Transcriptional control of cardiac conduction system development and function

van Weerd, J.H.

Publication date
2020
Document Version
Other version
License
Other

Chapter 4

A large permissive regulatory domain exclusively controls *Tbx3* expression in the cardiac conduction system

Jan Hendrik van Weerd*
Ileana Badi*
Malou van den Boogaard
Sonia Stefanovic
Harmen J.G. van de Werken
Melisa Gomez-Velazquez
Claudio Badia-Careaga
Miguel Manzanares
Wouter de Laat
Phil Barnett
Vincent M. Christoffels

* these authors contributed equally

Published in:
Circulation Research, 2014:115(4);432-41

ABSTRACT

Rationale: The evolutionary conserved *Tbx3/ Tbx5* gene cluster encodes T-box transcription factors that play crucial roles in the development and homeostasis of the cardiac conduction system (CCS) in human and mouse. Both genes are expressed in overlapping patterns and function in strictly tissue-specific and dose-dependent manners, yet, their regulation is poorly understood.

Objective: To analyse the mechanism underlying the complex regulation of the *Tbx3/ Tbx5* cluster.

Methods and Results: By probing the 3D architecture of the *Tbx3/Tbx5* cluster using high-resolution 4C-sequencing *in vivo*, we found that its regulatory landscape is in a preformed conformation similar in embryonic heart, limbs and brain. *Tbx3* and its flanking gene desert form a 1 Mbp loop between CTCF binding sites that is separated from the neighboring *Tbx5* loop. However, *Ctcf* inactivation did not result in transcriptional regulatory interaction between *Tbx3* and *Tbx5*. Multiple sites within the *Tbx3* locus contact the promoter, including sites corresponding to regions known to contain variations in the human genome influencing conduction. We identified an atrioventricular-specific enhancer and a pan-cardiac enhancer that contact the promoter and each other and synergize to activate transcription in the atrioventricular conduction system.

Conclusions: We provide a high-resolution model of the 3D structure and function of the *Tbx3/ Tbx5* locus and show that the locus is organized in a preformed, permissive structure. The *Tbx3* locus forms a CTCF-independent autonomous regulatory domain with multiple combinatorial regulatory elements that control the precise pattern of *Tbx3* in the CCS.

INTRODUCTION

The heart comprises working muscle and an integrated cardiac conduction system (CCS) that initiates and propagates the electrical impulse required for the coordinated contraction of the chambers. The development of the CCS is regulated by transcription factors that act in strictly stage, location and dose-dependent manners.[1,2] Disrupted function or expression of these factors may lead to disorders in development or function of the CCS, which are associated with morbidity, heart failure and sudden death.[3] The evolutionary conserved *Tbx3/Tbx5* gene cluster encodes T-box transcription factors that play key roles in the development and function of multiple organ systems including heart, limbs, lungs and liver.[4] Haploinsufficiency of *TBX5* causes Holt-Oram syndrome (OMIM [Online Mendelian Inheritance of Man]: 142900), a developmental disorder characterized by hand-heart defects,[5,6] whereas haploinsufficiency of *TBX3* causes the ulnar-mammary syndrome (OMIM: 181450), a developmental disorder characterized by limb, mammary gland, tooth and genital abnormalities.[7] Both genes play crucial roles in the development and patterning of the CCS. In the heart, *Tbx5* is expressed in a pattern including the atria and CCS and acts as an activator of genes promoting conduction.[8,9] *Tbx3* is expressed specifically in the CCS components and functions by imposing a pacemaker phenotype on cells within its expression domain.[10] Ectopic expression of Tbx3 in the myocardium induces a pacemaker phenotype.[10,11] A decrease in Tbx3 below critical levels results in lethal arrhythmias.[12] Therefore, the correct spatiotemporal pattern and dosage of expression of *Tbx3* and of *Tbx5* are critical for development and functional integration of the CCS.

The regulatory architecture of loci harboring genes encoding transcription factors or other developmental factors often comprises long-range regulatory elements, each responsible for a subset of the temporal, spatial and/or quantitative aspects of expression of the gene.[13] Genome-wide association studies (GWAS) suggest that a large number of noncoding variants underlie the increased risk to various common diseases, often by disrupting regulatory elements that control the expression of nearby genes.[14,15] GWAS revealed that common variants in the noncoding regions around *TBX3* and *TBX5* are associated with cardiac impulse conduction and depolarization properties,[16,17] whereas a congenital heart defect patient has been identified with a homozygous variant in an enhancer regulating *TBX5.*[18] These findings further underscore the critical importance of understanding the regulatory system controlling the expression pattern and dosage of Tbx3 and Tbx5.

The three-dimensional (3D) organization of chromatin has been recognized as an important regulatory layer for gene regulation, setting the stage for physical contact of the control sequences, such as enhancers, and their target promoters.[19] High-resolution circular chromosome conformation capture combined with sequencing (4C-seq) is a recent technology examining the spatial organization of DNA and measuring the contact frequencies of a chosen genomic site, or 'viewpoint', with the entire genome.[20] We assessed the 3D architecture of the *Tbx3/Tbx5* cluster using 4C-seq to gain insight into its regulation and to find potential genomic

targets that could underlie congenital defects. Our work provides a high-resolution model of the 3D structure and function of the *Tbx3/Tbx5* locus, which reveals that the *Tbx3* and *Tbx5* loci form autonomous regulatory domains that do not interact, and that the gene desert upstream of *Tbx3* harbors multiple regulatory elements that synergistically control the precise pattern of *Tbx3* in the CCS.

RESULTS

Physical separation of *Tbx3* and *Tbx5* regulatory domains

To gain insight into the regulation of *Tbx3*, we assessed the 3D architecture of the *Tbx3/Tbx5* cluster using high-resolution 4C-seq of embryonic day (E) 10.5 mouse hearts. The 4C-seq experiments show results typical for chromosome conformation capture experiments, in that most of the signal is close to the viewpoint, intrachromosomal contacts outnumber interchromosomal contacts, and contacts at distal sites tend to cluster.[20,21] To assess the conformation of the *Tbx3* locus and identify putative regulatory regions that contact *Tbx3* and may regulate its expression, the *Tbx3* promoter was used as viewpoint. The broad peak around the *Tbx3* promoter reflects contact by linear proximity, whereas more distant signal clusters (peaks) represent discernible interactions through looping. The capture profile reveals that the *Tbx3* promoter contacts multiple regions upstream of *Tbx3*, with interactions reaching up to neighboring gene *Med13l*, located 903 kbp upstream (Figure 1A). Analysis of these data reveals relatively strict boundaries of the contact profile, with most contacts being confined to the domain spanning approximately 900 kbp upstream to 50 kbp downstream of *Tbx3* (Figure 1A). Interestingly, analysis of chromosomal contacts with *TBX3* and *TBX5* as assessed by Hi-C in human fibroblasts[21] reveals, albeit in much lower resolution, a similar pattern with multiple regions in the *TBX3* domain contacting the gene. Again, contacts with the *TBX5* or *MED13L* domains are almost absent (Supplemental Figure 1).

The contact profile of the promoter of the adjacent *Tbx5* gene reveals an independent, neighboring, regulatory domain with boundaries near *Tbx3* and the neighboring *Rbm19* gene, located approximately 230 kbp downstream of *Tbx5* (Figure 1B). The two domains hardly overlap, indicating that sites in the *Tbx3* domain do not contact the *Tbx5* domain and vice versa. *Med13l* is isolated from these regulatory landscapes, as revealed by the contact profile of the *Med13l* promoter (Supplemental Figure 1). In contrast, the contact profile of *Rbm19*, downstream of *Tbx5*, partially overlaps with that of *Tbx5* but not with the *Tbx3* regulatory landscape (data not shown). This demonstrates that *Tbx3* forms a large independent regulatory domain not utilized by the neighboring genes *Tbx5*, *Med13l* or *Rbm19*. The strict boundaries of the contact profiles correspond with the location of CTCF-binding sites in heart and other tissues[22] (Figure 1D, Supplemental Figure 2A). CTCF is involved in chromatin looping and in insulation of loci.[23,24] We studied the interaction profile of one of the CTCF-binding sites in between *Tbx3* and *Tbx5.* This site physically interacts with distal CTCF-binding sites-containing regions located 974 kbp up- and 358 kbp downstream, forming chromatin loops that demarcate the *Tbx3* and *Tbx5* regulatory landscapes, respectively (Figure 1C). These observations suggest that CTCF may mediate the physical separation of the regulatory domains of *Tbx3* and *Tbx5* (Figure 1E), preventing enhancer sharing between these genes. To test this possibility, we analyzed the expression of *Tbx3* and *Tbx5* in embryos in which *Ctcf* was inactivated specifically in the cardiomyocytes. However, neither the level nor the pattern of *Tbx3* and *Tbx5* were

affected, indicating that CTCF is not required to prevent the separation of the regulatory sequences of *Tbx3* and *Tbx5* (Supplemental Figure 2B,C).

