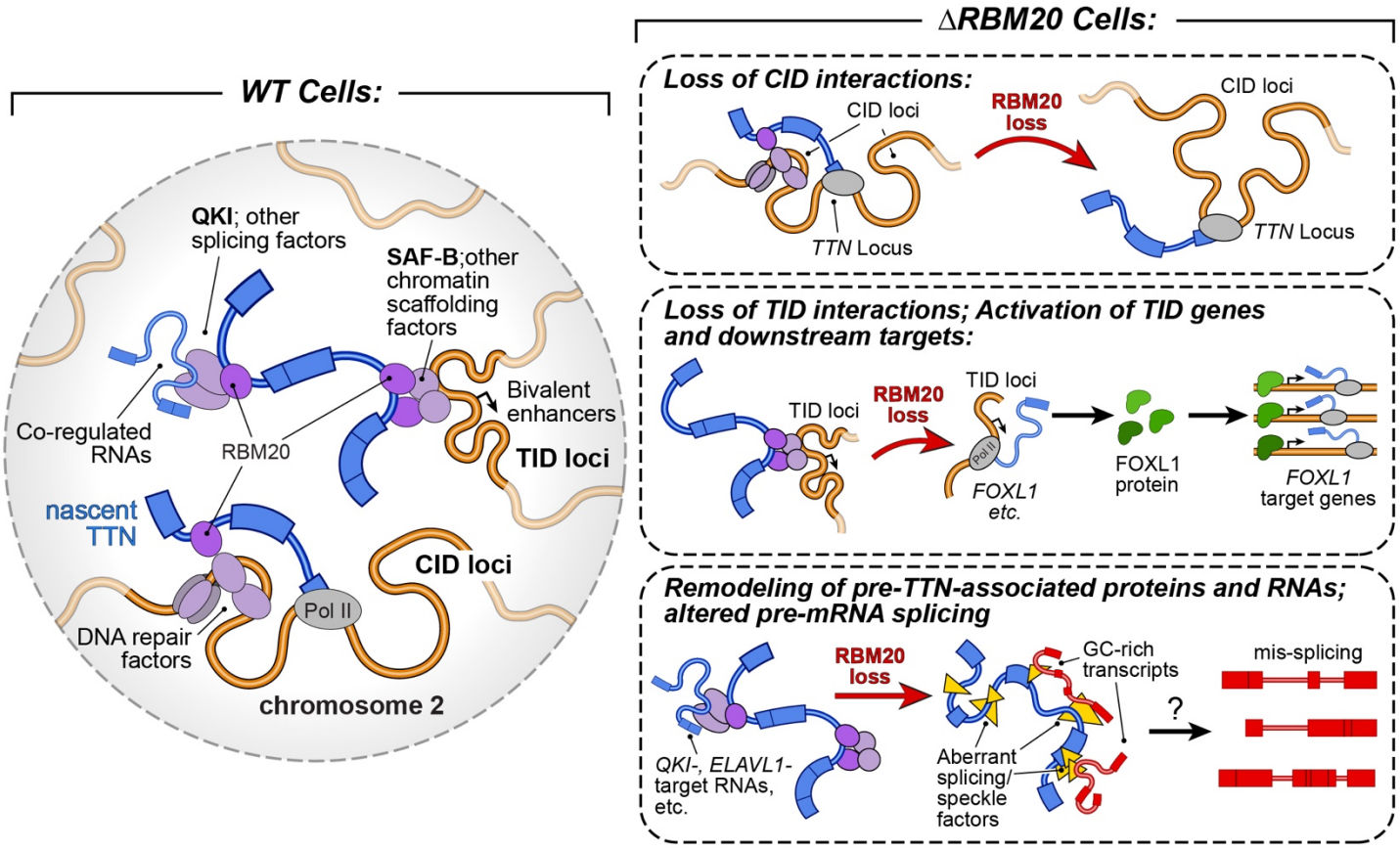

Figure 7. Molecular architecture of the *TTN* RNA factory. *Left*: overview of the compartment-wide interactions discovered by O-MAP. For simplicity, factory components are depicted as directly binding to pre-*TTN* and RBM20, though these may interact through higher-order assemblies or biomolecular condensates. Our data cannot discriminate between interactions with nascent transcripts and incompletely spliced intermediate species, both of which are depicted. "co-regulated RNAs:" QKI- and ELAVL1-bound transcripts, among others. *Right*: broad summary of architectural alterations induced by RBM20 loss.

mechanistic studies. Of note, in ΔRBM20 cells, nascent *TTN* appears incapable of affiliating with higher-order chromatin architecture, including other active loci on its own chromosome (**Fig. 2e–i**). It is unclear what influence RBM20 has on CID structure, which we hypothesize is largely determined by chromosomal A/B compartmentalization (**Fig. 2b**). That said, RBM20 loss appears to grossly disrupt TID organization, disassociating these domains from *TTN's* compartment (**Fig. 2g–i**), and reactivating TID-enclosed genes, with potential downstream effects on global gene expression (**Fig. 3**). Concomitantly, nascent *TTN* appears to compartmentalize into a vastly different subnuclear microenvironment, devoid of the unique proteomic markers observed in WT cells (**Fig. 6b**) and co-compartmentalized with a distinct ensemble of transcripts. Compared to their WT counterparts, these ΔRBM20-specific *TTN*-proximal transcripts appear disproportionately mis-spliced, (**Fig. 5e–i**) though few are direct RBM20 targets (**Fig. 5c**). We speculate that the denuded *TTN* transcript may help drive this aberrant splicing, potentially sequestering essential splicing factors and poisoning the biogenesis of other RNAs. A similar mechanism is thought to underlie the toxic RNA aggregates formed in repeat-expansion

disorders[80]. We note that our data are also consistent with a model in which aberrantly spliced *TTN* is passively sequestered into a speckle-like compartment, along with other mis-spliced transcripts.

The model emerging from this study significantly enhances our understanding of the molecular architecture and functional scope of the *TTN* RNA factory. Prior work has demonstrated that nascent *TTN* organizes a clique of RBM20 target genes, thereby coordinating these genes' biogenesis[13, 29]. Our work expands these findings, revealing that *TTN* transcripts concomitantly coordinate a hub of repressed chromatin and post-transcriptional RNA compartmentalization. This highlights O-MAP's ability to complement the data revealed by conventional methods (*e.g.*, Hi-C). Hi-C and O-MAP may preferentially enrich for chromatin interactions at different physical scales, since DNA proximity-ligation-based tools like Hi-C predominantly capture nanometer-scale interactions[26], while O-MAP relies on biotinyl radicals that diffuse over much longer distances (~100 nm)[42]. The two methods are also anchored on distinct molecular targets, since Hi-C captures direct chromatin-chromatin interactions (here, centered around the *TTN* locus), while O-MAP probes all chromatin, transcripts, and proteins adjacent to the target RNA (**Fig. 1b**).

These technical differences may also explain some of the complexities in our O-MAP-Seq results, which revealed a surprisingly imperfect overlap with our earlier RBM20 eCLIP data[22]. We do not believe that this discrepancy is an artifact of the O-MAP pipeline—both because this pipeline has been extensively validated on other target RNAs[39], and because the present data support the accuracy of our new results. Notably, the validation of O-MAP-Seq hits by RNA-FISH (**Fig. 4g**–**h**), and the observation that many of these hits represent putative targets for factory-resident RBPs (suggesting a straightforward mechanism for their compartmentalization, **Fig. 6f**–**g**) reinforce the reliability of our findings. Hence, we currently favor the notion that the *TTN* RNA factory comprises only part of the RBM20-associated transcriptome, and that many target transcripts may be regulated by the diffuse nucleoplasmic pool of RBM20, outside the factory[13, 16]. Moreover, while our intron-targeting approach was designed to capture nascent *TTN* transcripts, other *TTN*-derived species—including alternatively spliced isoforms retaining the targeted introns, improperly spliced transcripts undergoing surveillance or decay, and circular RNAs[15] —would, in principle, also be viable O-MAP targets. It is hence possible that O-MAP has captured additional compartments, or an ensemble of compartments, populated by alternatively spliced, *TTN*-derived transcripts. Hence, our analysis may have revealed a new dimension of genome compartmentalization by the *TTN* RNA at distinct stages of its lifecycle, complementary to the insights

previously obtained by Hi-C.

We also note that our revised architectural model is not entirely unprecedented, and indeed, that it resembles a class of subnuclear compartments termed nuclear stress bodies (nSBs)[7]. Like the *TTN* factory, nSBs are assembled on long, repetitive transcripts and form prominent nuclear foci of SAFB. These foci serve as recruitment sites for defined cohorts of RBPs, many of which are shared by the *TTN* factory (*e.g.* SRSF1, SRSF9), and appear to drive the post-transcriptional sequestration of these proteins' target transcripts[7, 76]. It is thus possible that the *TTN* RNA factory and nSBs represent foundational examples of a conserved assembly mechanism, which utilizes nascent transcripts to scaffold the formation of functionally distinct subnuclear regulatory compartments. This kind of RNA-mediated compartmentalization has been largely ascribed to noncoding RNAs[18]. But in the nucleus, all RNA is noncoding. That nascent pre-mRNA like *TTN* appears capable of "moonlighting" as a chromatin-architectural organizer raises the question of whether more mRNAs may play a similar role in other developmental and disease contexts.

