Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory

Bertero, Fields et al.

Supplementary Information

Guide to content:

Supplementary Note 1

Supplementary Figures and Figure legends

Supplementary Figure 1. Validation of hESCs and cardiac differentiation.

Supplementary Figure 2. Genome-wide contact maps across differentiation.

Supplementary Figure 3. Intra-chromosomal interactions vary over differentiation by compartment.

Supplementary Figure 4. Expression changes are related to A/B compartment changes.

Supplementary Figure 5. hESCs and hiPSCs show similar changes in chromatin organization during cardiac differentiation.

Supplementary Figure 6. TADs are dynamically regulated across differentiation.

Supplementary Figure 7. Local accessibility is dynamic across differentiation and reflects cell type specific transcription factors.

Supplementary Figure 8. Cardiac specific genes are enriched in B to A regions.

Supplementary Figure 9. Proximity between *TTN* and RBM20 target genes depends on *TTN* transcription and RBM20 expression.

Supplementary Figure 10. Generation of gene edited hESC-CM to investigate the mechanisms regulating *TTN trans* interactions.

Supplementary Figure 11. Uncropped gel images.

Supplementary Tables

Supplementary Table 1. Hi-C sequencing statistics.

Supplementary Table 2. HiC-Rep scores between samples.

Supplementary Table 3. RNA-seq sequencing statistics.

Supplementary Table 4. Biological process gene ontology terms associated with differentially expressed genes peaking in cardiomyocyte stage.

Supplementary Table 5. Biological process gene ontology terms associated with differentially expressed genes peaking in cardiomyocytes in regions of the genome that transition B to A.

Supplementary Table 6. Jaccard index scores for TAD boundary overlap**.**

Supplementary Table 7. ATAC-seq sequencing statistics.

Supplementary Table 8. Biological process gene ontology terms associated with differentially expressed genes peaking in cardiomyocyte stage and *trans*-associated with *TTN*.

Supplementary Table 9. Primer sequences.

Supplementary References

Supplementary Note 1

We also assessed chromatin dynamics in a second pluripotent stem cell line of distinct sex and origin compared to female RUES2 hESCs. For this, we performed *in situ* DNase Hi-C at various time points of cardiac differentiation of WTC11 human induced pluripotent stem cells (hiPSCs), a commonly used line derived from reprogramming of skin fibroblasts from a healthy male donor¹. Differentiating hiPSCs show the expected progression of lineage marker expression and result in highly pure hiPSC-CM (Supplementary Figure 5a-b). Of note, however, expression of a number of cardiac factors is less robust than observed for hESC-CM (Supplementary Figure 5a), indicating that cardiac maturation is slower in this hiPSC line. Accordingly, while compartment changes for MES and CP genes were remarkably similar, a subset of compartment transitions specific to the CM stage were only partially complete in hiPSC-CM (Supplementary Figure 5c-d). This indicates that similarly to gene expression and chromatin epigenetic changes, nuclear architecture dynamics can be influenced by the variability in the differentiation capacity of various hPSCs². Nevertheless, we noted that the global patterns of chromatin organization change observed in hESCs were largely reproduced in hiPSCs (Supplementary Figure 5c-e), indicating that these are conserved across multiple hPSC types of different origin and sex. Considering the stronger cardiogenic potential of RUES2 hESCs, we focused our subsequent analyses on this cell line.

3

Supplementary Figures and Figure Legends

Supplementary Figure 1. Validation of hESCs and cardiac differentiation. (**a**)

Representative flow cytometry plot of staining for TNNT2 (cTnT) of day 14 hESC-derived CMs. The gating strategy based on isotype IgG negative control staining is indicated. (**b**) Quantification of percentage cTnT positive in 11 independent differentiations; mean ± s.e.m. is shown. All are greater than 89% cTnT positive. (**c**) Karyotype analysis of RUES2 hESCs showing normal 46 X,X pattern. (**d**) RT-qPCR of stage specific markers of differentiation. mean ± s.e.m. is shown; n = 3 independent differentiations. Source data are provided as a Source Data file.

Supplementary Figure 2. Genome-wide contact maps across differentiation. (**a-e**)

Heatmap of genome-wide contact matrices between chromosomes for each time point across

differentiation and fetal heart. Blue box marks chromosome pair with decreasing signal. Red boxes mark chromosome pairs with increasing signal across differentiation. Few whole chromosome changes are observed across time points save for chromosome 15, which increases its association with the smaller chromosomes in CMs.