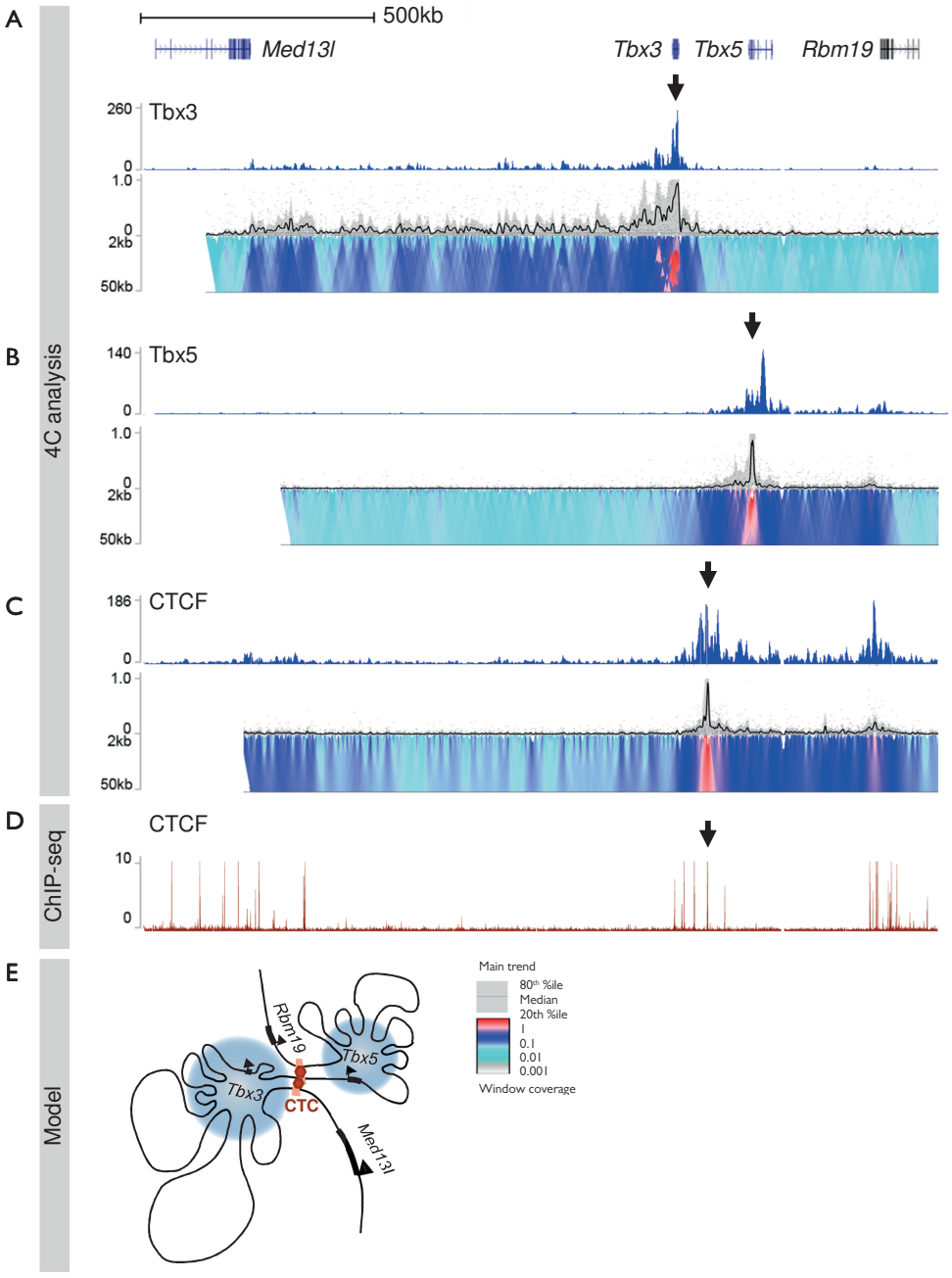
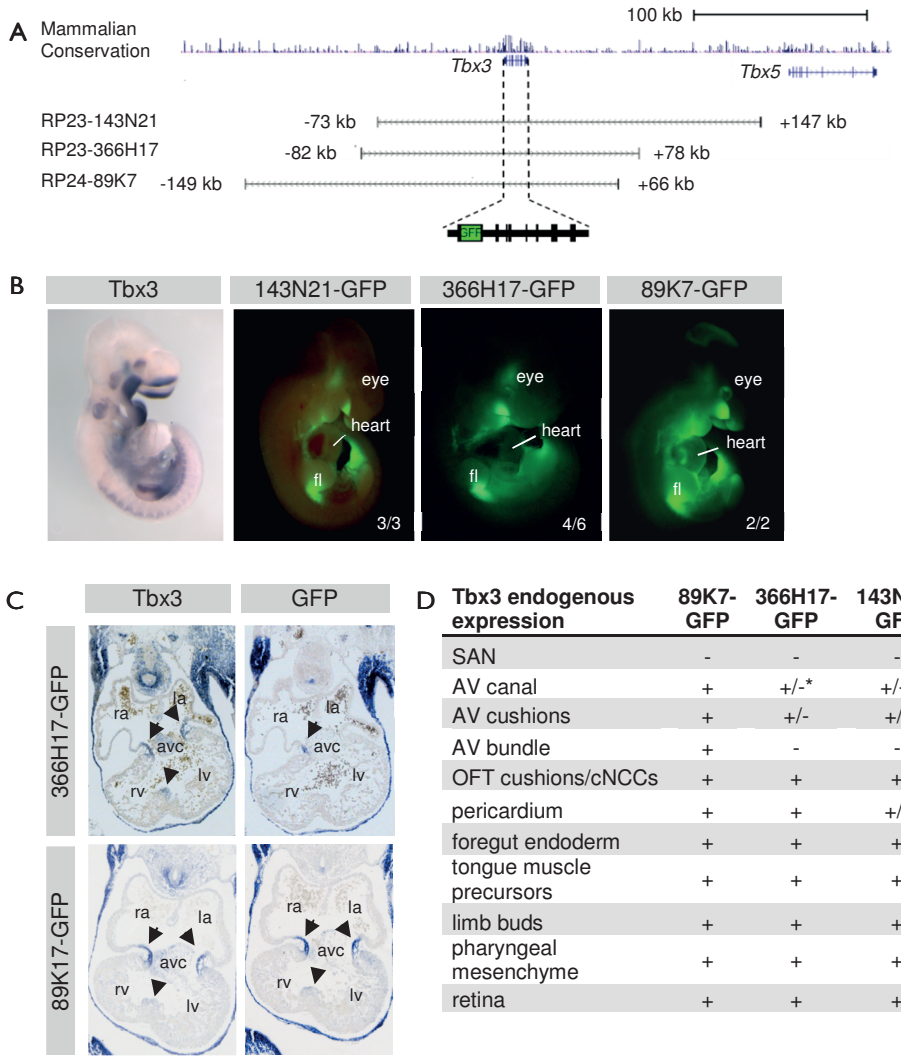


Figure 1. Separation of the *Tbx3* and *Tbx5* regulatory domains. (**A-D**) Contact profiles of the *Tbx3* and *Tbx5* loci as determined by 4C reveal genomic regions that physically contact the *Tbx3* promoter in embryonic heart cells. The upper track shows the UCSC genome browser traces of a running window approach for viewpoints of the *Tbx3* promoter (**A**), *Tbx5* promoter (**B**) and CTCF site (**C,D**). The second track shows the respective normalized contact intensities (grey dots) and their truncated mean trends (black line) for all viewpoints in and around *Tbx3* and *Tbx5*. Means are computed for 5 kbp windows and the grey band displays the 20–80% percentiles for these windows. The scaling-plot below depicts data points across differently scaled window sizes (from 2 kbp (top row) to 50 kbp (bottom)). The data points and the colors are scaled to the maximum mean value of a 5 kbp window size.[20] Local changes in colour codes indicate regions enriched for captured sequences, which correspond to the promoter-enhancer contacts described. The contact profile of *Tbx5* (B) reveals a complementary pattern to that of *Tbx3* (A), with hardly any overlap in contacting regions. (**C**) The contact profile of the CTCF-binding site in between *Tbx3* and *Tbx5* (marked by arrow in (D)) as determined by 4C reveals interaction with neighboring CTCF-binding site islands (marked by asterisks). (**D**) Distribution of CTCF binding sites as determined by ChIP-seq in heart cells.[30] CTCF-binding sites are located in between *Tbx3* and *Tbx5* and near flanking genes *Med13l* and *Rbm19*, colocalizing with the boundaries in contact profiles of *Tbx3* and *Tbx5* promoters. Arrow marks CTCF site used for 4C. (**E**) Schematic overview of the *Tbx3* and *Tbx5* loci, depicting the physical separation of the *Tbx3* and *Tbx5* regulatory domains.

Active enhancers are marked by acetylation of histone H3 on lysine K27 (H3K27ac).[25] Chromatin immunoprecipitation (ChIP)-sequencing with an antibody against H3K27ac in both embryonic chamber and atrioventricular canal (AVC) myocardial cells[26] revealed that the region flanking *Tbx3* is extensively acetylated on H3K27 in AVC cells, corresponding to the AVC-specific expression of *Tbx3* (Supplemental Figure 3). In contrast, the region flanking *Tbx5* is acetylated in both chamber myocardium and AVC, corresponding to its expression in both compartments. The boundary between the complementary H3K27ac signatures colocalizes roughly with the boundary of the contact profiles for *Tbx3* and *Tbx5*. Taken together, our data indicate that the *Tbx3* locus forms an independent regulatory domain which is physically separated from that of *Tbx5*, and that this domain contains multiple regions that contact the *Tbx3* gene.

GFP-modified BACs recapitulate atrioventricular conduction system expression of *Tbx3*

While 4C-seq physically wires gene promoters to potential enhancer sequences, physical contact does not necessarily reflect functional interaction.[27] To identify regulatory elements within the *Tbx3* locus, we generated transgenic mice with modified bacterial artificial chromosomes (BACs). BAC RP23-143N21 contains genomic sequences from -73 kbp to +147 kbp relative to the transcription start site of *Tbx3*, BAC RP23-366H17 contains sequences from -82 kbp to +78 kbp and BAC RP24-89K7 contains sequences from -149 to + 66 kbp. To investigate whether regulatory elements required for cardiac Tbx3 expression are located within the regions covered by these BACs, we modified them by inserting a green fluorescent protein (GFP)-encoding reporter gene at the translation start site of *Tbx3* (143N21-GFP, 366H17-GFP, 89K7-GFP; Figure 2A), also resulting in the inactivation of *Tbx3* within the BAC, negating any potential gain of Tbx3



D

Tbx3 endogenous expression	89K7-GFP	366H17-GFP	143N21-GFP
SAN	-	-	-
AV canal	+	+/-*	+/-*
AV cushions	+	+/-	+/-
AV bundle	+	-	-
OFT cushions/cNCCs	+	+	+
pericardium	+	+	+/-
foregut endoderm	+	+	+
tongue muscle precursors	+	+	+
limb buds	+	+	+
pharyngeal mesenchyme	+	+	+
retina	+	+	+

Figure 2. GFP-modified BACs recapitulate AV conduction system expression of Tbx3. (**A**) Tbx3 locus with regions spanned by BACs. A GFP cassette was inserted in the translational start site of *Tbx3* to obtain BACs 143N21-GFP, 366H17-GFP and 89K7-GFP. (**B**) Whole-mount in-situ hybridization on E11.5 embryo showing endogenous *Tbx3* expression pattern. GFP expression patterns in F0 embryos of 143N21-GFP, 366H17-GFP and 89K7-GFP reveals recapitulation of many aspects of extracardiac *Tbx3* expression pattern. Numbers depict number of embryos with reproducible GFP expression patterns among the total number of independent transgenic embryos. (**C**) In-situ hybridization on sections through 366H17-GFP and 89K7-GFP reveals Tbx3 and GFP expression patterns. Arrows and arrowheads depict expression in AVC and interventricular region, respectively. (**D**) Table depicting GFP-positive regions in 89K7-GFP, 366H17-GFP and 143N21-GFP embryos. av; atrioventricular, avc; atrioventricular canal, fg; foregut, fl; forelimb, la; left atrium, lv; left ventricle, ra; right atrium, rv; right ventricle, san; sinoatrial node, oft; outflow tract, cnccs; cardiac neural crest cells.

function in transgenic mice. Analysis of the GFP expression pattern in F0 embryos of embryonic day (E) 11.5 revealed that all three modified BACs are able to recapitulate many aspects of the extracardiac *Tbx3* expression pattern. GFP expression was found in the retina, limb buds, mammary glands and pharyngeal mesenchyme (described earlier for 366H17-GFP[28]), corresponding to the endogenous *Tbx3* expression pattern (Figure 2B). *In situ* hybridization revealed that 143N21-GFP and 366H17-GFP recapitulate *Tbx3* expression in the foregut endoderm and ventral body wall (Figure 2C). Cardiac GFP expression was detected in the AVC myocardium, but in a domain smaller than that of *Tbx3* itself. The GFP expression domain included ventral, right-sided and dorsal aspects (primordial atrioventricular (AV) node) of the AVC. The *Tbx3*-positive AV bundle primordium (interventricular ring) did not express GFP, indicating that the region from -82 kbp to +147 kbp of *Tbx3* lacks DNA sequences for *Tbx3* activation in the complete CCS. GFP expression was also observed in *Tbx3*-positive cardiac neural crest cells distal from the outflow tract and the outflow tract cushions (neural crest-derived cells) (Supplemental Figure 4). 89K7-GFP, spanning an additional 60 kbp region upstream of the other BACs, drove GFP expression in the entire AVC, recapitulating the pattern of *Tbx3*. Furthermore, GFP expression was observed in the interventricular ring, indicating that the region -149 kbp to -82 kbp upstream of *Tbx3* harbors regulatory elements required for *Tbx3* expression in the prospective AV bundle. These data reveal that the genomic region of -82 kbp to +147 kbp of *Tbx3* harbors regulatory elements that drive AV activity of *Tbx3*, whereas the region from -149 kbp to -82 kbp of *Tbx3* harbors additional regulatory sequences driving the full AVC and bundle expression pattern of *Tbx3*. We observed no GFP expression in the sinoatrial node, indicating that the regulatory elements driving *Tbx3* expression in this region are absent from the regions spanned by the tested BACs.