This characterization study highlights O-MAP's unique power to capture facets of nuclear architecture that are opaque to most conventional approaches. Conventional approaches toward interrogating nuclear structure have focused on nm-scale chromatin-chromatin interactions, using proximity-ligation-based methods like Hi-C[2]. Though powerful, these methods have difficulty probing the µm-scale, compartment-level interactions that define the *TTN* RNA factory[29] and many other subnuclear compartments, and of course, they do not provide information on RNA or protein content within the compartment[6]. In contrast, applying O-MAP to a compartment-resident transcript (nascent *TTN*) enabled us to readily capture higher-order TID interactions (**Fig. 2**) and ensembles of co-compartmentalized RNAs (**Fig. 4**), revealing a host of new regulatory phenomena. Previous efforts at characterizing these kinds of compartmental interactions have relied on tools like subcellular fractionation[81], affinity purification of a tagged compartmental protein[32], or targeted *in situ* proximity-biotinylation of that protein[4]. These methods simultaneously survey all structures of a given type *en masse,* and hence cannot identify the unique molecular features that might define an individual compartment or locus. The novel O-MAP strategy we have implemented here overcomes this longstanding technical roadblock, enabling us to characterize, for the first time, the distinct molecular microenvironment surrounding a single genomic locus of interest. We anticipate that this flexible approach can be applied in myriad other settings.

DATA AVAILABILITY AND ACCESSION CODES

Mass spectrometry data have been deposited to the MassIVE repository with the accession number MSV000096046, and to PRIDE, with the accession number PXD056627. RNA- and DNA-sequencing data have been deposited to the Gene Expression Omnibus under SuperSeries accession number GSE234811. Raw imaging files and custom code used in this work is available upon request. This manuscript made use of public datasets from ENCODE, accession numbers ENCSR584MHY and ENCFF135PCO.

ACKNOWLEDGEMENTS

We thank A. Tsue for technical assistance with O-MAP, H. Reinecke and J. Klaiman for assistance with cardiomyocyte differentiation, N. Peters and members of J. Scott's laboratory for assistance with imaging, S. Henikoff for the generous donation of reagents, members of Y. Sancak's lab for their mitochondrial expertise, I. Ulitsky for insight into O-MAP-Seq data analysis, and members of the Shechner, Bertero, and Murry labs for their support, thoughtful advice, and critiques. This work was supported by National Institutes of Health Grants R01HL160825 (to E.E.K., A.F., C.E.M., and D.M.S.), R01GM138799 (to E.E.K. and D.M.S.), T32GM007750 (to E.E.K.), UM1HG011531 (to B.H., and W.S.N.), R35GM137916 (to C.K.C. and B.J.B.). E.E.K. was further supported by the NSF DEB2016186 and AHA 902616; A.B. received funding from the European Research Council (ERC) under the European Union’s Horizon Europe research and innovation programme (Grant agreement No. 101076026; Project Acronym TRANS-3); C.E.M also was supported by a grant from the Robert B. McMillen Foundation. Imaging at the UW Keck Center was supported by NIH S10 OD016240 and the UW Student Technology Fee (STF). This work used an EASY-nLC1200 UHPLC and Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer purchased with funding from a NIH S10OD021502. NGS data analysis was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington, and supported by the UW Student Technology Fund.

AUTHOR CONTRIBUTIONS

E.E.K. contributed to experimental design, performed most wet lab experiments, and oversaw initial data analysis. A.F. and A.B. assisted with cardiomyocyte differentiation and O-MAP probe design. D.M.M., R.F.,

D.K.S., and S-E.O. assisted with mass-spectrometry experiments. C.L., S.B., B.H., and W.S.N. performed follow-up omics data analysis. M.L. assisted with imaging experiments. C.K.C. and B.J.B. assisted with DNA-FISH experiments and image analysis. D.M.S., A.B., and C.E.M. initially conceived of this work; D.M.S. directed research. All authors contributed to writing.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests. E.E.K, B.J.B, and D.M.S. have filed for a patent concerning the use of oligonucleotide-directed proximity-labeling to elucidate and visualize subcellular interactions *in situ.* C.E.M. has equity in Sana Biotechnology and StemCardia, Inc; neither organization participated in this study. The remaining authors declare no competing interests.

CORRESPONDING AUTHORS

Correspondence to: Alessandro Bertero (alessandro.bertero@unito.it); Charles Murry (charles.murry@med.usc.edu); David M. Shechner (shechner@uw.edu)

MATERIALS AND METHODS

Cell culture and tissue sections

RUES2 hESCs (RUESe002-A; WiCell) and WTC-11 hiPSCs (a gift of Bruce R. Conklin; also available from Coriell GM25256) were maintained on Matrigel (Corning; 354277) in Essential 8 media (Thermo Fisher; A1517001). Media was exchanged every other day. Cells were passaged with Versene (Thermo Fisher; 15040066) into low confluent wells supplemented with 10 μM Y-27632 (Stem Cell Technologies; 72302).

hiPSC and hESC lines were differentiated into hiPSC- or hESC-cardiomyocytes (CMs) as described previously[13, 22]. Briefly, 4.5–9 x 10^4 cells were plated onto 12-well plates pre-coated with 80 μg/μl growth factor-reduced Matrigel in mTeSR1 (StemCell Technologies; 85850) and 10 μM Y-27632, for 24 hours. Cells were then treated with 1 μM CHIR-99021 (Cayman Chemical; 13122). On day 0, media was exchanged for RPMI-1640 with glutamine supplemented with 500 μg/ml Bovine Serum Albumin (BSA), 213 μg/ml ascorbic acid (RBA media; Millipore Sigma), and 5 μM CHIR-99021, and maintained for 48 h. On day 2, culture media was replaced with RBA media supplemented with 2 μM WNT-C59 (Selleck Chemicals; S7037) and maintained for 48 h. On day 4,

media was replaced with RBA. On day 6, cells were cultured in RPMI supplemented with B27 (Thermo Fisher; 17504044), with media exchanges every two days. To increase cell viability after cryopreservation and thawing, on day 13 cells underwent a 45-minute heat shock at 42 °C, using pre-warmed media. On day 14, cells were dissociated with TrypLE (Thermo Fisher; 12563011) with gentle trituration and were strained with 100 μM filters (Fisher Scientific; 22363549). Dissociated cells were frozen in CryoStor media (1 million cells/100 μL CryoStor; Millipore Sigma; C2874) and frozen slowly at -80 °C using a Mr. Frosty Freezing Container (Thermo Fisher; 5100-0001). The next day cells were stored in liquid nitrogen.

Formaldehyde-fixed and paraffin embedded fetal heart tissue sections were acquired from the National Disease Research exchange. The University of Washington Pathology Research Services Laboratory was used to cut blocks into 5 μm sections.

pre-TTN O-MAP main protocol

RUES2-CMs were used for both O-MAP omics and O-MAP imaging. For omics, CM stocks were thawed onto 12-well plates coated in Matrigel at 1.2 million cells/well in RPMI/B27, supplemented with 10 μM Y-27632 and 10% FBS. For imaging, 2.5 x 10^5 cells were plated onto 10mm glass bottom plates (Cellvis; D35-10-1.5-N). The next day media was exchanged for RPMI/B27 and exchanged every two days thereafter, until post-differentiation day 21. The core O-MAP protocol follows the same two-day protocol as previously described[39], but with slight optimizations for cardiomyocytes. Briefly, during the fixation step, formaldehyde concentration was increased to 4% (v/v; increased from the standard 2%), and endogenous peroxidase inactivation proceeded for 20 minutes (increased from the standard 10 minutes). The remainder of the protocol is largely unchanged from that described previously[39], and is summarized here.

pre-TTN O-MAP day 1

Cells were quickly washed with 1x DPBS (Thermo Fisher; 14190250) and fixed in 4% (v/v) formaldehyde (Electron Microscopy Sciences; 15710) in 1x PBS (Sigma; 6506) for 10 minutes at room temperature, without agitation. The fixation solution was aspirated and quenched by two five-minute washes with 250 mM glycine in 1x PBS, with rocking (3 rpm on a platform rocker; Everlast Rocker 247). Cells were then briefly washed with DPBS and permeabilized with 0.5% (v/v) triton-X 100 (Thermo Fisher; 28314) in PBS for 10 min with rocking.