Replicate: - Rep1 - Rep2

Supplementary Figure 3. Intra-chromosomal interactions vary over differentiation by compartment. (**a**) t-SNE projection of HiC-Rep scores for replicates across differentiation and fetal heart. (**b**) Empirical cumulative distribution function of sizes of compartments across differentiation. Bins were merged to generate sizes of consecutive A or B domains. (**c**) Delta heatmap of the contact matrix of a region of chromosome 2. Dotted lines highlight contacts between B compartment regions. (**d-f**) Compartmentalization saddle plots. Bins were divided in ten deciles and average normalized contact frequency was calculated genome-wide. Normalization was calculated based on all bins at a given distance for *cis* interaction. (d) *cis* contacts across differentiation. (e) *trans* contacts across differentiation. (f) *cis* and *trans* contacts for fetal heart samples. (**g**) Distance plot of A-A, B-B and A-B interactions for hESC and CM, values are normalized to all contacts at a given distance (raw plot from Fig. 1j). Source data are provided as a Source Data file.

Supplementary Figure 4. Expression changes are related to A/B Compartment changes. (**a**) Expression values for stage-specific and tissue-specific markers. Cardiac specific markers are expressed in CM and fetal heart, while fetal heart expresses markers of both cardiac fibroblasts (*POSTN*) and endothelial cells (*PECAM1* and *CDH5*). (**b**) Overlap between

differentially expressed genes peaking expression in CM and genes found in chromatin domains transitioning from B to A as hESCs differentiate into CM. (**c-e**) Gene tracks of Hi-C PC1 and RNA-seq reads of (c) *ELMO1*, (d) *CXCR4*, (e) *EOMES*. Source data are provided as a Source Data file.

Supplementary Figure 5. hESCs and hiPSCs show similar changes in chromatin organization during cardiac differentiation. (**a**) RT-qPCR at the indicated time points of

hESC or hiPSC differentiation into cardiomyocytes (see Fig. 1a). PSC: pluripotent stem cell (day 0); ME: mesendoderm (day 2); CP: cardiac progenitor (day 5); CM: cardiomyocyte (day 14). Expression is relative to the housekeeping gene $HPT1$ and mean \pm s.e.m. is shown. $n = 4$ independent differentiations. p-values are calculated by two-way ANOVA followed by Holm-Sidak's multiple comparisons for hiPSC vs hESC; * < 0.05; ** < 0.01; **** < 0.0001. (**b**) Flow cytometry for TNNT2 (cTnT) in CM derived from the two independent hiPSC differentiations analyzed by *in situ* DNase Hi-C. (**c**) PC1 scores for 500 Kb-wide genomic bins computed from the Hi-C contact map at the indicated time points of hPSC differentiation; A and B compartments are shown in red and blue, respectively. Representative ME-, CP-, and CMpeaking genes that show similar compartment transitions in hESCs and hiPSCs are reported (*EOMES*, *BMPER* and *NEBL*, respectively). *TTN* is an example of a compartment transition that is not completed in hiPSC-CM by day 14 of differentiation, in line with its lower expression compared to hESC-CM (panel a). (**d**) t-SNE plot of PC1 scores. Samples at the same stage of differentiation cluster closely and are grouped within ovals. R1/2 = first/second biological replicate. (**e**) Delta compartmentalization saddle plot in *cis* contacts hiPSC-CM vs hiPSCs. Bins were assigned to ten deciles based on PC1 score, average observed/expected distance normalized scores for each pair of deciles were calculated. Source data are provided as a Source Data file.

12

Supplemenary Figure 6. TADs are dynamically regulated across differentiation. (**a-b**) t-

SNE projection of Jaccard distance for TAD boundaries based on (a) DI method and (b)