Two putative enhancers physically contact the *Tbx3* promoter

To identify potential enhancers within the *Tbx3* BACs conferring tissue-specific expression, we analyzed the contact profile of the promoter by 4C in embryonic heart, brain and limb cells. Contact profiles of all three cell types revealed a highly similar pattern with multiple sites contacting the *Tbx3* promoter (Figure 3A). Two regions showed strong interaction with the promoter, one approximately 25 kbp upstream of the promoter and present in all three modified BACs, and one at approximately 90 kbp upstream of the promoter and present in only BAC 89K7-GFP. To identify putative enhancers within the contact regions, we evaluated ChIP-sequencing datasets[29-32] of the genome-wide occupancy profiles of cardiac transcription factors Gata4, Nkx2-5, Tbx5, Tbx3, the enhancer-associated histone acetyltransferase p300, RNA Polymerase II (Pol2) and histone modifications H3K27ac and monomethylation of histone H3 on lysine K4 (H3K4me1) (Figure 3B). Based on these occupancy profiles two candidate regions were identified upstream of *Tbx3*. Putative enhancer A (eA) is located 19 kbp upstream of *Tbx3*, within the proximal contact region and is occupied by Tbx5, Gata4 and Nkx2-5 and marked by H3K4me1. Putative enhancer B (eB), which shows occupancy by all factors except Pol2, is located

95 kbp upstream of *Tbx3* within the distal contact region that is present only in BAC 89K7 (Figure 3B,C, Supplemental Figure 5). Tbx3, H3K27ac, p300 and Pol2 occupancy of eA and Pol2 occupancy of eB were not called by Model-based Analysis of ChIP-Seq (MACS). However, inspection of the raw data tracks revealed distinguishable peaks for all these factors on both sites (Supplemental Figure 6A). The weak signal might be explained by the fact that the ChIP-seq datasets are derived from whole hearts, whereas the CCS in which *Tbx3* is active comprises only a small fraction of the total cardiac cell population. The association of both enhancers with the enhancer-associated histone modifications H3K27ac and H3K4me1 was confirmed by ChIP-quantitative PCR (ChIP-qPCR). Both putative enhancers showed enrichment of H3K27ac and H3K4me1 occupancy over a negative control region, revealing their possible role as transcriptional enhancers (Supplemental Figure 6B).

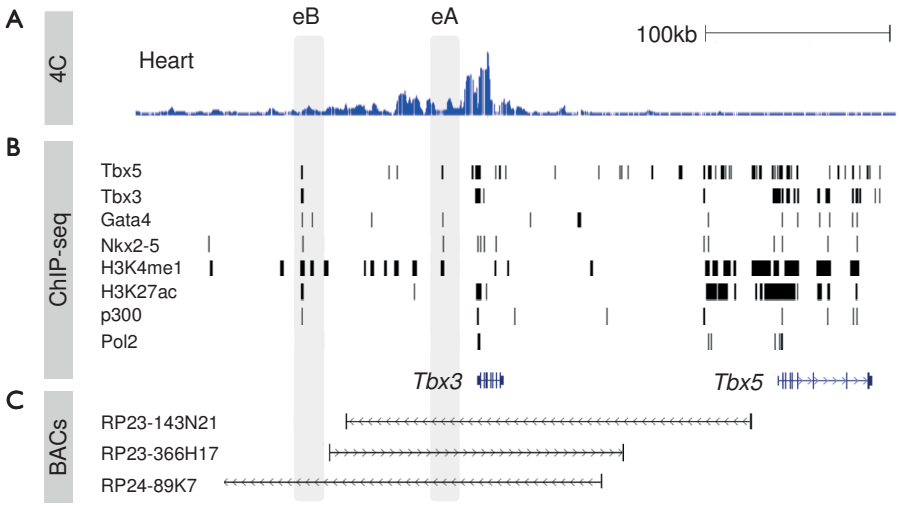


Figure 3. Identification of two putative enhancers contacting *Tbx3*. (**A**) Contact profile in heart with Tbx3 promoter as viewpoint as assessed by 4C reveals multiple sites contacting *Tbx3*. Y-axis depicts contact tag count. (**B**) Occupancy profiles of the *Tbx3* locus of cardiac transcription factors Tbx5, Tbx3, Gata4 and Nkx2, histone modifications H3K4me1 and H3K27ac and enhancer-associated proteins p300 and Pol2 and as assessed by ChIP-seq. Peaks were called using the MACS peak calling algorithm. (**C**) Two potential enhancers are identified which are present in all three modified BACs (eA) or exclusively in 89K7-GFP (eB).

We used putative eA and eB as viewpoints in 4C-seq analysis, and found that eA strongly interacts with both eB and the *Tbx3* promoter. The same holds for eB, albeit to a much lesser extent (Figure 4 and Supplemental Figure 7). These findings reveal that both enhancers and the Tbx3 promoter are spatially clustered together. Comparison of contact profiles from heart, limb and brain indicated that the chromosomal conformation of the *Tbx3* locus is not tissue-specific, since the contact profiles for all three cell types with different viewpoints reveal a striking similarity

(Figure 4). Only minor differences in contact profiles are observed between heart and limb. In brain, slightly more appreciable differences are observed in the contact profiles. The region upstream of *Tbx3,* encompassing both eA and eB, shows more enhancer-promoter contacts in heart as compared to brain. Furthermore, the *Tbx3* promoter contacts *Tbx5* and several flanking intergenic regions in brain, exceeding the boundaries marked by CTCF binding sites, and vice versa (Figure 4B). The separation between both domains is therefore less strict in embryonic brain.

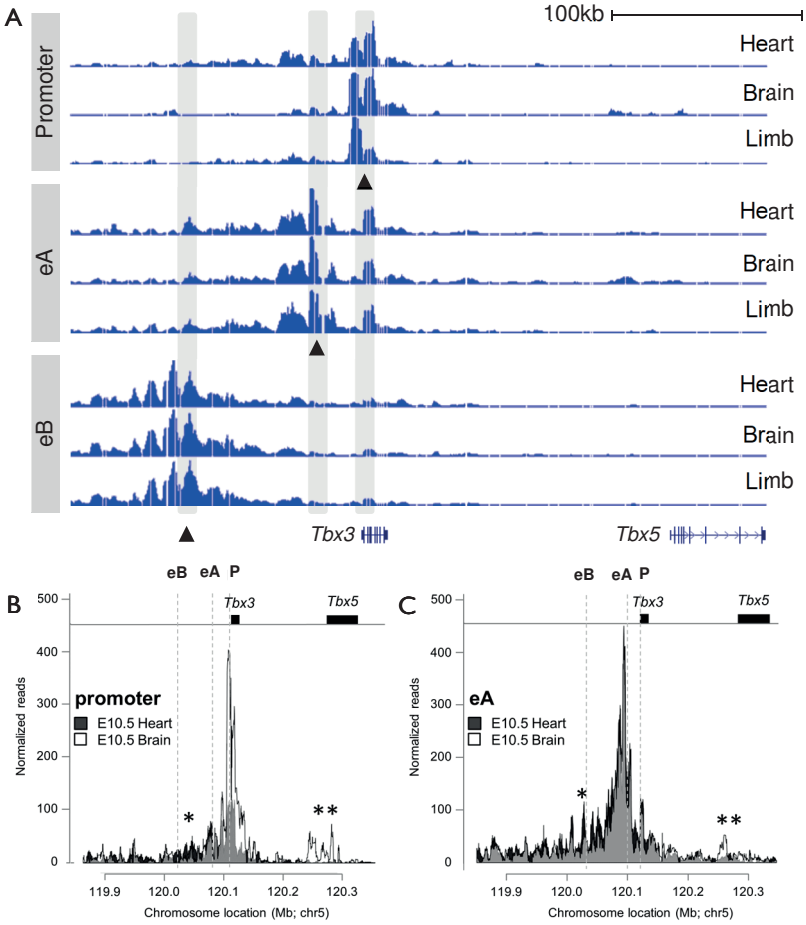


Figure 4. eA and eB physically interact with *Tbx3* in a tissue-independent manner. (**A**) Contact profiles for *Tbx3* promoter, eA and eB as assessed by 4C in different cell types. Arrows indicate points of view for the depicted tracks. Both eA and eB physically contact the *Tbx3* promoter and vice versa. Asterisks indicate contacts close to *Tbx5* with eA and the Tbx3 promoter. (**B,C**) Overlap of heart and brain contact profiles for Tbx3 promoter (B) and eA (C) reveal minor differences. In heart, more contacts are observed in the region upstream of Tbx3 for both viewpoints (asterisks), whereas in brain more contacts with the Tbx5 promoter are observed (double asterisks).

eA and eB are required for CCS expression of *Tbx3*

To investigate whether putative enhancers eA and eB are involved in the regulation of *Tbx3* expression, fragments containing the enhancers were deleted from 89K7-GFP to construct ΔeAeB-89K7-GFP. Inspection of GFP expression in E10.5 F0 ΔeAeB-89K7-GFP embryos revealed that extracardiac *Tbx3* expression was largely unaffected by deletion of both fragments (Figure 5). GFP expression was detected in the limb buds, eye placode, pharyngeal arches and cardiac neural crest cells in a pattern that is similar to 89K7-GFP, indicating that *Tbx3* expression in these regions is not mediated by eA and eB. However, GFP expression in the heart is completely abolished in ΔeAeB-89K7-GFP embryos. Immunohistochemistry revealed that GFP expression in the AVC, AV bundle, foregut endoderm and ventral body wall is absent in ΔeAeB-89K7-GFP embryos (Figure 5). These findings reveal putative enhancers eA and eB are required for CCS expression of *Tbx3*.

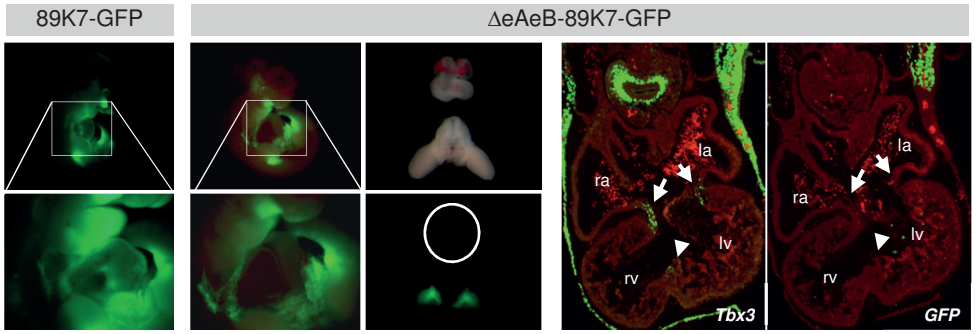


Figure 5. eA and eB are required for AVCS expression of *Tbx3*. 89K7-GFP recapitulates (extra)cardiac expression pattern of *Tbx3*. Deletion of both eA and eB from 89K7-GFP (ΔeAeB-89K7-GFP) leads to complete loss of cardiac GFP expression, whereas extracardiac expression is unaffected. Rectangles depict regions shown in greater magnification below. Circle depicts loss of GFP expression in the heart. Arrows depict atrioventricular canal, arrowheads depicts interventricular region.