Three five-minute DPBS washes were then performed to prevent over-permeabilization. To inactivate endogenous peroxidases, samples were then incubated with 0.5% H_2O_2 (v/v, in PBS) for 20 minutes with rocking. Two five-minute washes with DPBS followed to prevent H_2O_2 damage to cells. Cells were then equilibrated in Formamide Wash Buffer (40% (v/v) deionized formamide (VWR, 0314); 2X SSC (Thermo Fisher; AM9765); 0.1% Tween-20 (Thermo Fisher; 28320) for five minutes with rocking. Wash buffer was then aspirated, and to each well was added 50 µL of Primary Probe Mix, comprising 0.1 µM aggregated primary oligo probe pool (**Supplementary Table 1**), in 1x Hybridization Buffer (40% (v/v) deionized formamide; 2x SSC; 0.1% (v/v) Tween-20; 10% (w/v) dextran sulfate (SIGMA; D8906); in nuclease-free water). A clean 12-well glass coverslip was then applied to spread the Primary Probe Mix across the samples. A kimwipe soaked in nuclease-free water was placed between the wells, and plates were sealed with Parafilm to create a hybridization chamber. Plates were then incubated for 8 hours at 42°C without rocking.

pre-TTN O-MAP day 2

After primary hybridization, ~1 mL pre-warmed 30% Formamide Wash Buffer was added to the samples. Coverslips were then removed, and cells were allowed to wash for five-minutes rocking. Two additional five-minute washes with fresh 30% Formamide Wash Buffer followed. To each well was added 50 µL of O-MAP Secondary Probe Mix, comprising 0.1 µM HRP-oligo conjugate (**Supplementary Table 1**) in 30% Hybridization Buffer. Samples were covered with a clean coverslip and incubated for one hour at 37°C in the dark. For pre-*TTN* O-MAP imaging, endogenous biotin blocking was performed (see "*pre-TTN O-MAP Imaging*," below). Otherwise, to each well 1 mL of PBST (0.1% (v/v) Tween-20, in 1X PBS) was added. Coverslips were then removed, and samples washed for 15 minutes in PBST at room temperature, with gentle rocking. Samples were washed three times in PBST (15 minutes each), aspirated. To initiate *in situ* biotinylation, samples were treated with 1 mL of Labeling Solution (0.8 µM biotinyl-tyramide (Sigma; SML2135); 1 mM H_2O_2, in 1x PBST) for 45 minutes at room temperature, protected from light. The biotinylation reaction was quenched by the addition of 1 mL of PBST-q (10 nM sodium azide (Sigma; 71289), 10 nM sodium ascorbate (Sigma; A4034) in 1X PBST), for three five-minute washes, with gentle rocking.

pre-*TTN *O-MAP Imaging

After secondary hybridization with HRP-conjugated oligo, samples were washed three times in PBST; five minutes per wash. To block the signal from endogenous biotin, samples were incubated for 30 minutes with 1 mL 1% nuclease-free BSA (VWR; 97061-420) in PBST, at room temperature with rocking, followed by a 15-minute incubation with 1 mL Neutravidin solution (10 µg/mL neutravidin (Thermo Fisher; 31000), 1% (w/v) nuclease-free BSA, in 1x PBST), at room temperature with rocking. After three five-minute washes in PBST, samples were incubated with 1-mL of biotin solution (10 µg/mL D-biotin (Thermo Fisher; B20656) in 1x PBST) to saturate neutravidin. Samples were then washed three times in PBST before proceeding with *in situ* biotinylation and quenching as described above. Following biotinylation, samples were incubated with 1 mL of 1 µg/mL neutravidin-Dy550 (ThermoFisher; 84606) in 1% BSA PBST for one hour at room temperature with rocking. Samples were then washed three times with PBST, stained with DAPI, and either imaged immediately or mounted in Vectashield (Vector Labs; H-1900-10) and stored at 4°C before use.

Combined pre-*TTN *O-MAP, RNA-FISH , and immunofluorescence

In combined O-MAP/RNA-FISH experiments (*e.g.*, **Fig. 1d**) primary probes were split into two sub-pools, equipped with orthogonal "landing pad" sequences that enable simultaneous recruitment of a fluorescent imaging oligo (for RNA-FISH) and HRP-conjugate (for O-MAP). These were introduced in a single primary hybridization step. Cells were processed as described in ***pre-*TTN *O-MAP day 1***, followed by endogenous biotin-blocking and *in situ* biotinylation, as described in ***pre-*TTN *O-MAP imaging***, above. After quenching biotinylation, samples were equilibrated 30% Formamide Wash buffer (1 mL, three five-minute washes, Room temperature, with rocking). This buffer was aspirated and each well was treating with 50 µL of FISH secondary probe mix, comprising 0.1 µM Alexa Fluor 647-conjugated secondary oligo (IDT; **Supplementary Table 1**) in 30% Formamide Hybridization Buffer. Samples were covered with a clean coverslip, incubated for one hour at 37°C in the dark, then washed three times with PBST. Before proceeding to immunofluorescence, samples were blocked with 5% (w/v) nuclease-free BSA in 1X PBST for 30 minutes. Samples were then incubated with rabbit-RBM20 (Thermo Fisher; PA5-57404, used at 1:800 dilution) in 5% BSA in 1X PBST for 1 hour 45 minutes. Samples were washed three times in 1X PBST (5 minutes per wash), and then incubated with Atto488-conjugated goat anti-rabbit (Thermo Fisher; R37116, used at 1:300 dilution) in 1% BSA in 1XPBST,

supplemented with 1 µg/mL neutravidin-Dy550. Samples were washed four times in 1 x PBST, stained with DAPI, and immaged immediately in 1 x PBST.

When performing pairwise fluorescent imaging (*e.g.*, combined O-MAP/RNA-FISH; O-MAP /immunofluorescence, RNA-FISH/immunofluorescence) the O-MAP protocol preceded the RNA-FISH or immunofluorescence protocol and RNA-FISH preceded the immunofluorescence protocol. Antibodies used: rabbit-TTN (Myomedix; Z1Z2, used at 1:250 dilution); rabbit-QKI (Sigma; HPA019123, used at 1:125 dilution); rabbit-SAFB (Sigma; HPA020076, used at 1:50 dilution); rabbit-H3K27Me3 (CST; 9733, used at 1:800 dilution); rabbit-H3K4Me3 (Active Motif; 39159, used at 1:500 dilution); rabbit-TrxR2 (Thermo Fisher; PA5-29458, used at 1:1000 dilution); and mouse-ATP5A (Abcam; ab14748, used at 1:500 dilution).

pre*-TTN *O-MAP-ChIP

O-MAP-ChIP was performed essentially as previously described[39], but with 24 sonication cycles during the cell-lysis/chromatin-shearing step. Briefly, O-MAP-labeled cells (3.6 x 10^6 per replicate; two biological replicates per experimental condition) were harvested in PBSTq using cell scrapers, pelleted at 3,000 x g for 10 minutes at 4°C, and media was aspirated. Cell pellets were flash-frozen in liquid nitrogen and stored at –80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1mL CLB (20 mM Tris pH 8.0, 85 mM KCl, 0.5% (v/v) NP-40) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail (Thermo Fisher; 87786) and 10 mM sodium azide, for 10 minutes. Lysates were clarified by centrifugation at 3,000 x g for five minutes at 4 °C, and supernatant was aspirated. Pellets were resuspended in fresh CLB, pelleted and aspirated twice more, for a total of three washes. Pellets were then lysed by gentle pipetting in 1 mL of NLB (10 mM Tris-HCl pH 7.5, 1% (v/v) NP-40, 0.5% (w/v) sodium Deoxycholate, 0.1% (w/v) SDS) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were then placed into an ice-cold thermal block and sheared with a Branson Digital Sonifier outfitted with a double stepped microtip, at 10-12 Watts over 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals. A total of 24 intervals were performed per sample, and after every 5 intervals the thermal block was placed on ice for 2 minutes to maintain temperature. This resulted in an average shearing size of approximately 200–400 bp, as gauged by agarose gel electrophoresis.

Sheared samples were clarified by centrifugation at 15,000 x g for 10 minutes at 4°C, and supernatants were moved to fresh tubes. 10 µL was removed from each sample and saved as "input," and to the remaining were added a 100 µL slurry of streptavidin-coated magnetic beads (Thermo Fisher; 88817), that had been pre-equilibrated by two washes in NLB. Samples rotated end-over-end for 2 hours at room-temperature. Beads were collected by magnetic separation, aspirated, and washed with the following series of buffers (1 mL each, 5-minutes per wash, at room-temperature, with end-over-end rotation): (1) NLB, supplemented with 5 mM EDTA, 10 mM sodium azide and protease inhibitors (1x Halt EDTA-free Protease Inhibitor Cocktail) and 150 mM NaCl; (2) NLB, supplemented with 5 mM EDTA, 10mM sodium azide and protease inhibitors, (3–4) two washes in High Salt Buffer (1 M KCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA), (5–6) two washes in Urea Buffer (2 M Urea, 10 mM Tris-HCl pH 7.5, 5 mM EDTA), (7) 10 mM Tris-HCl pH 7.5, 1% (w/v) SDS, (8–9) two washes in TE Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Washed beads were resuspended in 100 µL of Elution Buffer (2% (v/v) N-lauryl sarcosine (VWR; 100218-364), 10 mM EDTA, 5 mM DTT, in 1x PBS, supplemented with 200 µg proteinase K (Thermo Fisher; AM2548). Samples were then incubated at 65°C with shaking at 700 rpm for one hour in a Mixer HC, transferred to 0.2 mL tubes, and incubated overnight at 65°C in a thermocycler. Samples were transferred to 1.5 mL tubes and extracted by addition of an equal volume of phenol pH 6.6, followed by two equal volume extractions with absolute chloroform. Samples were supplemented with 1 µg GlycoBlue (Thermo Fisher; AM9516) and NaCl to 300 mM final, and precipitated with ethanol, overnight at –20°C. Samples were then pelleted by centrifugation at 15,000 x g for 30 minutes at 4 °C, washed twice with ice-cold 80% ethanol and resuspended in 20 µL nuclease-free water. Residual RNA was removed by supplementation of 10 µg RNase Cocktail Enzyme Mix (Thermo Fisher; AM2286) and incubation at 37 °C for 1 hour. Finally, DNA was purified by phenol extraction and ethanol precipitation as described above, and resuspended in 20 µL of nuclease-free water.