insulation score method across differentiation. Jaccard distance is calculated as 1- (intersection/union) for TAD boundaries within 80 Kb. TAD boundaries cluster by replicate. (**c**) TAD size and number within A and B compartments across differentiation for insulation method. Boxplots present the median and 25th and 75th percentile, with the whiskers extending to 1.5 times the inter-quartile range. n represents the number of TADs. n = TAD number; p-values by Wilcoxon test, *** < 0.001. (**d**) Enrichment of TAD boundaries between hESC and CM state within A/B compartment dynamics for insulation method. Log₂ values are observed/TAD union set; p-values by chi-squared test for the indicated TAD boundary set overlaps; * < 0.05. (**e**) Distance between TAD boundaries and CTCF peaks. (**f**) Expression of nearest genes to TAD boundaries that are either stage specific or shared between hESC and CM for insulation method. Box and whisker plots as in panel c. n =127 genes for hESC (lost) boundaries, 85 genes for CM (gained) boundaries, and 1477 genes for shared boundaries; p-values by onesample, two-sided t-test relative to a hypothesized mean value of 0 (no expression change vs hESC), * < 0.05, *** < 0.001. (**g**) Gene track of DI score, DI-determined TADs, insulation scoredetermined TADs, and RNA-seq data for the *KCNN2* locus in hESCs and CM. Source data are provided as a Source Data file.

Supplementary Figure 7. Local accessibility is dynamic across differentiation and reflects cell type specific transcription factors. (**a**) Density plot of ATAC read pair sizes

across differentiation for both replicates. (**b**) Heatmap of the correlation of ATAC reads across all ~138,000 peaks in the union set. Clustering pairs replicates and order the samples by differentiation state. (**c**) Normalized reads per 1 Kb for ATAC hotspots within A or B compartment across differentiation. Boxplots present the median and 25th and 75th percentile, with the whiskers extending to 1.5 times the inter-quartile range. n = ATAC peaks number; pvalues by Wilcoxon test, *** < 0.001. (**d**) Gene track across a stretch of chromosome for ATAC and A/B compartment, showing enrichment of reads within A compartment compared to B. (**e**) Distance to nearest TSS between stage-specific peaks and total peaks across differentiation. Box and whisker plots as in panel c. n = ATAC peaks number; p-values by Wilcoxon test, *** < 0.001. (**f**) Top enriched motifs identified within stage specific peaks across differentiation, as determined by DREME and TOMTOM, including POU family (OCT4) in hESC and MES, GATA family in CP and CM, and MEF family in CM. (**g**) ChIP-seq reads across CM-specific ATAC in B-A regions for GATA4, NKX2-5, TBX5 (CM) and GATA4 (ESC). (**h**) Gene track of RNA-seq, ATAC-seq, and ChIP-seq across the *ACTN2* locus. Source data are provided as a Source Data file.

Supplementary Figure 8. Cardiac specific genes are enriched in B to A regions. (**a**) Plot of

gene density for compartment dynamics across differentiation. Log₂ observed/expected is

normalized to genome fraction per compartment type. (**b**) Nearest gene distance (TSS-TSS) for upregulated genes peaking in CM stage subdividing by B to A compartment and heart development genes (GO term). Boxplots present the median and 25th and 75th percentile, with the whiskers extending to 1.5 times the inter-quartile range. $n =$ gene number; p-values by Wilcoxon test, ** < 0.01, *** < 0.001. (c) Log₂ ratio of distance to the nearest gene divided by gene size for upregulated genes peaking in CM stage subdividing by B to A compartment and heart development genes (GO term). Box and whisker plots as in panel b. Gene number and average \log_2 ratio are indicated (top and bottom, respectively); p-values by Wilcoxon test, $*$ < 0.05. (**d**) Cumulative distribution plot of the rank of heart tissue expression for each gene based on RNA-seq data from Protein Atlas. 1 indicates highest relative expression across tissues is in heart. Subdivided by upregulated genes peaking in CM stage in B to A compartment or not and non-upregulated genes. p-values by K-S test, ** < 0.01, *** < 0.001. (**e**) FPKM values for RBM20 target and *TTN trans*-associated genes *CACNA1C* (chr 12) and *CAMK2D* (chr 4) during differentiation. (**f**) Cumulative score of *trans* associations between upregulated and constitutive A compartment RBM20 target genes in hESC and CM, indicated by red line, compared to a background model of 1,000 random permutations of all upregulated, constitutive A compartment genes. The resulting random shuffling p-value is indicated. Source data are provided as a Source Data file.