Evolutionary conserved eA and eB are sufficient to drive atrioventricular *Tbx3* expression

To elucidate whether both enhancer fragments on their own are sufficient to drive atrioventricular Tbx3 expression, we isolated fragments of approximately 2 kbp spanning eA or eB, cloned them in a LacZ-reporter vector with a minimal promoter (eA-LacZ and eB-LacZ, respectively) and assessed their *in vivo* activity. F0-screening of transient transgenic eA-LacZ E10.5 embryos revealed that eA is sufficient to drive LacZ expression in the AVC, strong in the ventral, right and dorsal aspects, weak at the left side (Figure 6A). This cardiac expression pattern corresponds to the GFP expression as observed in BACs 143N21-GFP and 366H17-GFP, in

which eA but not eB is located. Consistent LacZ expression was also observed in the pharyngeal arches, whereas ventral body wall expression was not detected in any of the embryos.

eB-LacZ embryos showed a variable expression pattern. Two embryos showed cardiac expression in only few spots in the heart not correlating with the pattern of endogenous *Tbx3* expression, whereas two other embryos showed cardiac expression throughout the entire heart (Figure 6B). This indicates that on its own eB is cardiac-specific, is weak and influenced by repressive effects of the genomic region at the site of its integration, and, within the heart, lacks spatial information. Interestingly, when coupled to eA and tested in transient transgenic mouse embryos (eAeB-LacZ), robust expression was observed throughout the entire AVC and prospective AV bundle (interventricular ring; Figure 6C). Strong ventral body wall expression was also observed in eAeB-LacZ embryos, in contrast to either eA-LacZ or eB-LacZ. This expression pattern in eAeB-LacZ embryos is highly similar to the GFP-expression we observed in 89K7-GFP embryos, corresponding to the presence of both eA and eB in BAC 89K7. These data imply that eA and eB do not act as independent modules, but act in synergy, whereby eB enhances the activity of eA and complements the pattern of eA in the AV conduction system precursors (Figure 6G).

To investigate whether the mechanism of Tbx3 regulation in the AV conduction system is evolutionary conserved we tested the human and chicken orthologues of eA in vivo. Human eA-LacZ showed an expression pattern comparable to its murine counterpart, with expression in the major portion of the AVC, mainly in the ventral, right and dorsal side (Figure 6D). Weak activity was also observed in the prospective AV bundle. Chicken eA-LacZ shows similar AV-ring expression, however, expression in the prospective AV bundle was not detected in any of the embryos (Figure 6E). The chicken eA-LacZ construct also showed clear expression around the fore limbs, possibly reflecting Tbx3 expression in the zone of polarizing activity implicated in limb bud development, which was not detected in the human or murine orthologue of eA. The inter-species consistency in AVC activity of eA indicates that aspects of the Tbx3-mediated formation of the CCS are evolutionary conserved.

Tbx3 regulation in the AVC by eA is mediated by BMP signaling.

We recently found that several AV canal-specific enhancers are GATA-binding site dependent and are synergistically activated by Gata4 and Smads (downstream BMP-signaling effectors), which recruit histone acetyltransferases such as p300.[26,33] Since the enhancers here are occupied by Gata4 and p300, we hypothesized that they may be activated by Gata4 and BMP-signaling. In transient transfection assays, we found that eA and eB are not activated by Gata4 or BMP effectors alone. Co-transfection with Gata4 and BMP effectors, however, resulted in a synergistic induction of eA (Figure 6F). Transfection of two sub-fragments of eA revealed that the regulatory information of this enhancer is harbored within eA2. Consistently, this sub-fragment also drives AV canal expression in transgenic embryos (not shown). The eB enhancer fragment was not strongly activated. Together these data are consistent with the observed activity patterns of the

two enhancer fragments *in vivo*, in which eA provides AV canal specificity whereas eB is weak and does not show specificity within the heart.

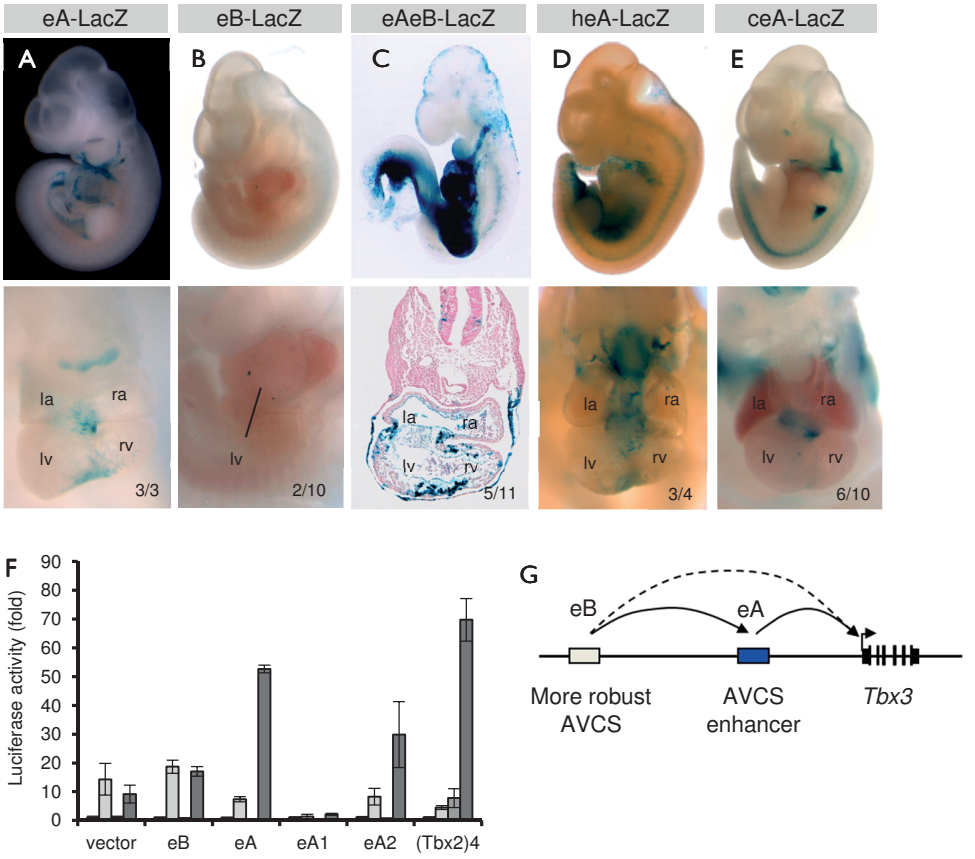


Figure 6. eA and eB are sufficient to drive AVC *Tbx3* expression and are evolutionary conserved. (**A**) Transient transgenic eA-LacZ embryos (E10.5) reveals that eA is sufficient to drive Tbx3 expression in the AVC and part of the interventricular septum. (**B**) eB-LacZ expression is observed in only few spots in the heart or throughout the entire heart (not shown). (**C**) eA and eB synergize to drive robust Tbx3 expression in the AVC and interventricular septum. (**D**,**E**) Human (D) and chicken (E) orthologues of eA drive expression in the AVC indicating evolutionary conservation of eA. (**F**) Luciferase reporter assay on eA and its two subfragments, eB and a previously described tandemly repeated Tbx2 enhancer [26]. Cos-7 cells were co-transfected with Gata4 and Smad expressing vectors. Bars represent average ± s.d. (n=3). (**G**) Model depicting the synergy between eA and eB to drive AVC expression of *Tbx3.*

DISCUSSION

The T-box transcription factors *Tbx5* and *Tbx3* are crucial for the development and function of the heart and its electrical components. In this study, we provided a high-resolution model of the 3D conformation of the evolutionary conserved *Tbx3/Tbx5* genomic locus using high-resolution 4C-seq *in vivo*. We found that *Tbx3* and *Tbx5* are physically organized into separate loci and that the genomic organization is independent of tissue-type. We identified two enhancers within the *Tbx3* domain that contact each other and the *Tbx3* promoter, and that synergize to drive *Tbx3* expression in the AV conduction system.

Tbx3 and *Tbx5* form an evolutionary conserved cluster of developmental transcription factor genes originating from a primitive T-box gene.[34] Studies on the transcriptional regulation of clustered developmental genes, like the *Irx* and *Hox* clusters, revealed that regulatory DNA elements are not uniquely associated with single promoters.[35,36] Instead, genes within these clusters extensively share enhancers, coordinating the regulation of their expression pattern. It has been proposed that enhancer sharing explains the conservation of the genomic organization of such clusters during evolution.[35,36] Our data indicate that, contrary to the *Irx* and *Hox* gene clusters, the *Tbx3/Tbx5* cluster is divided into two separate regulatory domains. Interdomain contacts appear almost absent in all tissue types we tested, indicating that *Tbx3* and *Tbx5* are regulated independently. This also implies that genomic variation affecting regulatory elements of either one of the genes does not directly influence the expression pattern and level of the other.

The strict boundaries between the *Tbx3* and *Tbx5* domains are marked by CTCF binding sites. CTCF-bound sites can function as boundary regions, interacting with each other to physically segregate the chromosome into topologically associated domains (TADs) within which enhancer-promoter contacts have been shown to be particularly frequent.[23,37,38] CTCF binding sites have been suggested to isolate genes or gene clusters from neighboring transcriptional interference. However, in cardiac-specific *Ctcf* mutants we did not observe any ectopic expression of *Tbx3* in the larger *Tbx5* expression domain or any change in the level of expression of either gene. These data indicate that although CTCF binding marks regions involved in looping and locus separation, CTCF itself is not required to transcriptionally insulate the loci from neighboring regulatory activity. Consistently, inactivation of *Ctcf* in erythroid cells or limb buds did not lead to ectopic activation of genes adjacent to the *β-globin* locus or *Hoxd* locus, respectively.[24,39]

Tissue-specific gene expression is mediated by specific interactions of the promoter with distal tissue-specific enhancers, and as such a tissue-specific 3D chromosomal conformation is expected for different cell types. Contrary to this, however, our data indicate that the chromosomal architecture of the *Tbx3/Tbx5* locus is strikingly similar between different tissue types. Contact profiles in heart and limb cells reveal a fixed chromosomal architecture with multiple sites contacting the *Tbx3* promoter, including eA and eB. In brain cells, the contact profile appears somewhat different, displaying slightly more *Tbx3-Tbx5* interdomain contacts than other

tissues tested (Figure 4B,C). *Tbx3* is expressed in the developing heart and limbs, whereas its expression in the developing brain is very limited or absent. The observed variation in the topology of the *Tbx3* locus might thus correspond to its inactivity in this tissue. We propose that the genomic contacts of eA, eB and the *Tbx3* promoter are mostly tissue-independent and organized in a preformed, permissive chromatin structure, in which functional enhancer-promoter interactions are facilitated by the rigid, cell-type invariant structural enhancer contacts.[19] The enhancer regions and the promoter are in close proximity irrespective of tissue type, which thus facilitates CCS-specific transcription factor complexes in the activation of *Tbx3* transcription in the CCS. In other cell types the CCS-specific enhancers are not activated but remain in close proximity to the promoter.

Analysis of the chromosomal interactions in human fibroblasts assessed by Hi-C[40] reveals a similar pattern as observed in mouse. Multiple regions upstream of *TBX3* contact the gene and these regions do not overlap with regions contacting *TBX5* or *MED13L* (Supplemental Figure 1). This indicates that the genomic organization of the *Tbx3/Tbx5* locus is not only conserved between tissues, but also between mammalian species, which reveals a robust conservation of the topological *Tbx3/Tbx5* domains. Although CTCF appeared to be not required for the insulation of the *Tbx3* regulatory domain, it has been well-described as a factor frequently present at the boundaries of topologically associated domains. The highly similar occupancy profiles of CTCF between different tissue types (Supplemental Figure 2A) therefore further indicates that the 3D organization of the *Tbx3* locus is largely conserved and cell type-independent.

GWAS have identified common genetic variants in the human *TBX3/ TBX5* locus that influence CCS function.[16,17] *Tbx3* and *Tbx5* are required for functional development, maturation and homeostasis of the CCS in a highly dosage-sensitive manner, indicating that minor changes in regulatory sequences could potentially have consequences for CCS function and homeostasis.[12,41] Based on the contact profile of the *Tbx3* and *Tbx5* promoters and their separate regulatory domains, we propose that all the regulatory elements mediating *Tbx3* or *Tbx5* expression, respectively, are confined within their respective domains marked by CTCF binding sites. This indicates that the functional effect of genetic variation is limited to either the *TBX3* or *TBX5* domain, respectively, and within those domains to the regions contacting the respective promoters. Our data therefore facilitates the assignment of function to genomic variants identified in GWAS.