Pre*-TTN *O-MAP-ChIP library preparation and sequencing

DNA samples were processed for sequencing and sequenced as previously described[39]. Briefly, samples were quantified using a NanoDrop One and 300 ng of DNA from each sample was used for library preparation, using the NEBNext Ultra II DNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB; E7645 and E7335, respectively), according to the manufacturer's instructions. Two biological replicates were used per condition,

and each library was assigned a unique index. Library concentrations were quantified using the NEBNext Library Quantification Kit for Illumina (NEB; E7360), and independently validated using an Agilent 4200 TapeStation with a “DNA High Sensitivity” kit (Fred Hutch Genomics Core). Libraries were pooled in equimolar concentrations to 20 nM aggregate in nuclease-free water, with no more than eight libraries per pool. Each pool was subjected to 150 cycles of paired-end sequencing on an Illumina HiSeq 4000, run in high output mode (Azenta Life Sciences).

DNA SABER-FISH Probe Design

DNA SABER-FISH[45] probes (**Supplementary Table 3**) were designed using the online PaintSHOP application[82], with a target pool size of 1000 probes per locus. Briefly, segments of DNA from the center of each target locus (see, "*pre-TTN O-MAP-ChIP data analysis*," below) were uploaded to PaintSHOP as a .BED file. DNA Probe Design parameters were as follows: Repeat=allowed; Max Off Target Score=100; Max K-mer count=5; Min Prob Value=0; Optimize Set=show; Number of probes per target=1000, Enable=off. For the Append Sequences function the following parameters were used: Design Scheme=DNA Probe Design (5´ Outer Primer Sequence=Append; Orientation=Forward; Format=Same for all probes; Select Sequence Set=Custom; ATCCTAGCCCATACGGCAATG), (5´ Bridge Sequence=Append; Orientation=Forward, Format=Unique for each target; Select Sequence Set=Kishi et al. 2019 Bridges); (5´ Inner Primer Sequence=Append, Orientation=Forward; Format=Same for all probes; Select Sequence Set=Custom; TGAATAGCAGCGGTGGCAAAC); (3´ Appending Scheme=Primer/Bridge/Universal); (3´ Outer Primer Sequence=Append; Orientation=Reverse Complement; Format=Same for all probes; Select Sequence Set=Custom; GTATCGTGCAAGGGTGAATGC).

DNA-FISH Probe Amplification

Probes were purchased from Twist Bioscience as a single pool (Oligo pool, Tier 3L ssDNA, >6,001—12,000 oligos (151-200nt)), and Z resuspended in Tris-EDTA, to 6.0 ng/μL final. The probes were amplified by PCR, using Phusion HF (Thermo Fisher; F531S) and primers Forward 5´–ATCCTAGCCCATACGGCAATG–3´ and Reverse 5´–TAATACGACTCACTATAGGGGTATCGTGCAAGGGTGAATGC–3´, which adds a T7 promoter sequence. This reaction ran on a ProFlex PCR System (ThermoFisher) using the following parameters: Initial denaturation: 98°C, 5s; Denaturation: 98°C, 5s; Annealing: 58°C, 30s; Extension: 72°C, 15s; Final extension:

72°C, 5 min. Denaturation, annealing, and extension were performed for 10 cycles. DNA products were then purified using the Monarch PCR and DNA Cleanup Kit (NEB; T1030S), following the manufacturer's protocol, and eluted into a final volume of 30 µL. These purified products (10 µL per reaction) were then used as templates for *in vitro* RNA synthesis using the HiScribe T7 High Yield RNA Synthesis Kit (NEB; E2040S), in a final reaction volume of 30 µL, following the manufacturer's provided protocol. Reactions were performed at 37°C for 5–16 hours, in a thermocycler. Single-stranded DNA FISH probes were then synthesized by Reverse-Transcription using the Maxima H Minus RT enzyme (ThermoFisher; EP0751). RT reactions (70 µL, total) contained the entire 30 µL of unpurified T7 transcription reaction, supplemented with 1 mM of each dNTP, 1x Maxima H Minus RT Buffer, 14.28 µM RT primer, 4 µL RNaseOUT (Thermo Fisher; 10777019), and 1000 U Maxima H Minus RT enzyme. Reactions were incubated at 50°C for two hours in a thermocycler, and template RNA was removed by addition of 50 µL Alkaline Lysis Buffer (500 mM NaOH; 250 mM EDTA), followed by incubation for 12 minutes at 95°C, in a thermocycler. The final DNA probe pool was purified using the Oligo Clean and Concentrator Kit (Zymo Research; D4060) following the manufacturer's instructions, eluted into 20 µL nuclease-free water, at stored at –20°C until use.

Polymerase Exchange Reaction (PER)

To amplify the DNA SABER-FISH signal, 5´–Bridge oligos (IDT; **Supplementary Table 1**) were extended as 10 nt concatemers (enabling them to recruit more fluorescent imager oligos), using the Polymerase Exchange Reaction (PER), as described previously[39, 45]. To remove contaminating dGTP, a 40 µL reaction comprising 1x PBS; 12.5 mM $MgSO_4$; 375 µM each of dTTP, dATP, and dCTP; 125 nM each Clean.G oligo and the corresponding template hairpin (**Supplementary Table 1**); and 0.5 µL Bst LF polymerase (NEB; M0275S) were pre-incubated at 37°C for 15 minutes in a thermocycler. To these, 10 µL of the corresponding bridge oligo (4 µM, final) were added, and samples were incubated at 37°C for 2 hours, followed by 20 minutes at 80°C to inactive the polymerase. PER extension was confirmed by polyacrylamide gel electrophoresis (hand-poured 10% TBE-Urea, stained with 1x SYBER Gold, Thermo Fisher S11494). Extended probes were purified using the Oligo Clean and Concentrator Kit, according to manufacturer's instructions, and eluted in 10 uL nuclease-free water.

DNA SABER-FISH

DNA SABER-FISH probe pools were concentrated prior to use, by combining 1.2 μM of the purified ssDNA pool and 1.2 μM of the corresponding PER-extended bridge oligo, and speed-vacuuming to dryness (Savant; DNA Speed Vac DNA120). Dried oligo pellets were stored at –20°C prior to use. Pellets were then resuspended in a buffer comprising 50% (v/v) Formamide, 2X SSC, and 0.1% Tween-20, supplemented with 0.4 μg/mL RNase-A. These Hybridization Solution samples were placed on ice while samples were processed for DNA-FISH.

Cells were briefly washed in 1 mL 1X DPBS then fixed with 1 mL fixative (4% (v/v) PFA in 1X PBS), for 10 minutes at room temperature, without agitation. Fixative was aspirated and cells were washed with 1 mL DPBS for 1 minute, and then permeabilized with 1 mL 0.5% (v/v) Triton X-100 in 1X PBS for 10 minutes, at room temperature. Cells were then washed with 1mL 1X DPBS for two-minutes, and then incubated in 1 mL of 0.1 N HCl in 1X PBS for five minutes at room temperature. Cells were then washed twice with 2X SSCT (1 mL; 1 minute per wash; room temperature) and once with 50% Formamide Wash Buffer (1 mL; 2 minutes; room temperature). Samples were then aspirated, and incubated with 1 mL pre-warmed 50% Formamide Wash Buffer, and incubated at 60°C for 20 minutes. Wash buffer was aspirated and 50 μL of hybridization solution (*see above*) was added. Samples were denatured at 78°C for 4 minutes, resting on a pre-equilibrated thermal block in a water bath, sealed, and incubated at 37°C overnight.

The next day, samples were briefly rinsed twice with pre-warmed 60°C 2x SSCT, followed by four 5-minute washes with pre-warmed 60°C 2x SSCT, and two 5-minute washes with room-temperature 2x SSCT (1 mL per wash). Samples were then rinsed with 1X PBS, aspirated, and treated with 50 μL of Fluorophore solution (1 μM each AlexaFluor 647- and Atto 488- conjugated probes, in 1X PBS, **Supplementary Table 1**). Samples were protected from light and incubated for one-hour at 37°C. Samples were then rinsed with 1 mL 1X PBST, followed by two 5-minute washes in 1X PBST, and two 5-minute 1X PBS washes (1 mL each). Samples were imaged immediately, in 1X PBS.