18

Supplementary Figure 9. Proximity between *TTN* **and RBM20 target genes depends on** *TTN* **transcription and RBM20 expression**. (**a**) On the left, normalized maximal distance per diploid cell between the indicated loci in hESC vs hESC-CM (the number of cells is indicated). Box and whiskers plots present aggregated data from two independent cultures, and indicate median, 25th and 75th percentile, and the 10-90 percentile range. p-values by Kruskal-Wallis test followed by Dunn's multiple comparisons vs hESC (unless otherwise indicated); ns ≥ 0.05; *** < 0.001. On the right, proximity between the indicated loci in individual cells (defined as the minimal distance between the centers of the 1 μ m-wide FISH spots being less than twice their diameter). Mean ± standard deviation from two independent cultures; no statistical analysis. (**b**) On the left and right, same graphs as in panel a but for hESC-CM maintained in standard

culture conditions (Control) or treated with 5 µM Actinomycin D (ActD). In the middle, normalized distance of each locus from the nuclear periphery (the number of loci is indicated). All p-values calculated as in panel a but for ActD vs Control. (**c**) Same graphs as in panel a, but for hESC-CM obtained from the indicated gene edited hESC lines or wild-type (WT) control cells. Data from two independent cultures for all conditions but for RBM20 KO g2. All p-values calculated as in panel a but for each condition vs WT; * < 0.05. Source data are provided as a Source Data file.

Supplementary Figure 10. Generation of gene edited hESC-CM to investigate the mechanisms regulating *TTN trans* **interactions.** (**a**) Schematic of the genomic region encompassing the *TTN* promoter and the first coding exons. The position of CRISPR/Cas9 single guide RNAs (sgRNAs) employed to delete the promoter or to induce a frameshift mutation in the first coding exon is indicated; arrows depict primers used for genomic PCR. (**b**) Genomic PCR for the *TTN* promoter in selected hESC clones (Cl.) isolated after CRISPR/Cas9-

mediated gene editing with the indicated sgRNA combination; wild type (WT) hESCs were used as positive control. The analysis confirms homozygous deletion of a ~1.7 Kb genomic region in all gene edited clones (*TTN* ∆Prom). (**c**) Electropherograms from Sanger sequencing of genomic PCR products from exon 2 of *TTN* in wild type or *TTN* knockout hESCs (KO; induced by sgRNA *TTN* ex2). The 20 bp-sgRNA targeting sequence is highlighted, and the protospacer adjacent motif (PAM) site is underlined A homozygous "C" insertion leads to a p.Val15AlafsTer21 mutation. (**d**) G-banding karyotype of *TTN* knockout and *TTN* ∆Prom hESCs; a normal 46 XX female karyotype is confirmed. (**e**) Schematic of the *RBM20* locus. The position of two sgRNAs employed to induce a frameshift mutation in the second exon is indicated. (**f**) As in c, but for genomic PCR products from exon 2 of *RBM20.* RBM20 knockout lines have homozygous insertions of a single nucleotide leading to p.Arg178ThrfsTer230 and p.Glu264Ter mutations for sgRNA1 and sgRNA2, respectively. (**g**) As in d, but for the two *RBM20* knockout hESC lines. (**h**) Representative flow cytometry for TNNT2 (cTnT) in cardiomyocytes derived from the indicated hESCs, indicating that all gene edited lines could be successfully differentiated in highly-pure cardiomyocytes.

Supplementary Figure 11

Supplementary Figure 11. Uncropped gel images. Unmodified Western blot and DNA

agarose gel electrophoresis images from the indicated figure panels. The cropping masks are shown by dotted boxes, and molecular weights are indicated.

Supplementary Tables

Supplementary Table 1. Hi-C sequencing statistics.

Supplementary Table 2. HiC-Rep scores between samples.

Supplementary Table 3. RNA-seq sequencing statistics.

Supplementary Table 4. Biological process gene ontology terms associated with differentially expressed genes peaking in cardiomyocyte stage.

Supplementary Table 5. Biological process gene ontology terms associated with differentially expressed genes peaking in cardiomyocytes in regions of the genome that transition B to A.

Supplementary Table 6. Jaccard index scores for TAD boundary overlap.

Supplementary Table 7. ATAC-seq sequencing statistics.

Supplementary Table 8. Biological process gene ontology terms associated with differentially expressed genes peaking in cardiomyocyte stage and *trans***-associated with** *TTN***.**

Supplementary Table 9. Primer sequences.

Supplementary References

- 1. Kreitzer, F. R. *et al.* A robust method to derive functional neural crest cells from human pluripotent stem cells. *Am. J. Stem Cells* **2,** 119–31 (2013).
- 2. Ortmann, D. & Vallier, L. Variability of human pluripotent stem cell lines. *Curr. Opin. Genet. Dev.* **46,** 179–185 (2017).