Identification of enhancers driving CCS-specific Tbx3 expression is crucial if one is to fully understand the mechanisms underlying CCS development and function. We report the identification of two enhancers within the *Tbx3* domain that synergistically drive the AV conduction system expression of *Tbx3*. Major components of the *Tbx3*$^{+}$ AV conduction system are the slow-conducting AV node and rings, and the fast-conducting AV bundle and bundle branches. These components derive from the embryonic AV canal and crest of the interventricular septum, respectively, which are interconnected throughout development.[42] The AV canal component is evolutionarily conserved in vertebrates.[43] Our data reveal that the

regulatory sequences regulating *Tbx3* in these respective components are separated in the genome, and suggest that eA is responsible for driving *Tbx3* expression in the conserved AVC component through a BMP signaling-mediated mechanism. Although on its own weak, eB appears to synergize with eA to drive robust *Tbx3* expression in the AVC and AV bundle precursor region in the interventricular septum. We speculate that eB may have been recruited during the evolution of the AV bundle, a structure, as yet, only found in birds and mammals. These elements may come together to form a single regulatory unit, leading to an expression pattern that is wider than the sum of each isolated element alone.[44,45] The contribution of presumably weak elements such as eB could therefore be to provide robustness to the complex regulatory structure.

Taken together, our data show that the *Tbx3* locus forms a conserved, independent regulatory unit physically separated from neighboring loci. The chromosomal topology of this unit is in a preformed and permissive organization, which is tissue independent and evolutionary conserved. Multiple sites within this unit synergize to control the precise expression of *Tbx3* in the CCS. This study provides insight into the tight relation between chromatin topology and the strict regulation of the evolutionary conserved *Tbx3/Tbx5* gene cluster.

MATERIALS AND METHODS

Circular chromosome conformation capture

4C was performed as previously described.[20] In short, fore- and midbrain, limbs and heart were dissected from E10.5 embryos and processed as described to obtain single-cell samples. Approximately 10^7 cells were used for fixation and lysis, and cross-linked chromatin was digested with DpnII (New England Biolabs, cat.no. R0543M). Samples were treated with T4 DNA ligase (Roche, cat.no. 10799009001) to obtain large DNA circular molecules, encompassing multiple restriction fragments held together by cross-links. After cross-link removal by heating, DNA was digested with Csp6I (Fermentas, cat.no. ER0211). Final ligation of restriction fragments yields circular DNA templates for 4C PCR amplification. For all experiments, 200 ng of the resulting 4C template was used for the subsequent PCR reaction, of which 16 (total: 3.2 µg of 4C template) were pooled and purified for next-generation sequencing. The PCR products were purified using two columns per sample of the High Pure PCR Product Purification Kit (Roche cat. no. 11732676001). Inverse PCR primers specific for fragment of interest (viewpoint) are listed in Supplemental Table 1. Primers were designed and data analysis for mapping and normalizing 4C-seq data was performed as described in ref. [20]. In short, the rich fragment pool (around 6-8 fragment ends per 1 kbp) was used to generate statistically robust semiquantitative contact maps in the 10 kbp to 1 Mbp region surrounding the viewpoint. Moreover, the high-resolution restriction site grid was used to quantify contact intensities in genomic windows varying in size from a few kilobases to tens of kilobases. Chromosomal contacts in human fibroblasts were assessed using Hi-C data generated by Jin *et al.* [40]

BAC modification and generation of transgenic mice

BACs RP24-89K7 and RP23-143N21 were obtained from C57BL/6J mouse BAC libraries (CHORI, BACPAC Resources) and modified to generate 89K7-GFP and 143N21-GFP. A GFP reporter gene cassette was inserted in the translational start site of Tbx3 according to the two-step BAC modification protocol as described in ref. [46]. Generation of the 366H17-GFP mouse line was described previously.[28] To obtain ΔeAeB 89K7-GFP, eA and eB enhancer sequences were sequentially deleted from GFP-89K7 using the abovementioned protocol with shuttle vectors in which both fragments were absent. Modified BAC DNA was purified with the NucleoBond PC 20 kit (Macherey-Nagel) according to manufacturer's instructions and injected into mouse pronuclei of FVB/N mice. Microinjected zygotes were implanted into pseudo-pregnant females (E0.5) and transgenic embryos were isolated in PBS at the indicated day, visualized by fluorescence microscopy, then fixed in 4% paraformaldehyde in PBS for 4 h and processed for *in situ* hybridization or immunohistochemistry. At least two independent transgenic lines per BAC construct were analyzed to define the GFP expression pattern and to account for variation in expression pattern due to position effect. The sequences of primers used for cloning and genotyping are available upon request. All animal work was approved by the Animal

Experimental Committee of the Academic Medical Center, University of Amsterdam, and carried out in compliance with Dutch government guidelines.

Ctcf conditional knockout

The *Ctcf* floxed allele and *Nkx2.5*-Cre line were previously described.[47,48] Mice were bred in the core animal facility in the Centro Nacional de Investigaciones Cardiovasculares in accordance with national and European legislation. Three pools of 2 to 4 wildtype and *Ctcf* KO hearts were used for qRT-PCR. 500 ng of total RNA was reverse transcribed using random primers (Applied Biosystems). Expression changes were normalized to actin. PCR was performed in an AB7900_384 (Applied Biosystems) using SYBR Green (Applied Biosystems) as reporter. qRT-PCR primers are available upon request. Homo- or heterozygous mice for the *Ctcf*-floxed allele that did not carry the *Nkx2.5*-Cre were used as controls.

In-situ hybridization

After fixation in 4% PFA in PBS, embryos were embedded in paraplast and sectioned at 10 µm. Non-radioactive *in situ* hybridization on sections was performed as previously described.[49,50] Probes for Tbx3 and GFP have been described in [28]. Sections were observed with a Zeiss Axiophot microscope and photographed with a Leica DFC320 Digital Camera.

Immunohistochemistry

After fixation, embryos were embedded in paraplast and sectioned at 8-10 µm for immunohistochemistry. Anti-Tbx3 (E-20, sc-31656) goat polyclonal (1:200; Santa Cruz Biotechnology Inc.), anti-GFP (ab5450) goat polyclonal (1:200; Abcam) primary antibodies and Alexa Fluor 488 donkey anti-goat IgG, Fluor 568 donkey anti-goat IgG or Fluor 680 donkey anti-goat IgG (1:250; Invitrogen) secondary antibodies were used. Stained sections were photographed with a fluorescence microscope Leica DM6000.

ChIP-sequencing

ChIP-sequencing data sets for Nkx2-5, Gata4, Tbx5 and Tbx3 were described in refs. [29, 31]. ChIP-seq datasets for p300, Pol2, CTCF and DNase hypersensitivity sites were generated by the ENCODE consortium.[30] ChIP-sequencing datasets for H3K27ac are described in ref. [26]. Peaks were called using the peak caller MACS, using corresponding input control tissue, and by an in-house peak calling algorithm dubbed Occupeak.[51]

Identification of consensus binding sites

Consensus binding sites were identified using JASPAR[52] and TRANSFAC[53] software. Sequences were tested for enrichment of binding sites for cardiac transcription factors Nkx2-5, Gata4, T-box factors and Bmp signaling effector Smads. Relative profile score threshold for JASPAR was

set to 85%. TRANSFAC parameters were set at the default setting for vertebrate sequences and the minFP setting to minimize false positive scores.

Chromatin immunoprecipitation

E14.5 embryos were dissected, and the hearts were isolated and collected in cold PBS. ChIP was performed using the True MicroChIP kit (Diagenode) according to the manufacturer's protocol. The antibodies used for ChIP were anti-H3K27ac (0.5µg, Abcam ab4729) and anti H3K4me1 (0.5µg, Abcam ab8895). Quantitative PCR was performed on a Roche LightCycler 480 System using SYBR Green detection. Fold enrichment indicates the ratio of ChIPed DNA to a negative control region, normalized for input DNA. Primer sequences are provided in Supplemental Table 2.

In vivo enhancer screening assay

Mouse eB (chr5:120,024,484-120,025,600; NCBI37/mm9), 3.4kbp-eB (chr5:120,024,225-120,027,612), eA (chr5:120,100,494-120,102,560), eAb (chr5:120,101,525-120,102,566), human eA (chr12:115,140,877-115,142,783; NCBI36/hg18) and chicken eA (chr15:12,509,382-12,511,893; WUGSC2.1/galGal3) enhancer sequences were PCR-amplified and cloned in the HindIII restriction site of the Hsp68-LacZ reporter vector.15 eA was subcloned into the SmaI site of eB-Hsp68-LacZ vector to obtain the eAeB-Hsp68-LacZ construct. After vector backbone excision, linearized fragments were injected into the pronucleus of 0.5-day-old fertilized FVB/N eggs, which were transferred into the oviducts of CD-1 pseudopregnant foster females (Cyagen Biosciences). Embryos were harvested and stained with X-gal to detect LacZ activity.

In vitro reporter assay

eA (chr5:120,100,494-120,102,560), eA1 (chr5:120,100,494-120,101,620), eA2 (chr5:120,101,526- 120,102,560), eB (chr5:120,024,484-120,025,600) and eAeB were cloned in the pGL2-SV40 luciferase construct containing a minimal promoter (Promega). $(Tbx2)_4$ construct was previously described.[26] The constructs were transfected in Cos-7 cells using PEI as transfection reagent (DNA:PEI ratio 1:4). Constructs were co-transfected with expression constructs for Gata4, Smad1, Smad4 (all cloned in pcDNA3.1) and constitutively active Bmp-receptor Alk3 (Alk3CA; pCS2.BmpR1a.CA). Cell extracts and luciferase assays were performed according to the manufacturer's protocol (Promega). Mean luciferase activities and standard deviations (s.d.) were plotted as fold activation compared to the empty pGL2-SV40 expression plasmid. Transfections were carried out in triplo.

Acknowledgements

We thank Corrie de Gier-de Vries, Vincent Wakker and Karel van Duijvenboden for their technical assistance, and Niels Galjart (Erasmus MC, Rotterdam) for providing the *Ctcf*-floxed mice. This work was supported by grants from the Rembrandt Institute of Cardiovascular Sciences, Academic Medical Center, Amsterdam (to V.M. Christoffels), the Netherlands Heart Foundation (NHS grant 2010B205, to V.M. Christoffels) and by the European Community's Seventh Framework Program contract ("CardioGeNet" 223463, to V.M. Christoffels). Work in the lab of Miguel Manzanares was funded by the Spanish Government (grant BFU2011-23083), the CNIC Translational Program (grant CNIC-08-2009) and the ProCNIC Foundation.