FFPE Tissue Section Preparation for RNA-FISH and Immunofluorescence

Tissue sections were prepared using the FFPE Sample Preparation and Pretreatment protocol from the RNAscope 2.5 assay (ACD Bio), using glass Coplin jars. Briefly, to deparaffinize sections slides were placed in Xylenes (Sigma; 534056) for 5-minutes at RT, followed by a second 5-minute Xylene incubation, in a fresh jar. Sliders were then immediately placed in fresh 100% EtOH for 1 minute at RT with agitation, followed by another

1-minute incubation in 100% EtOH, in a fresh jar. Slides were then dried for 3 minutes on a Thermostat slide warmer (No. 26020) with its thermostat set to 4.5. Then, ~5–8 drops of RNAscope Hydrogen Peroxide solution were added to cover the entire section, and samples were incubated for 10 min at room temperature. Slides were then washed twice for 1 minute in nuclease-free water, followed by submerging in boiling RNAscope 1X Target Retreival Reagents solution for 10-minutes. Slides were immediately placed in room temperature nuclease-free water for 1 minute, followed by another 1-minute water wash, and a 1-minute EtOH wash. Slides were dried for 3 minutes on a Thermostat slide warmer, as above. An Immedge hydrophobic barrier pen (Vector labs; H-4000) was then used to create a hydrophobic barrier around each section, and samples were dried for 3-minutes at RT. Five drops of RNAscope Protease Plus was then added to entirely cover each section. Sections were then incubated in hybridization chambers at 40°C for 40 minutes, washed twice with room temperature nuclease-free water (1 min per wash), and twice for two-minutes with 150 µL 30% Formamide Wash Buffer. 150 µL of primary RNA-FISH probes were then added. Thereafter, combined RNA-FISH and immunofluorescence were performed as described above.

pre-TTN O-MAP-Seq

O-MAP labeled cells (4.8 x 10^6 cells; four biological replicates per experimental condition) were harvested by scraping in PBSTq. Samples were then pelleted by centrifugation at 3,000 x g for ten minutes at 4°C, aspirated, flash-frozen in liquid nitrogen, and stored at –80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1000 µL ice cold RIPA Buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (w/v) SDS, 0.5% (w/v) sodium Deoxycholate, 1% (v/v) Triton X-100, 5 mM EDTA, 0.5 mM DTT), supplemented with 1x EDTA-Free Halt Protease Inhibitor Cocktail, 0.1 U/µL RNase-OUT, and 10mM sodium azide, and rocked end-over-end at 4°C for five minutes. Cells were placed in an ice-cold thermal block and then sheared using a Branson Digital Sonifier 250 outfitted with a double stepped microtip, at 10–12 Watts for 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals, five intervals total. After every five intervals the thermal block was re-equilibrated on ice for two-minutes. Following sonication, lysates were clarified by centrifugation at 15,000 x g for ten minutes at 4°C. Supernatants were placed in fresh tubes and diluted with 1 mL Nuclear Lysis Buffer (NLB: 25 mM Tris-HCl pH 7.5, 150 mM KCl, 0.5% (v/v) NP-40, 5 mM EDTA, 0.5 mM DTT), supplemented with 1X Halt Protease Inhibitor Cocktail, 0.1 U/µL RNase-OUT and 10 mM sodium azide. From each sample 5% was removed

as “input,” and to the remainder were added 100 µL of a Pierce Streptavidin Magnetic Bead slurry that had been equilibrated by two washes in 1:1 RIPA:NLB, supplemented with 10mM sodium azide, 0.1 U/µL RNase-OUT, and 1X Halt Protease Inhibitor Cocktail. Samples were incubated for 2 hours at room temperature with end-over-end agitation. Beads were collected by magnetization, and then washed with the following series of buffers (1 mL each, 5 min per wash at room-temperature with end-over-end agitation). All buffers were supplemented with 1x EDTA-Free Halt protease inhibitor cocktail, 0.1 U/µL RNase-OUT, and 0.5 mM DTT, unless otherwise noted: (1) RIPA, supplemented with 10 mM sodium azide; (2) RIPA alone (3) High Salt Buffer (1 M KCl, 50 mM Tris-HCl pH 7.5, 5 mM EDTA) (4) Urea Buffer (2 M Urea, 50mM Tris-HCl pH 7.5, 5 mM EDTA) (5) 1:1 RIPA:NLB, without protease inhibitors (6) NLB, without protease inhibitors, (7–8) two washes in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA), without protease inhibitors.

RNA was isolated from both input and O-MAP-enriched samples by proteolysis in 100 µL Elution Buffer (2% (v/v) N-lauryl sarcoside, 10 mM EDTA, 5 mM DTT, 200 µg proteinase K, in 1xPBS). Reactions were shaken at 700 rpm in a Mixer HC (USA Scientific) for 1 hour at 42°C, followed by 1 hour at 60°C. RNA was then extracted once with 1 volume of phenol pH 4.3, and twice thereafter with an equal volume of absolute chloroform. Reactions were supplemented with 1µL Glycoblue and NaCl to 300 mM, and ethanol precipitated at –20°C overnight. RNA was harvested by centrifugation at 15,000 x g for 30 minutes at 4°C, washed twice with 70% ethanol, and resuspended in 40 µL nuclease-free water. To each sample, 5 µL DNA Digestion Buffer and 5 µL DNase I (Zymo Research; R1013) was added, gently mixed by inversion, and then incubated for 15-minutes at room temperature. RNA was then purified using an RNA Clean and Concentrator-5 Kit (Zymo Research; R1013), according to manufacturer’s instructions. RNA was eluted in 15 µL nuclease-free water, and its concentration was measured using a Nanodrop One (Thermo Fisher).

pre-TTN O-MAP-Seq library preparation and sequencing

RNA samples were processed for sequencing as described previously[39] Briefly, ribosomal RNA was first depleted by RNase-H digestion, using a pool of DNA oligonucleotides antisense to rRNA, as previously described[39, 83]. 1 µg of the antisense probe pool was added to 1 µg of RNA in 200 mM NaCl, 100 mM Tris-HCl, pH 7.4, at a final volume of 10 µL. Samples were heated at 95°C for 2 minutes and then slowly cooled to 45°C at a rate of –0.1°C/s, using a ProFlex PCR system. Then, reactions were supplemented with 10 µL of RNase H mix (10 U

Hybridase Thermostable RNase H (Lucigen; H39500), 20 mM MgCl2) that was pre-heated to 45°C. Samples were incubated at 45°C for 30 minutes then placed on ice. RNA was purified by acidic phenol-chloroform extraction and ethanol precipitation. Samples were resuspended in 40 µL of nuclease-free water, and 5 µL of DNA Digestion Buffer and 1 µL of DNase I (Zymo Research; R1013) were added. Samples were mixed by gentle inversion and then incubated at room temperature for 15 minutes. RNA was purified again by the RNA Clean and Concentrator-5 Kit, eluting into 10 µL nuclease-free water. Samples were quantified on a Nanodrop One. Sequencing libraries were synthesized from 300 ng RNA using the NEBNext Ultra II Directional RNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB; E7760 and E7735), according to manufacturer's instructions. Four biological replicates were used for each condition and each library was given a unique index. Libraries were quantified using the NEBNext Library Quantification Kit for Illumina, following manufacturer's instructions, and library quality was confirmed using an Agilent 4200 TapeStation with a "DNA High Sensitivity" kit (Fred Hutch Genomics Core). Libraries were pooled in equimolar concentrations to 20 nM aggregate concentration in nuclease-free water, with no more than 8 libraries per pool. The pool underwent 150 cycles of paired-end sequencing, on one lane of an Illumina HiSeq 4000 per pool, run in high output mode (Azenta Life Sciences).

pre-TTN O-MAP-ChIP data analysis

Deep sequencing reads were trimmed using TrimGalore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with parameters -q 30 --phred33, and mapped to the GRCH38 genome using Bowtie2 version 2.4.47 (Ref.[84]). Duplicate reads were removed with the Picard MarkDuplicates function (http://broadinstitute.github.io/picard). Coverage bigWig files were generated using deepTools version 3.5.19 (Ref. [85]) with a binsize of 1 and normalizing with RPKM. Log_2(Fold Change) bigWig files were generated by comparing O-MAP ChIP data to matched input samples. Regions of enrichment were called using Enriched Domain Detector (EDD)[86], using default settings and blacklisting regions from the ENCODE Blacklist[87]. Epigenomic analysis of enriched DNA loci was performed using ChromHMM version 1.22 (Ref. [46]). The OverlapEnrichment function was called using a chromatin state file for RUES2 cardiomyocytes[88] (Accelerating Medicines in Partnerships; accession DSR313BZF). Fold enrichment of each epigenomic signature for all DNA loci was plotted as a heatmap using seaborn version 0.10.11 (Ref. [89]). Transcription factor data were pulled from

the Human Transcription Factor database version 1.011 (Ref. [90]), and those encoded in pre-*TTN* O-MAP TIDs underwent gene ontology using Metascape[51].

Expression analysis for TID-enclosed and FOXL1 target genes (**Figs. 3b,d,** and **e**) used O-MAP-Seq input samples (*see below*) to measure the global transcriptome. Data were processed as follows. Demultiplexed FASTQ files were aligned to the reference GRCh38 genome using STAR 2.7.11b (Ref. [91]), in GeneCounts mode. Output files (ReadsPerGene.out.tab) were imported as data frames in R 4.3.3, and a unique counts table for each sample was created and normalized as TPM (Transcripts Per Million) using the NormalizeTPM function (ADImpute package[92]). The gene length file used for the normalization was obtained with gtftools in –l mode with the same GTF file used for the genome indexing. The column "mean" was then selected for the normalization. Genes of interest (TIDs enclosed genes; FOXL1 regulated genes) were selected using the dplyr package, and for each group of a melt table of filtered counts was created using reshape2 (Ref. [93]). Genes were grouped by biological condition (WT and RBM20 KO) and the median was calculated with the median function (stats package). A table with median values per gene and condition were plotted as density plots with the geom_density function in ggplot2 (Ref. [94]). Statistical significance was calculated with a Kolmogorov-Smirnov (KS) test, using the ks.test function (statspackage) in both "greater" and "less" modes.