REFERENCES

1. Munshi NV. Gene regulatory networks in cardiac conduction system development. *Circ Res*. 2012;110:1525-1537.
2. Christoffels VM, Smits GJ, Kispert A and Moorman AF. Development of the pacemaker tissues of the heart. *Circ Res*. 2010;106:240-254.
3. Wolf CM and Berul CI. Inherited conduction system abnormalities--one group of diseases, many genes. *J Cardiovasc Electrophysiol*. 2006;17:446-455.
4. Naiche LA, Harrelson Z, Kelly RG and Papaioannou VE. T-Box Genes in Vertebrate Development. *Annu Rev Genet*. 2005;39:219-239.
5. Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soults J, Grayzel D, Kroumpouzou E, Traill TA, Leblanc-Straceski J, *et al.* Mutations in human TBX5 (corrected) cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet*. 1997;15:30-35.
6. Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis AR, Yi CH, Gebuhr T, Bullen PJ, Robson SC, Strachan T, *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet*. 1997;15:21-9.
7. Bamshad M, Lin RC, Law DJ, Watkins WS, Krakowiak PA, Moore ME, Franceschini P, Lala R, Holmes LB, Gebuhr TC, *et al.* Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat Genet*. 1997;16:311-315.
8. Bruneau BG, Nemer G, Schmitt JP, Charron F, Robitaille L, Caron S, Conner DA, Gessler M, Nemer M, Seidman CE and Seidman JG. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell*. 2001;106:709-721.
9. Arnolds DE, Liu F, Fahrenbach JP, Kim GH, Schillinger KJ, Smemo S, McNally EM, Nobrega MA, Patel VV and Moskowitz IP. TBX5 drives Scn5a expression to regulate cardiac conduction system function. *J Clin Invest*. 2012;122:2509-2518.
10. Hoogaars WM, Engel A, Brons JF, Verkerk AO, de Lange FJ, Wong LY, Bakker ML, Clout DE, Wakker V, Barnett P, *et al.* Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria. *Genes Dev*. 2007;21:1098-1112.
11. Bakker ML, Boink GJ, Boukens BJ, Verkerk AO, van den Boogaard M, den Haan AD, Hoogaars WM, Buermans HP, de Bakker JM, Seppen J, *et al.* T-box transcription factor TBX3 reprograms mature cardiac myocytes into pacemaker-like cells. *Cardiovasc Res*. 2012;94:439-449.
12. Frank DU, Carter KL, Thomas KR, Burr RM, Bakker ML, Coetzee WA, Tristani-Firouzi M, Bamshad MJ, Christoffels VM and Moon AM. Lethal arrhythmias in Tbx3-deficient mice reveal extreme dosage sensitivity of cardiac conduction system function and homeostasis. *Proc Natl Acad Sci USA*. 2011;109:E154-E163.
13. Sakabe NJ, Savic D and Nobrega MA. Transcriptional enhancers in development and disease. *Genome Biol*. 2012;13:238.
14. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190-1195.
15. Ward LD and Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol*. 2012;30:1095-1106.

16. Pfeufer A, van Noord C, Marciante KD, Arking DE, Larson MG, Smith AV, Tarasov KV, Muller M, Sotoodehnia N, Sinner MF, *et al.* Genome-wide association study of PR interval. *Nat Genet*. 2010;42:153-159.

17. Sotoodehnia N, Isaacs A, de Bakker PI, Dorr M, Newton-Cheh C, Nolte IM, van der Harst P, Muller M, Eijgelsheim M, Alonso A, *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet*. 2010;42:1068-1076.

18. Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC and Nobrega MA. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum Mol Genet*. 2012;21:3255-3263.

19. de Laat W and Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*. 2013;502:499-506.

20. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods*. 2012;9:969-972.

21. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289-293.

22. Stamatoyannopoulos JA, Snyder M, Hardison R, Ren B, Gingeras T, Gilbert DM, Groudine M, Bender M, Kaul R, Canfield T, *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol*. 2012;13:418.

23. Phillips JE and Corces VG. CTCF: master weaver of the genome. *Cell*. 2009;137:1194-211.

24. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N and de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 2006;20:2349-2354.

25. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA*. 2010;107:21931-21936.

26. Stefanovic S, Barnett P, van Duijvenboden K, Weber D, Gessler M and Christoffels VM. GATA-dependent regulatory switches establish atrioventricular canal specificity during heart development. *Nat Commun*. 2014;5:3680.

27. Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG and Higgs DR. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J*. 2007;26:2041-2051.

28. Horsthuis T, Buermans HP, Brons JF, Verkerk AO, Bakker ML, Wakker V, Clout DE, Moorman AF, 't Hoen PA and Christoffels VM. Gene expression profiling of the forming atrioventricular node using a novel Tbx3-based node-specific transgenic reporter. *Circ Res*. 2009;105:61-69.

29. He A, Kong SW, Ma Q and Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci USA*. 2011;108:5632-5637.

30. Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, Cline MS, Karolchik D, Barber GP, Clawson H, *et al.* ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res*. 2012;40:D912-D917.

31. van den Boogaard M, Wong LY, Tessadori F, Bakker ML, Dreizehnter LK, Wakker V, Bezzina CR, 't Hoen PA, Bakkers J, Barnett P and Christoffels VM. Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J Clin Invest*. 2012;122:2519-2530.

32. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009;457:854-858.

33. Dai YS and Markham BE. p300 Functions as a coactivator of transcription factor GATA-4. *J Biol Chem*. 2001;276:37178-37185.

34. Agulnik SI, Garvey N, Hancock S, Ruvinsky I, Chapman DL, Agulnik I, Bollag R, Papaioannou V and Silver LM. Evolution of mouse T-box genes by tandem duplication and cluster dispersion. *Genetics*. 1996;144:249-254.

35. Tena JJ, Alonso ME, Calle-Mustienes E, Splinter E, de Laat W, Manzanares M and Gomez-Skarmeta JL. An evolutionarily conserved three-dimensional structure in the vertebrate Irx clusters facilitates enhancer sharing and coregulation. *Nat Commun*. 2011;2:310.

36. Duboule D. Vertebrate *hox* gene regulation: clustering and/or colinearity? *Curr Opin Genet Dev*. 1998;8:514-518.

37. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013;153:1281-1295.

38. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS and Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376-380.

39. Soshnikova N, Montavon T, Leleu M, Galjart N and Duboule D. Functional analysis of CTCF during mammalian limb development. *Dev Cell*. 2010;19:819-830.

40. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA and Ren B. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013.

41. Mori AD, Zhu Y, Vahora I, Nieman B, Koshiba-Takeuchi K, Davidson L, Pizard A, Seidman CE, Seidman JG, Chen XJ, Henkelman RM and Bruneau BG. Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. *Dev Biol*. 2006;297:566-586.

42. Aanhaanen WT, Mommersteeg MT, Norden J, Wakker V, de Gier-de Vries C, Anderson RH, Kispert A, Moorman AF and Christoffels VM. Developmental origin, growth, and three-dimensional architecture of the atrioventricular conduction axis of the mouse heart. *Circ Res*. 2010;107:728-736.

43. Jensen B, Boukens BJ, Postma AV, Gunst QD, van den Hoff MJ, Moorman AF, Wang T and Christoffels VM. Identifying the evolutionary building blocks of the cardiac conduction system. *PLoS One*. 2012;7:e44231.

44. Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F and Duboule D. A regulatory archipelago controls Hox genes transcription in digits. *Cell*. 2011;147:1132-45.

45. Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H and Shiroishi T. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell*. 2009;16:47-57.

46. Gong S, Yang XW, Li C and Heintz N. Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res*. 2002;12:1992-1998.

47. Heath H, Ribeiro de Almeida C, Sleutels F, Dingjan G, van de Nobelen S, Jonkers I, Ling KW, Gribnau J, Renkawitz R, Grosveld F, Hendriks RW and Galjart N. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J*. 2008;27:2839-2850.

48. Stanley EG, Biben C, Elefanty A, Barnett L, Koentgen F, Robb L and Harvey RP. Efficient Cre-mediated deletion in cardiac progenitor cells conferred by a 3'UTR-ires-Cre allele of the homeobox gene Nkx2-5. *Int J Dev Biol*. 2002;46:431-439.

49. Moorman AFM, Houweling AC, de Boer PAJ and Christoffels VM. Sensitive nonradioactive detection of mRNA in tissue sections: novel application of the whole-mount in situ hybridization protocol. *J Histochem Cytochem*. 2001;49:1-8.

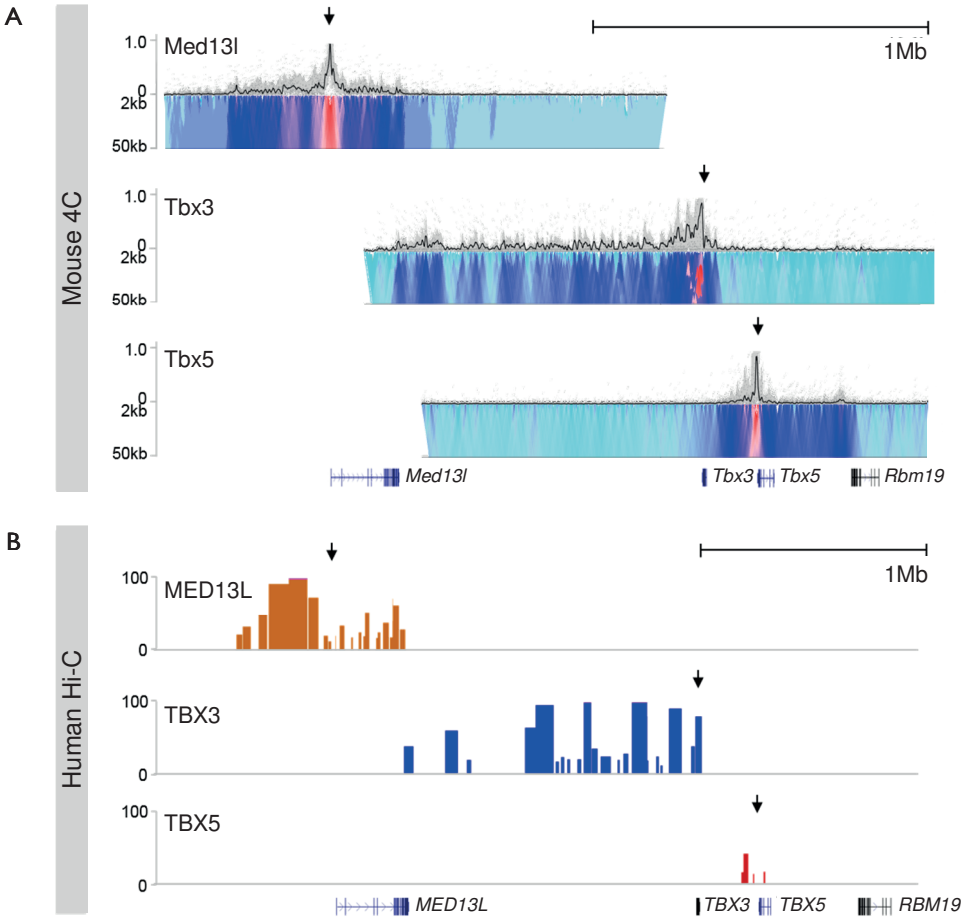
50. Luna-Zurita L, Prados B, Grego-Bessa J, Luxan G, Del MG, Benguria A, Adams RH, Perez-Pomares JM and de la Pompa JL. Integration of a Notch-dependent mesenchymal gene program and Bmp2-driven cell invasiveness regulates murine cardiac valve formation. *J Clin Invest*. 2010;120:3493-3507.

51. de Boer BA, van Duijvenboden K, van den Boogaard M, Christoffels VM, Barnett P and Ruijter JM. OccuPeak: ChIP-seq peak calling based on internal background modelling. *PLoS One*. 2014;9:e99844.

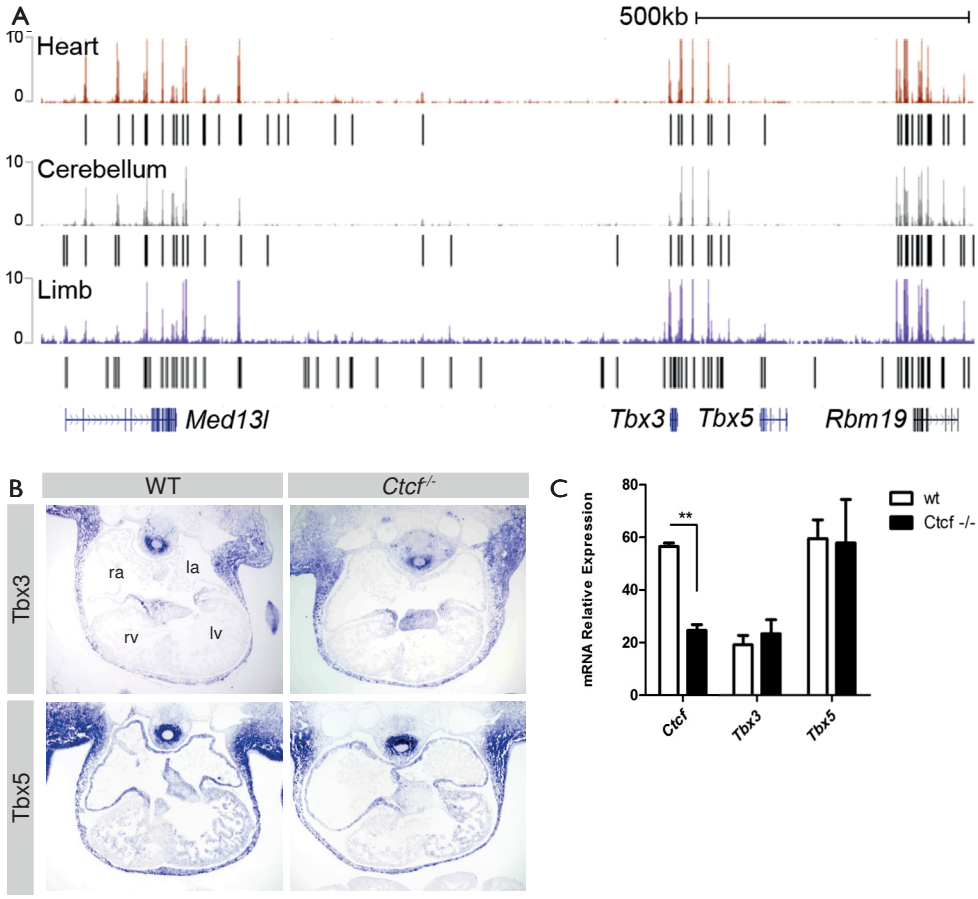
52. Wasserman WW and Sandelin A. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet*. 2004;5:276-287.

53. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*. 2003;31:374-378.

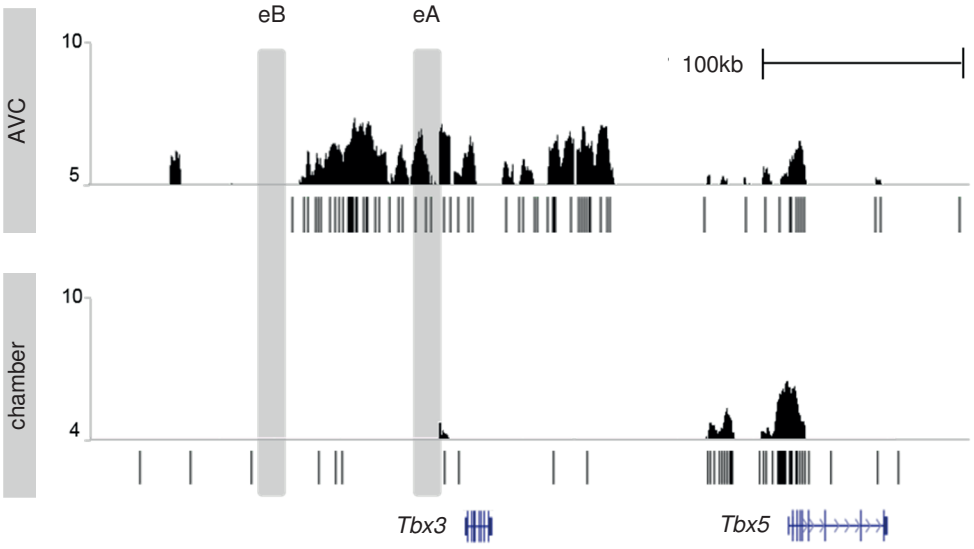
SUPPLEMENTAL FIGURES



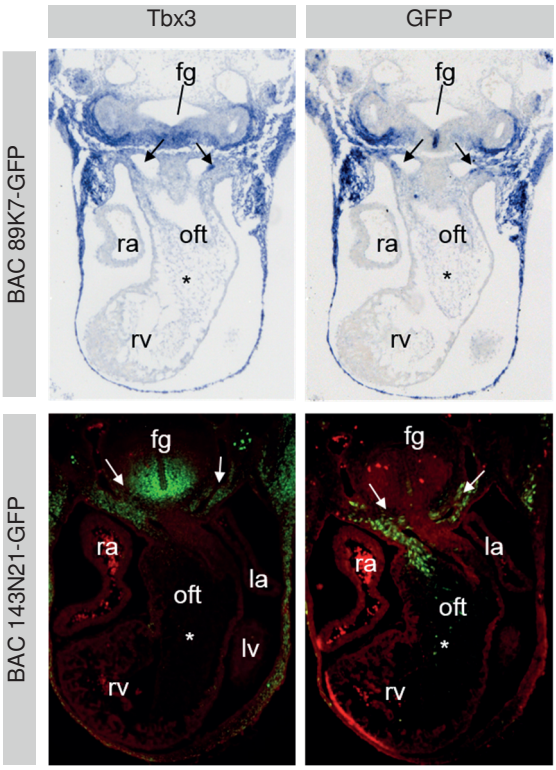
Supplemental Figure 1. Conservation of topological *Tbx3/ Tbx5* domains. Contact profiles with *Med13l*, *Tbx3* and *Tbx5* promoters as viewpoints in mouse heart (upper tracks, 4C data) and human fibroblasts (lower tracks, Hi-C data generated by Jin *et al.* [40]). Arrows depict point of view for each track. Similar to the mouse genome, contacts with TBX3 in human are restricted to the upstream domain and do not overlap with regions contacting *MED13L* (upstream) and *TBX5* (downstream).



Supplemental Figure 2. cKO of *Ctcf* in cardiomyocytes does not result in altered gene expression. (**A**) UCSC browser views of ChIP-seq data (ENCODE consortium (Rosenbloom *et al.*, 2012)) depicting occupancy of CTCF in heart, cerebellum and limb (E14.5). Lower traces depict peaks called by MACS, indicating the high similarity of occupancy profiles between the three tissue types. (**B**) *In-situ* hybridization on sections through WT and *Ctcf*$^{-/-}$ embryos at E10.5 reveal no misexpression of *Tbx3* and *Tbx5* in the heart. (**C**) qPCR validation of *Ctcf*, *Tbx3* and *Tbx5* mRNA levels. *Ctcf* mRNA levels are significantly reduced in *Ctcf*$^{-/-}$ embryos, whereas *Tbx3* and *Tbx5* levels were unaffected.



Supplemental Figure 3. Cell-type specific H3K27 acetylation associates with gene activity. Occupancy profiles of H3K27ac in the *Tbx3* and *Tbx5* locus as assessed by ChIP-seq in AVC and chamber myocardium. Y-axis in upper tracks depicts normalized tag count. Lower tracks depict peaks called by our in-house peak calling algorithm called OccuPeak [51].



Supplemental Figure 4. GFP-modified BACs recapitulate *Tbx3* expression in neural crest-derived cell. *In-situ* hybridization and immunohistochemistry on sections through 89K7-GFP, 143N21-GFP and 366H17-GFP (not shown) embryos at E11.5 reveal GFP expression in cardiac neural crest cells flanking the outflow tract (arrows) and the outflow tract cushions (neural crest-derived cells; asterisks).

>Tbx3_eA (2067 bp; chr5:120,100,494-120,102,560 [mm9])

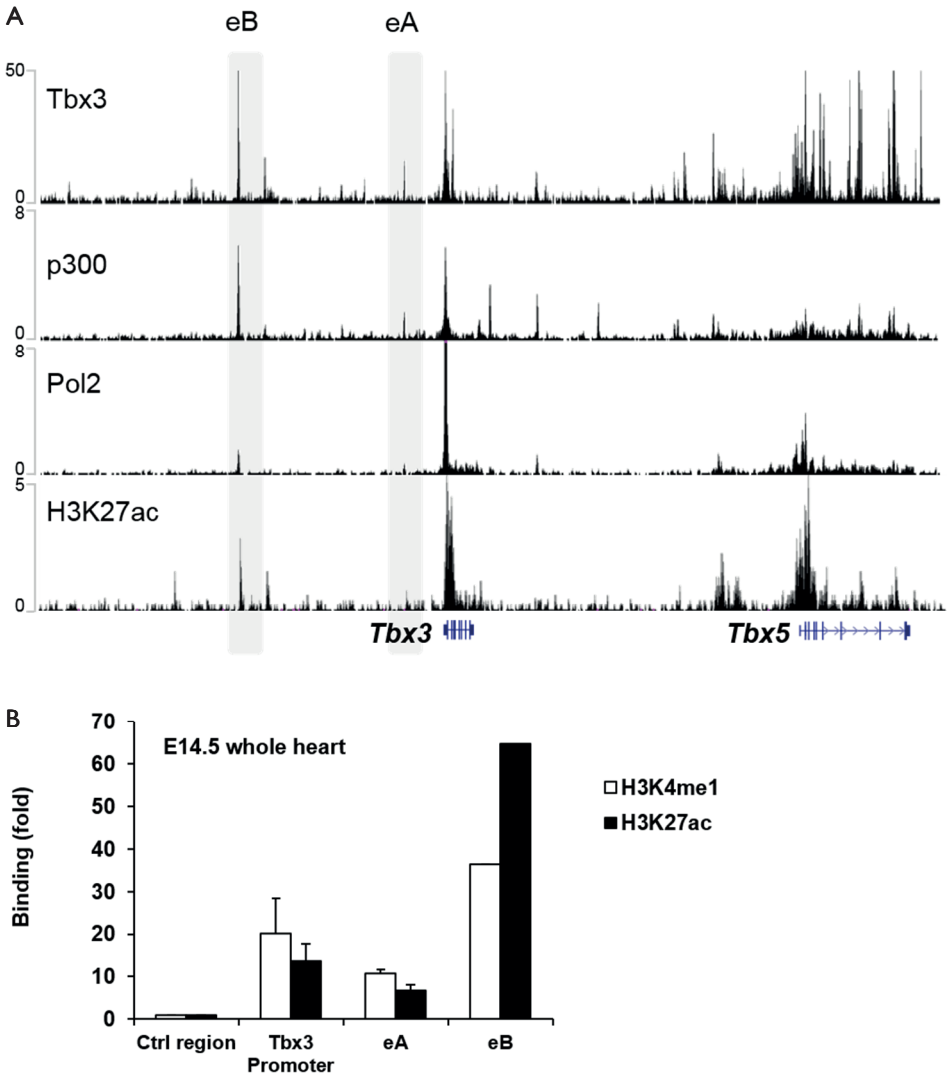
GCACTACCACTGCCCTGCACATTTTAAATATTTATTACTTAAAATATTTATTCATTCATTTATGAGGGGTGGTAGGAGGG
GGAGAAAGAGATGTGGGGGGAGGGGGAGAGAGATGTGAGGGGAGGGAAGAGAGAGATGTGGGGGGAGGGTGCTTCAGCAC
ATCTGTGGAACTCAGAAGATGACTTTGTAGAACAAATTCTTTCCTTCCACCTTTACGGGGTCCTAGGGATTGAACTCTGG
CTGTCTCGATTGTGTGGCAAGCTGCTTCATCTTTCGAGCCATCCCACAGGTCCAAGAAGGGTATTTTAAAAAATTATTTA
TTTTGTGCATGTGTGCACACATGTAGGACACATGGGGCAGTCAGAGGACAATCCCAGGAGTCAGTTCTTTCCTTCCAACG
CGTGAGTCCCAAAGAAGACACATGGGACATTAGGTTCGACTGTAAATGTCTTTAACCGGCTTCCAAGAAAGGTATTTTAA
ATGAGAATCTTTAATGTAGCAAGAAAAAGATTCTTACATTATTTCTTCTTCCTCTGTCTCTATCTCTGTCTCTGTCTCTC
TCTCTCTCTCTGTCTCTCTCTCACCCCTCCCCTCTTTCTCTCAAATACTTATATTTGCTTTGTTTTCATAAACAGCTGAT
AGTTTGTTTTAAAAGAACAGAGGGAGGGAGGGGTGCCTACAGGGCAGAGAGAGAGAGAGGTGGAGTGTGAGTTGTACAAG
CGCTCGCTCGCTCCCGCATTCAGGCCTCACTAGCAAGTAATAAACAAGGACCTGGTTCAGGTACAGACAGCTGGGAGGGG
GATAGGGAGGAATGCGAAAGCAGATTTACACCTGTCCTCACAGTTGCTATGAGGACGCTTAAGGGCTCTAGAAATGACCC
TTGGTGCCGACTTGTCTGGAAGCCATCAGCCTGAGTTTTGAAGCAGGAATCTTCAAGTGGGCAGCAGGAGAGTAAGCTCA
GCCCCACCCCCATCCCCCAGGCCCCAGGCCTGCCCTCCCCCCCCCACCACCATGAGGTTCCACAGAGCTGAACCTCC
CACAAGGCAAGGTAGTCCAGTCTCCTATCCAGCCCTCTCCTTTTCTCTCAGTTGATAGCAGCACCCGTTTTCACTCCTGC
TGAGGACTGGGCTCTGGGCAGGATTAGCAGCCGGCTCCTAAACCACCCTTGTCTTCCCTCAGGAGCTCTGTCTGGCCCCA
GTCCAGCCAGCTCCTTCCCCTACCATCAGCCGCCCTGGCCCACTGTCCAGACAGATGAGTGACTGACAATACACAGACCT
CCCTGGAATCCTGCTTCTGGTCACCACACACCGCTGGCCTGCAGGAGCCGGAGCCCAGGCAGCGCTGCCCTGCCTGAAAC
TTGTTACCACCTCTCAGAGGCTGTTGAAAGCCCCACTCAGCTGAGCCCTTCAGGATGGGGACCCTGGAGCCAGGGCACTC
AGAAGGGATGTTTGACACCTGCCCTGGGGACTCTGAGTCCCTGACTCTCTGCCCCTGGCCCCATAGAGGTCTAGGAGGGT
TGAAAGTGGGTAGATTACAAACTGTGCAGATACCAGAGAGGACATGGTGGGGAGACCTAATACTTTGCAGGGATCAGAAG
AGGATGCTGGCCCAGCCTCCATTGCCTTATTGATAAGATGAGGAGGTGGTAGAAAGACCATGTTGTGTGTACTATGAATG
TTTCAGTTGCTCAGAGCATTAAAAAAAGTATGTGTGTGTCTCTCAGTTGTTAGAACATTTACATATATATATACATACAT
ATATATATATATACACATATACATATACATACATATATATATGCATGCTCTGTTTTGTGTGTGTGTGTATGTCTGTGCAC
ATATCACATGTGTAAGTGACTGCCTGTAGAAGCCAGAGAGGGCATAAGATCCCTTGGAGCTGAATATATAAGCAGCTATG
AGCTGGTCTTTGTAGGTGCTGGGATCTGAATTTGGGTCCCCTGGAAGAGTAGCCATTGCTCTCAACCGTTGACCCATCTC
TCTAGCCCCCTGCACACCTTTAAATAACCTACTCATTTTATCAAAACACATCATGGTTGACCAAGA