Data analysis

Raw RNA-seq FASTA files were aligned to GRCH38 using HISAT2 version 2.2.11 (Ref. [95]), in paired-end setting with default parameters. The resulting SAM files were converted to BAM format and sorted using Samtools version 1.15.1 (Ref. [96]). Bigwig files for visualizing strand specific information were created using deepTools version 3.5.1 (Ref. [85]) with parameters: --filterRNAstrand forward/reverse --binSize 1 --normalizeusingBPM. Mapped reads were quantified using StringTie version 2.2.1 (Ref. [97]) and the StringTie output was prepared for differential expression analysis using the prepDE.py function. The resulting count matrices were used for differential expression analysis using DESeq2 (Ref. [98]) with an FDR cutoff of 0.05. For differential analysis between WT and ΔRBM20 phenotypes, the DESeq2 likelihood ratio test was used to calculate (streptavidin-enriched transcripts for WT / streptavidin-enriched transcripts for ΔRBM20) / (whole-cell transcriptome for WT / whole-cell transcriptome for ΔRBM20). For computing differentially enriched transcripts between WT and ΔRBM20 cells, DESeq2 was performed for WT whole-cell transcriptome compared to ΔRBM20 whole-cell

transcriptome using standard settings. The rMATS package[99] was used for alternative splicing analysis, utilizing the BAM files generated by HISAT2 with the parameters: -t paired –readLength 150. Low coverage splicing events were filtered to include only transcripts with an average of 5 or more reads between replicates of each group. Differential events were further filtered for an FDR of 0.05 and a change in PSI between the two groups > 10% (δ PSI > 0.1). Sashimi plots for each rMATS event type (skipped exon, retained intron, mutually exclusive exons, alternative 5' splice site, and alternative 3' splice site) were generated using rmats2sashimiplot (https://github.com/Xinglab/rmats2sashimiplot) using the BAM files generated by HISAT2 for coordinate location and the rMATS events file for splicing event information. Transite[71] was used for differential RNA-binding protein profiling of transcripts enriched at TTN foci in WT versus ΔRBM20 cells using the Transcript Set Motif Analysis with the parameters: --matrix-based --Homo sapiens --svg --Transite motif database --maximum binding sites 5. The oRNAment database[100] was used to identify transcripts with QKI binding motifs. Volcano plots were generated using EnhancedVolcano version 1.12.0 (https://github.com/kevinblighe/EnhancedVolcano). All statistical analyses (Fisher's exact test, hypergeometric distribution test, or Student's t-test, where appropriate) were performed in R or in python using the ggplot2 (Ref. [94]), seaborn[89], or matplotlib[101] modules. Gene ontology analysis was performed using Metascape[51], Enrichr [102], and GSEA[103].

Quantitative RT–PCR

Quantitative RT–PCR (qRT–PCR) was performed as previously described[104]. Growth media was aspirated, and RNA was extracted with TRIZol reagent (Invitrogen; 15596026; 1 mL per well of well of a 12-well plate), according to the manufacturer's protocol. Contaminating DNA was removed with DNase I and RNA was purified using RNA Clean and Concentrator-5 Kit as described above (*see, "pre*-TTN *O-MAP-seq"*), eluting into 30 µL nuclease-free water. RNA samples were reverse transcribed using SuperScript IV Reverse Transcriptase (Thermo Fisher; 18090010), priming with random hexamers according to the manufacturer's protocol. cDNA was diluted with nuclease-free water, mixed with gene specific or isoform specific primers, supplemented with Rox-normalized PowerUp SYBR Green Master Mix (Thermo Fisher; A25777), and aliquoted into 384-well plates, 10 µL per well. qPCR was performed on a QuantStudio 5 Real-Time PCR System (Thermo Fisher; A34322) in both biological and technical quadruplicate. Gene expression was calculated using the 2-ΔΔCt method using the housekeeping gene RPLP0. Splicing isoform gene expression was calculated using the 2-ΔCt method as previously described2;

2^-(Ct spliced isoform – Ct constitutive isoform)[105]. Primer sequences for qPCR are listed in (**Supplementary Table 5**)

pre*-TTN *O-MAP-MS sample preparation

O-MAP labeled cells ($2.4x10^6$ per replicate; five biological replicates per experimental condition) were harvested in PBSTq without sodium ascorbate using cell scrapers, and pelleted at 3,000 x g for 10 minutes at 4°C. PBSTq was aspirated and cells were flash-frozen in liquid nitrogen and stored at –80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1mL CLB (20 mM Tris pH 8.0, 85 mM KCl, 0.5% (v/v) NP-40) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were clarified by centrifugation at 3,000 x g for five minutes at 4 °C followed by aspiration of supernatant. Samples were resuspended with CLB, pelleted and aspirated twice more. Pellets were then lysed by gentle pipetting in 1 mL of NLB (10 mM Tris-HCl pH 7.5, 1% (v/v) NP-40, 0.5% (w/v) sodium Deoxycholate, 0.1% (w/v) SDS) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were then placed into an ice-cold thermal block and sheared with a Branson Digital Sonifier outfitted with a double stepped microtip, at 10–12 Watts over 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals. A total of 4 intervals were performed per sample; in between samples the thermal block was re-equilibrated on ice.

Samples were supplemented with 47.47 µL of 20% (w/v) SDS, to a final concentration of 1%, and boiled at 95°C for one hour to reverse formaldehyde crosslinks. Samples were passively cooled to 50 °C before placing on ice, and then sonicated for one more cycle, using the same setting described above. Samples were then clarified by centrifugation at 15,000 x g for 10-minutes at 4°C, and supernatants were moved to fresh tubes. To each tube was added 75 µL of Pierce Streptavidin-coated magnetic beads that had been equilibrated by two NLB washes. Samples were rotated end-over-end for 2 hours at room-temperature, and beads were collected by magnetization and aspirated. Beads were then washed with 250 µl of NLB, supplemented with 5 mM EDTA, 10 mM sodium azide, 1x Halt EDTA-free Protease Inhibitor Cocktail, and 150 mM NaCl. Beads were then moved to fresh PCR-strip tubes (Simport Scientific; T320-2N) to minimize sample loss. The following series of washes were then performed (200 µl each, 1 minute per wash, at room-temperature, with end-over-end rotation): (1–4) four washes of NLB, supplemented with 5 mM EDTA, 10mM sodium azide and protease inhibitors; (5–8) four

washes of 1X PBS in nuclease-free water; (9–16) four washes of 1 M KCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA; (17–24) eight washes of 2 M Urea, 10 mM Tris-HCl pH 7.5, 5 mM EDTA; (25–28) four washes of 200 mM EPPS pH 8.5 in nuclease-free water. After the final wash the beads were moved to fresh PCR strip tubes, aspirated, and resuspended in 50 µl 200 mM EPPS pH 8.5.

Samples were transferred to a Mixer HC unit (USA Scientific) outfitted with custom "thermal couples" fashioned by wrapping a small square of aluminum foil around the end of a 1 mL pipette tip. Samples were supplemented with TCEP (Thermo Fisher; 77720) to a final concentration of 10 mM, and incubated at room temperature for 1 hour, with shaking at 700 rpm. Samples were alkylated with iodoacetamide (Sigma; I1149-5G) supplemented to a final concentration of 20 mM, and incubated in the dark for one-hour at room temperature with shaking at 700 rpm. Alkylation was quenched by the addition of dithiothreitol (Sigma; 43816) to 5 mM and incubation at room temperature for 15 minutes, with shaking at 700 rpm. To each sample was added 300 ng of mass spectrometry grade LysC (FujiFilm; 121-05063), and samples were incubated at 37°C for three hours, with shaking at 700 rpm. Samples were then supplemented with 300 ng sequencing grade modified trypsin (Promega; V5113) and incubated for exactly 16 hours at 37°C, with shaking at 700 rpm.

Beads were removed by magnetic separation, and the clarified samples were moved to fresh PCR strip tubes and frozen at –80 °C to inactivate LysC and trypsin. Samples were stored at –80°C for no longer than three days. Samples were thawed to room temperature, and to each were added 0.5 µL of 1% HPLC-grade formic acid to achieve pH < 3, as measured using pH strip paper. Samples were then desalted using C18 StageTips, largely as described previously[39]. Briefly, each StageTip was activated with 50 µL of HPLC-grade methanol, followed by addition of 50 µL of Stage Tip B (80% aq. acetonitrile; 0.1% trifluoroacetic acid) and then 50 µL of Stage Tip A (0.1% TFA). Samples were then loaded, washed with 50 µL of Stage Tip A, and stored at 4°C until use.

O-MAP-MS data acquisition and analysis

Peptides were eluted from StageTips by addition of 50 µL 45% acetonitrile, 0.1% TFA, into a 96-well plate. Samples were dried and resuspended in 10 µL Stage Tip A. Peptides were separated on an EASY-nLC 1200 System (Thermo Fisher Scientific) using 20 cm long fused silica capillary columns (100 µm ID, laser pulled in-house with Sutter P-2000, Novato CA) packed with 3 µm 120 Å reversed phase C18 beads (Dr. Maisch,

Ammerbuch, DE). The liquid chromatography (LC) gradient was 90 min long with 5–35% B at 300 nL/min. LC solvent A was 0.5% (v/v) aq. acetic acid and LC solvent B was 0.5% (v/v) acetic acid, 80% acetonitrile. MS data were collected on a Thermo Fisher Scientific Orbitrap Fusion Lumos using a data-independent acquisition (DIA) method with a 120K resolution Orbitrap MS1 scan and 12 m/z isolation window, 15K resolution Orbitrap MS2 scans for precursors from 400-1000 m/z.