>Tbx3_eB (1117 bp; chr5:120,024,484-120,025,600 [mm9])

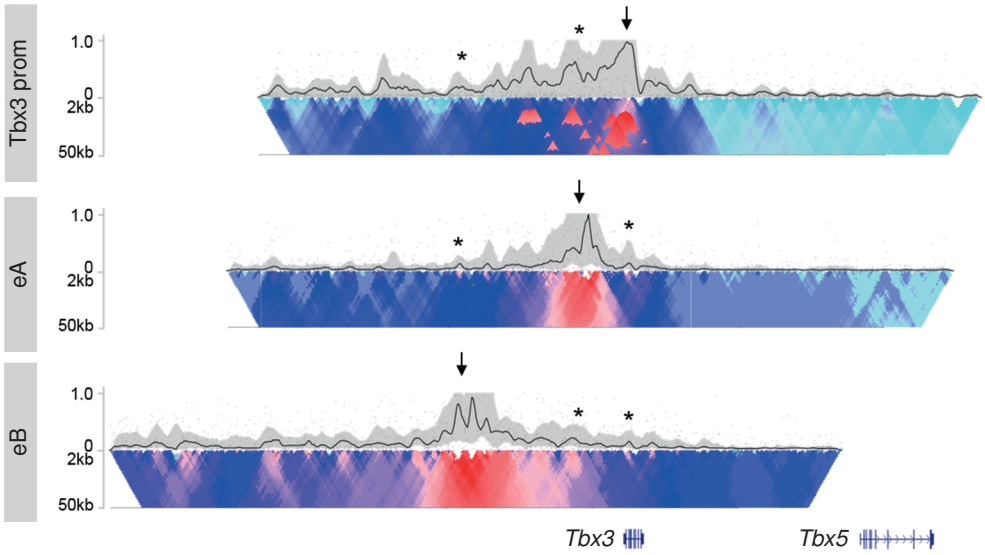
CACCCAATCTGAGAGCTTTTGCTACAGTGGCCCTATGTCACACTGACCTTACAGAACGAGAGCCCGGAAGGTAGTGGTGT
ATGCTGGGCCTTAGGCATTCAGGAGAGTAGTCAGATGGAGGTCTGTCTGGACTGGACTGAGACCATGTCTCAAACACGAT
GATGATGATGGTGGTGGTGGTGGTGATGATGATGATGGTGGTGATGATGATGGTGGTGGTGGTGGTGGTGGTGG
TGGTGATTAATCACTAACAGAGCTGAGATATCGTCTCAGAGGCCCCCTCTCCTCAGACAGCCTCTTTTCTGCTTCCACCC
AGTTTGTACAATGGTTCTCTAGAGGTTTAGTCTAGAGGAGGAGGTGTCCATAGTTATAAAAATCCTCTATATTTTACCAG
ATAGGGGGTGGAAACCGAGTCTGAGGAGAAACTTGCCAGCCATACACTGGGCTGGCCAATAAATCAGCAGCCAGGACTCT
TTATCTGTGGTGAAGGCCCAGGCCCATGAGGACATGAAGTAAACACCCCGTATGCTGTAGGCACCACTTGAATGCTTCAT
TTCTCAGCCAAGACCCTGCCCCCATCTCCCTCCAGCCCCCCCATCCATCACCGCTGCAGAGTAAAACTGGGTACAATTCC
TACATCCAGCACCCATGCTTCAAGGCCAGCTGGGCATGAGGAGAAGCCTGCCCCTCCCCGAACTGACCTTCTGAAATGCC
CTTTTCCCCAGTGAGTGAGGGCTGGCAGAGCAGGCCAGGCTCGGGGTTAGGCCTTGATAACCCGAGATCAAAGCGGCCTG
TCAGAGGTGTCAACTGAGAGGCAGCCAAGACAAGCGACACACACGGGCAGCTTTATCTCCCCGCTGGCTGGCGGGGAAGC
ACCTCGGTTCCCGGAAACTATTTCTGGACAGGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAACG
CGTCTATGGCTGAGACTGCTGGTCATGCTAAGCCAGTCCAAATCTGTTCAGGGACTCTGTGCGCCAGACGAGGGCCCTTT
CTCTCTAGGCTCTCTCCTCATTCATCCCCAAGTTCCTGTCTTCACAGGTGCCTGAGCACTGCGCAACAGAAGTTTGT

Consensus binding sites for Gata4, Nkx2-5, Tbx3/5 and Smad identified by JASPAR[52] and TRANSFAC.[53]

Supplemental Figure 5. Overview of transcription factor binding sites in eA and eB. Enrichment of binding sites for cardiac transcription factors Nkx2-5, Gata4, T-box factors and Bmp signaling effector Smads as revealed by JASPAR [52] and TRANSFAC [53] software.



Supplemental Figure 6. Transcription factor occupancy and histone modifications of eA and eB. (**A**) UCSC browser view of ChIP-sequencing datasets, revealing the occupancy of eA and eB by Tbx3, p300, Pol2 and H3K27ac. Y-axis depict normalized tag count. (**B**) ChIP-qPCR confirming eA and eB are marked by H3K4me1 and H3K27ac. Y-axis depicts fold activation over a negative control region. Bars represent mean±s.d. Experiments are performed in duplo.



Supplemental Figure 7. Physical interactions of *Tbx3*, eA and eB. Contact profiles of *Tbx3* promoter, eA and eB as viewpoints (arrows) reveal both enhancers and the *Tbx3* promoter physically contact each other (asterisks). The legend for the color-coded scaling plots can be found in the legend of Figure 1.

SUPPLEMENTAL TABLE 1. 4C-seq primer sequences.

Name	Sequence (5'—3')
Tbx3_promoter_F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGAATGTCCTGGCTGATTTGATC
Tbx3_promoter_R	CAAGCAGAAGACGGCATACGAGTGATGTGAGGGGTCAGAAA
eA_F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGACTGGATGTAGGTGTAAGATC
eA_R	CAAGCAGAAGACGGCATACGACAGAAGGCTCCTATGATGAT
eB_F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGACTTGAGACCAAACCCTGATC
eB_R	CAAGCAGAAGACGGCATACGAGCCTGGTATAAACTCCTCCT
Tbx5 promoter_F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCCCTGCCTTGGTGATGATC
Tbx5 promoter_R	CAAGCAGAAGACGGCATACGAGAGAGCCGGTAAAGTCTTTG
CTCF_F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGAATGAAGAGTTGGCAAAGATC
CTCF_R	CAAGCAGAAGACGGCATACGAAGAGAGTTTCTCCAACCATG
Rbm19 promoter_F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGAAAGTTCTTGCTGGGATC
Rbm19 promoter_R	CAAGCAGAAGACGGCATACGAAGGAATGAAGACACATGTCC
Med13l promoter_F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAGAAGAGAGAAAGTCAGATC
Med13l promoter_R	CAAGCAGAAGACGGCATACGACTTTGCCATATGAGGTCATT

SUPPLEMENTAL TABLE 2. qPCR primer sequences.

Name	Sequence (5'—3')
Tbx3_Promoter_F	CAGCACTCGACCTGTGAAAA
Tbx3_Promoter_R	ATTGGCTCTTTGACGCTTTC
Tbx3_eA_F	CAGCTGAGCCCTTCAGGAT
Tbx3_eA_R	GGTGTCAAACATCCCTTCTGA
Tbx3_eB_F	TCCCCGAACTGACCTTCTGA
Tbx3_eB_R	CGTGTGTGTCGCTTGTCTTG
negative control region 1*_F	TCCAGACACACTGAACAGCAC
negative control region 1_R	TCCCCGAGATCAAGTGCTAC
negative control region 2**_F	CCAGACACACTGAACAGCAC
negative control region 2_R	TGGCAGCAGGTTCTGGAATC

* (chr4: 147397477 – 147397580)
** (chr4:147397478 – 147397598)