Data .raw files were converted to. mzML using MSConvert 3.0.21251-d2724a5 and spectral libraries built using MSFragger-DIA[106] (FragPipe version 19.1) with quantification through DIA-NN version 1.8.232. The database search was against the UniProt human database (updated September 17th, 2021) containing 20420 sequences and 20420 reverse-sequence decoys. For the MSFragger analysis, both precursor and initial fragment mass tolerances were set to 20 ppm. Spectrum deisotoping, mass calibration, and parameter optimization were enabled. Enzyme specificity was set to “strictrypsin” and up to two missed trypsin cleavages were allowed. Oxidation of methionine, acetylation of protein N-termini, −18.0106 Da on N-terminal Glutamic acid, and −17.0265 Da on N-terminal Glutamine and Cysteine were set as variable modifications. Carbamidomethylation of Cysteine was set as a fixed modification. Maximum number of variable modifications per peptide was set to 3.

FragPipe/DIA-NN output files were processed and analyzed using the Perseus software package v1.5.6.0. Expression columns (protein MS intensities) were log2 transformed and normalized by subtracting the median log2 expression value from each expression value within each MS run. For statistical testing of significant differences in expression, a two-sample Student’s t test was applied.

Image acquisition

Fluorescence widefield microscopy was performed on a Leica DM IL, equipped with a HC Fluotar 100x oil immersion objective with a 1.32 numerical aperture and planar correction (Leica; 11506527), a white LED light source (Leica; EL6000) and a DFC365 FX digital camera (Leica; 11547004). The following filter cubes were used: Texas Red (Leica TX2 ET; 11504180; used with Dylight-550 conjugates), Cy5 (Leica Y5 ET; 11504181, used for Alexafluor-647), GFP (Leica GFP ET; 11504174, used for Atto-488 and AF488), and DAPI (Leica DAPI ET; 11504204). Illumination intensity was adjusted using the light source manual control; acquisition times ranged from 40–1000 ms, as controlled by the Leica LASX software. Fluorescence confocal microscopy was performed

on a Leica SP8X microscope (UW Keck Imaging Center), outfitted with a HC CS2 63x oil immersion objective, with 1.40 numerical aperture with both planar and apochromatic correction. The pixel size was 0.18 µm for fixed cell distance quantification, 0.06 µm for fixed cell representative images, and 0.045 for FFPE distance quantification and representative images. Samples were illuminated using a 470–670 nm tunable White Light Laser system, with a typical laser power of 0.1% for DAPI, 3% for 550 nm, and 10% for 647 nm. Gain and offset settings were adjusted to avoid pixel saturation. Images were line-averaged twice, with an average pixel dwell time of 1.58 µs. A bit-depth of 16 was used and zoom factor of 1-3 was used for all images.

Image processing

Images were processed using Fiji[107] and ImageJ[108] and multicolor overlays were made using the screen setting in Adobe Photoshop. All confocal images are maximum projections of z-stacks. Brightness and contrast were adjusted for display purposes using Fiji and ImageJ or Adobe Photoshop. In all cases, contrast adjustment was applied to improve signal visibility, by changing the minimum (black) and maximum (white) values only. Automated despeckling was applied when necessary (e.g. in DNA-FISH images with weak, diffuse speckling between cells) to reduce residual background signal. To measure the 3D distances between DNA-FISH puncta for the TTN locus and DNA-FISH puncta for an experimental TID, nuclei were first segmented in Fiji. Each nucleus was then duplicated in 3D retaining all Z-slices for DAPI and for FISH signal. Then, for each 3D composite image, the minimum edge-edge distances between *TTN* FISH loci and TID FISH loci were measured using the object-based ImageJ Distance Analysis (DiAna) plug-in[109], using the Gaussian-filtering setting to remove noise. At least 50 cells per condition were analyzed; the exact number is listed in each figure. The same pipeline was used to measure the distance between nascent *TTN* O-MAP signal and interacting RNA-FISH signal. DiAna was also used for object-based colocalization, and the shuffle function was used to calculate the statistical significance of object-based colocalization and to generate the cumulative distribution plots.

REFERENCES

1. Bhat, P., Honson, D. & Guttman, M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat Rev Mol Cell Biol* **22**, 653-670 (2021).
2. Mirny, L. & Dekker, J. Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb Perspect Biol* **14** (2022).
3. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).

4. Chen, Y. & Belmont, A.S. Genome organization around nuclear speckles. *Curr Opin Genet Dev* **55**, 91-99 (2019).

5. Bhat, P. *et al.* Genome organization around nuclear speckles drives mRNA splicing efficiency. *Nature* **629**, 1165-1173 (2024).

6. Belmont, A.S. Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. *Cold Spring Harb Perspect Biol* **14** (2022).

7. Hirose, T., Ninomiya, K., Nakagawa, S. & Yamazaki, T. A guide to membraneless organelles and their various roles in gene regulation. *Nat Rev Mol Cell Biol* **24**, 288-304 (2023).

8. Chakalova, L. & Fraser, P. Organization of transcription. *Cold Spring Harb Perspect Biol* **2**, a000729 (2010).

9. Bertero, A. RNA Biogenesis Instructs Functional Inter-Chromosomal Genome Architecture. *Front Genet* **12**, 645863 (2021).

10. Quinodoz, S.A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757 e724 (2018).

11. Sabari, B.R., Dall'Agnese, A. & Young, R.A. Biomolecular Condensates in the Nucleus. *Trends Biochem Sci* **45**, 961-977 (2020).

12. Spector, D.L. & Lamond, A.I. Nuclear speckles. *Cold Spring Harb Perspect Biol* **3** (2011).

13. Bertero, A. *et al.* Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. *Nat Commun* **10**, 1538 (2019).

14. Guo, W. *et al.* RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med* **18**, 766-773 (2012).

15. Khan, M.A. *et al.* RBM20 Regulates Circular RNA Production From the Titin Gene. *Circ Res* **119**, 996-1003 (2016).

16. Li, S., Guo, W., Dewey, C.N. & Greaser, M.L. Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Res* **41**, 2659-2672 (2013).

17. Shevtsov, S.P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nat Cell Biol* **13**, 167-173 (2011).

18. Smith, K.P., Hall, L.L. & Lawrence, J.B. Nuclear hubs built on RNAs and clustered organization of the genome. *Curr Opin Cell Biol* **64**, 67-76 (2020).

19. Loda, A., Collombet, S. & Heard, E. Gene regulation in time and space during X-chromosome inactivation. *Nat Rev Mol Cell Biol* **23**, 231-249 (2022).

20. Duronio, R.J. & Marzluff, W.F. Coordinating cell cycle-regulated histone gene expression through assembly and function of the Histone Locus Body. *RNA Biol* **14**, 726-738 (2017).

21. Pourmorady, A.D. *et al.* RNA-mediated symmetry breaking enables singular olfactory receptor choice. *Nature* **625**, 181-188 (2024).

22. Fenix, A.M. *et al.* Gain-of-function cardiomyopathic mutations in RBM20 rewire splicing regulation and re-distribute ribonucleoprotein granules within processing bodies. *Nat Commun* **12**, 6324 (2021).

23. Kayvanpour, E. *et al.* Genotype-phenotype associations in dilated cardiomyopathy: meta-analysis on more than 8000 individuals. *Clin Res Cardiol* **106**, 127-139 (2017).

24. Refaat, M.M. *et al.* Genetic variation in the alternative splicing regulator RBM20 is associated with dilated cardiomyopathy. *Heart Rhythm* **9**, 390-396 (2012).

25. van den Hoogenhof, M.M.G. *et al.* RBM20 Mutations Induce an Arrhythmogenic Dilated Cardiomyopathy Related to Disturbed Calcium Handling. *Circulation* **138**, 1330-1342 (2018).

26. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362** (2018).

27. Dekker, J. *et al.* Spatial and temporal organization of the genome: Current state and future aims of the 4D nucleome project. *Mol Cell* **83**, 2624-2640 (2023).

28. Quinodoz, S.A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* **184**, 5775-5790 e5730 (2021).

29. Hristov, B.H., Noble, W.S. & Bertero, A. Systematic identification of interchromosomal interaction networks supports the existence of specialized RNA factories. *Genome Res* (2024).

30. Beagrie, R.A. *et al.* Multiplex-GAM: genome-wide identification of chromatin contacts yields insights overlooked by Hi-C. *Nat Methods* **20**, 1037-1047 (2023).

31. Dekker, J. *et al.* An integrated view of the structure and function of the human 4D nucleome. *bioRxiv*, 2024.2009.2017.613111 (2024).

32. Frottin, F. *et al.* The nucleolus functions as a phase-separated protein quality control compartment. *Science* **365**, 342-347 (2019).

33. Machyna, M. & Simon, M.D. Catching RNAs on chromatin using hybridization capture methods. *Brief Funct Genomics* **17**, 96-103 (2018).

34. Dejardin, J. & Kingston, R.E. Purification of proteins associated with specific genomic Loci. *Cell* **136**, 175-186 (2009).

35. Qin, W., Cho, K.F., Cavanagh, P.E. & Ting, A.Y. Deciphering molecular interactions by proximity labeling. *Nat Methods* **18**, 133-143 (2021).

36. Gao, X.D. *et al.* C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat Methods* **15**, 433-436 (2018).

37. Myers, S.A. *et al.* Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat Methods* **15**, 437-439 (2018).

38. Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J Cell Biol* **217**, 4025-4048 (2018).

39. Tsue, A.F. *et al.* Multiomic characterization of RNA microenvironments by oligonucleotide-mediated proximity-interactome mapping. *Nat Methods* (2024).

40. Beliveau, B.J. *et al.* OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc Natl Acad Sci U S A* **115**, E2183-E2192 (2018).

41. Raj, A. & Rinn, J.L. Illuminating Genomic Dark Matter with RNA Imaging. *Cold Spring Harb Perspect Biol* **11** (2019).

42. Oakley, J.V. *et al.* Radius measurement via super-resolution microscopy enables the development of a variable radii proximity labeling platform. *Proc Natl Acad Sci U S A* **119**, e2203027119 (2022).

43. Koelemen, J., Gotthardt, M., Steinmetz, L.M. & Meder, B. RBM20-Related Cardiomyopathy: Current Understanding and Future Options. *J Clin Med* **10** (2021).

44. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb Perspect Biol* **2**, a003889 (2010).

45. Kishi, J.Y. *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat Methods* **16**, 533-544 (2019).

46. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216 (2012).

47. Paige, S.L. *et al.* A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221-232 (2012).

48. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* **48**, D882-D889 (2020).

49. Oji, A., Choubani, L., Miura, H. & Hiratani, I. Structure and dynamics of nuclear A/B compartments and subcompartments. *Curr Opin Cell Biol* **90**, 102406 (2024).

50. Sha, K. & Boyer, L.A. The chromatin signature of pluripotent cells, in *StemBook* (Cambridge (MA); 2008).

51. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).

52. Wyles, S.P. *et al.* Modeling structural and functional deficiencies of RBM20 familial dilated cardiomyopathy using human induced pluripotent stem cells. *Hum Mol Genet* **25**, 254-265 (2016).

53. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174-D182 (2024).

54. Attar, S. *et al.* Programmable peroxidase-assisted signal amplification enables flexible detection of nucleic acid targets in cellular and histopathological specimens. *bioRxiv* (2023).

55. Barutcu, A.R. *et al.* Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol Cell* **82**, 1035-1052 e1039 (2022).

56. Liao, S.E. & Regev, O. Splicing at the phase-separated nuclear speckle interface: a model. *Nucleic Acids Res* **49**, 636-645 (2021).

57. Chen, L.L. & Carmichael, G.G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell* **35**, 467-478 (2009).

58. Dauksaite, V. & Gotthardt, M. Molecular basis of titin exon exclusion by RBM20 and the novel titin splice regulator PTB4. *Nucleic Acids Res* **46**, 5227-5238 (2018).

59. Luk, A., Ahn, E., Soor, G.S. & Butany, J. Dilated cardiomyopathy: a review. *J Clin Pathol* **62**, 219-225 (2009).

60. Olson, T.M., Michels, V.V., Thibodeau, S.N., Tai, Y.S. & Keating, M.T. Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. *Science* **280**, 750-752 (1998).

61. Olson, T.M. *et al.* Metavinculin mutations alter actin interaction in dilated cardiomyopathy. *Circulation* **105**, 431-437 (2002).

62. Rittenhouse, N.L. & Dowen, J.M. Cohesin regulation and roles in chromosome structure and function. *Curr Opin Genet Dev* **85**, 102159 (2024).

63. Bianchi, M.E. & Agresti, A. HMG proteins: dynamic players in gene regulation and differentiation. *Curr Opin Genet Dev* **15**, 496-506 (2005).

64. Huo, X. *et al.* The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation. *Mol Cell* **77**, 368-383 e367 (2020).

65. Dequeker, B.J.H. *et al.* MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* **606**, 197-203 (2022).

66. Carlson, S.M. *et al.* RBM25 is a global splicing factor promoting inclusion of alternatively spliced exons and is itself regulated by lysine mono-methylation. *J Biol Chem* **292**, 13381-13390 (2017).

67. Norman, M., Rivers, C., Lee, Y.B., Idris, J. & Uney, J. The increasing diversity of functions attributed to the SAFB family of RNA-/DNA-binding proteins. *Biochem J* **473**, 4271-4288 (2016).

68. Darbelli, L. & Richard, S. Emerging functions of the Quaking RNA-binding proteins and link to human diseases. *Wiley Interdiscip Rev RNA* **7**, 399-412 (2016).

69. Chen, X. *et al.* The Emerging Roles of the RNA Binding Protein QKI in Cardiovascular Development and Function. *Front Cell Dev Biol* **9**, 668659 (2021).

70. Tijsen, A.J. *et al.* Titin Circular RNAs Create a Back-Splice Motif Essential for SRSF10 Splicing. *Circulation* **143**, 1502-1512 (2021).

71. Krismer, K. *et al.* Transite: A Computational Motif-Based Analysis Platform That Identifies RNA-Binding Proteins Modulating Changes in Gene Expression. *Cell Rep* **32**, 108064 (2020).

72. Maatz, H. *et al.* RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J Clin Invest* **124**, 3419-3430 (2014).

73. Belotserkovskii, B.P., Tornaletti, S., D'Souza, A.D. & Hanawalt, P.C. R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair (Amst)* **71**, 69-81 (2018).

74. Arnould, C. *et al.* Chromatin compartmentalization regulates the response to DNA damage. *Nature* **623**, 183-192 (2023).

75. Prasanth, K.V. *et al.* Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249-263 (2005).

76. Ninomiya, K. *et al.* LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation. *EMBO J* **39**, e102729 (2020).

77. Caudron-Herger, M. *et al.* Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J* **34**, 2758-2774 (2015).

78. Fazal, F.M. *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473-490 e426 (2019).

79. Gregorich, Z.R., Zhang, Y., Kamp, T.J., Granzier, H.L. & Guo, W. Mechanisms of RBM20 Cardiomyopathy: Insights From Model Systems. *Circ Genom Precis Med* **17**, e004355 (2024).

80. Braz, S.O., Acquaire, J., Gourdon, G. & Gomes-Pereira, M. Of Mice and Men: Advances in the Understanding of Neuromuscular Aspects of Myotonic Dystrophy. *Front Neurol* **9**, 519 (2018).

81. Andersen, J.S. *et al.* Nucleolar proteome dynamics. *Nature* **433**, 77-83 (2005).

82. Hershberg, E.A. *et al.* PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. *Nat Methods* **18**, 937-944 (2021).

83. Adiconis, X. *et al.* Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* **10**, 623-629 (2013).

84. Langdon, W.B. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min* **8**, 1 (2015).

85. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).

86. Lund, E., Oldenburg, A.R. & Collas, P. Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Res* **42**, e92 (2014).

87. Amemiya, H.M., Kundaje, A. & Boyle, A.P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019).

88. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300-307 (2021).

89. Waskom, M.L. seaborn: statistical data visualization. *Journal of Open Source Software* **6** (2021).

90. Lambert, S.A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665 (2018).

91. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

92. Xu, L., Xu, Y., Xue, T., Zhang, X. & Li, J. AdImpute: An Imputation Method for Single-Cell RNA-Seq Data Based on Semi-Supervised Autoencoders. *Front Genet* **12**, 739677 (2021).

93. Wickham, H. Reshaping Data with the reshape Package. *Journal of Statistical Software* **21**, 1 - 20 (2007).

94. Wickham, H. in Use R!,, Edn. 2nd 1 online resource (XVI, 260 pages 232 illustrations, 140 illustrations in color (Springer International Publishing : Imprint: Springer,, Cham; 2016).

95. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915 (2019).

96. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).

97. Shumate, A., Wong, B., Pertea, G. & Pertea, M. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS Comput Biol* **18**, e1009730 (2022).

98. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

99. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* **111**, E5593-5601 (2014).

100. Benoit Bouvrette, L.P., Bovaird, S., Blanchette, M. & Lecuyer, E. oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. *Nucleic Acids Res* **48**, D166-D173 (2020).

101. Hunter, J.D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**, 90-95 (2007).

102. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

103. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

104. Shechner, D.M. Targeting Noncoding RNA Domains to Genomic Loci with CRISPR-Display: Guidelines for Designing, Building, and Testing sgRNA-ncRNA Expression Constructs. *Methods Mol Biol* **2162**, 115-152 (2021).

105. Kirschneck, C. *et al.* Valid gene expression normalization by RT-qPCR in studies on hPDL fibroblasts with focus on orthodontic tooth movement and periodontitis. *Sci Rep* **7**, 14751 (2017).

106. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D. & Nesvizhskii, A.I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **14**, 513-520 (2017).

107. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).

108. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675 (2012).

109. Gilles, J.F., Dos Santos, M., Boudier, T., Bolte, S. & Heck, N. DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods* **115**, 55-64 (2017).