Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2019 September 06.

Published in final edited form as:

Mol Cell. 2016 October 06; 64(1): 37–50. doi:10.1016/j.molcel.2016.08.010.

A G-rich motif in the lncRNA *Braveheart* interacts with a zinc finger transcription factor to specify the cardiovascular lineage

Zhihong Xue[1], **Scott Hennelly**[2,3], **Boryana Doyle**[4,5], **Arune A. Gulati**[1], **Irina V. Novikova**[2,6], **Karissa Y. Sanbonmatsu**[2,3], **Laurie A. Boyer**[1,*]

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

[2]Los Alamos National Laboratory, Theoretical Biology and Biophysics Group, Los Alamos, NM 87545

[3]New Mexico Consortium, Los Alamos, NM 87544

[4]Undergraduate Research Opportunities Program, Massachusetts Institute of Technology, Cambridge, MA 02139

[5]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139

[6]Pacific Northwest National Laboratory, Environmental Molecular Sciences Laboratory, Richmond, WA 99354

SUMMARY

Long non-coding RNAs (lncRNAs) are an emerging class of transcripts that can modulate gene expression; however, their mechanisms of action remain poorly understood. Here, we experimentally determine the secondary structure of *Braveheart* (*Bvht*) using chemical probing methods and show this ~590nt transcript has a modular fold. Using CRISPR/Cas9-mediated editing of mouse embryonic stem cells, we find that deletion of 11 nucleotides in a 5' asymmetric G-rich internal loop (AGIL) of *Bvht* (*bvht*^{dAGIL}) dramatically impairs cardiomyocyte differentiation. We demonstrate a specific interaction between this motif and cellular nucleic acid binding protein (CNBP/ZNF9), a zinc finger protein known to bind single-stranded G-rich sequences. We further show that CNBP deletion partially rescues the *bvht*^{dAGIL} mutant phenotype by restoring differentiation capacity. Together, our work shows that *Bvht* functions with CNBP through a well-defined RNA motif to regulate cardiovascular lineage commitment, opening the door for exploring broader roles of RNA structure in development and disease.

Keywords

Braveheart; cardiac; CNBP; long non-coding RNA; SHAPE

*Corresponding author: Laurie A. Boyer, lboyer@mit.edu, Tel: 617 324-3335.

AUTHOR CONTRIBUTIONS

S.P.H. designed and performed chemical probing experiments; K.Y.S., S.P.H., I.V.N. analyzed the chemical probing data and determined the structures. Z.X. and L.A.B. designed all other experiments and interpreted the results. Z.X., B.D., and A.A.G. performed these experiments. L.A.B. and Z.X. wrote the manuscript.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

INTRODUCTION

Long non-coding RNAs (lncRNAs) have emerged as important regulators of development and disease. These transcripts are typically >200 nucleotides in length and are often polyadenylated, capped, and alternatively spliced but lack coding potential (Ulitsky and Bartel, 2013). Although biochemical and biophysical studies of lncRNAs are in their early stages, proposed mechanisms of action include chromatin scaffolding, Polycomb complex (PRC2) recruitment to chromatin, mRNA decay, and decoys for proteins and miRNAs (Geisler and Coller, 2013; Quinn and Chang, 2015). Studies have highlighted diverse cellular roles for lncRNAs across eukaryotes such as X chromosome inactivation, genomic imprinting, cell-cycle regulation, embryonic stem cell (ESC) pluripotency, and lineage commitment (Flynn and Chang, 2014; Lee and Bartolomei, 2013). In metazoans, there are a growing number of lncRNAs that function in lineage commitment and differentiation with key examples in the cardiovascular system (Grote et al., 2013; Han et al., 2014; Klattenhoff et al., 2013) including many that show differential expression in cardiac disease (Fatica and Bozzoni, 2014; Rizki and Boyer, 2015). Thus, it remains a critical goal to understand how long non-coding transcripts contribute to regulation of cell fate and disease.

Comparative sequence analysis has facilitated RNA secondary structure predictions and has helped to reveal the functions of ribonuclease P and riboswitches (Gutell et al., 2002; Mian, 1997; Parsch et al., 2000). These structural predictions are also experimentally supported by chemical probing methods (e.g. inline, SHAPE, DMS), NMR, and X-ray crystallography (Mondragon, 2013; Noller, 1984; Serganov and Patel, 2007). In contrast, predicting lncRNA secondary structure has been more complicated because these transcripts appear to be rapidly evolving and generally display low sequence conservation (Ponting et al., 2009). Recently, chemical probing methods have been exploited for studying lncRNA secondary structure. For example, Selective 2′ Hydroxyl Acylation analyzed by Primer Extension (SHAPE) probing of *in vitro* transcripts showed that the lncRNAs SRA and HOTAIR display a complex structural organization that comprise a variety of elements comparable to well-folded RNAs like group II introns and ribosomal RNAs (Novikova et al., 2012; Somarowthu et al., 2015). Genome-wide probing of RNA secondary structure using DMS-seq or icSHAPE-seq has also been performed in living cells, revealing active unfolding of mRNA structures suggesting RNA structures contribute to global RNA processing and translation (Ding et al., 2014; Rouskin et al., 2014; Spitale et al., 2015). Most lncRNAs, however, are not sufficiently abundant for detection *in vivo* and *in vivo* secondary structure studies can be obfuscated in the cell by the binding of proteins to RNA. Overall, detailed analysis of the native structure of individual lncRNAs is still largely lacking and is necessary to gain deeper insights into their precise roles.

Our prior work identified the mouse lncRNA *Braveheart* (*Bvht*) that appears to act *in trans* to regulate cardiovascular lineage commitment (Klattenhoff et al., 2013). Given that lncRNAs are generally lowly conserved by sequence and that many of these transcripts are species specific (Johnsson et al., 2014; Ponting et al., 2009), RNA secondary structure is key for understanding their broader roles. To investigate the molecular mechanism of *Bvht* action, here, we determined the secondary structure of full-length *Bvht* (~590 nt) using SHAPE and DMS probing *in vitro* and find that the transcript is organized into a highly modular

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

structure including a 5’ asymmetric G-rich internal loop (AGIL). Using CRISPR/Cas9-mediated homology directed repair, we deleted this loop (denoted *bvht*^{dAGIL}) in mouse ESCs and show that the AGIL motif is necessary for cardiomyocyte (CM) differentiation. Similar to shRNA-mediated *bvht* depletion, key cardiac transcription factors fail to activate during the transition from nascent mesoderm to the cardiac progenitor state. Using a protein microarray platform, we demonstrate that the AGIL motif interacts with a small subset of factors including the heart-expressed zinc finger transcription factor cellular nucleic acid binding protein (CNBP/ZNF9) known to bind G-rich single-stranded nucleic acids (Calcaterra et al., 2010; Chen et al., 2007). Finally, we find that CNBP represses cardiomyocyte differentiation and that loss of CNBP partially rescues the *bvht*^{dAGIL} phenotype, suggesting that these factors function together to specify the cardiovascular lineage. Our results show how a small RNA motif in *Bvht* can direct cell fate and demonstrate that structural studies combined with genetic perturbation can provide critical insights into lncRNA function.

RESULTS

Braveheart is organized into a highly modular structure

RNA can form complex structures that have catalytic activity or that act as scaffolds for the binding of metal ions, small molecules, nucleic acids, and proteins (Mondragon, 2013; Noller, 1984; Serganov and Patel, 2007). To obtain the secondary structure of *Bvht*, we used the shotgun secondary structure determination strategy (3S) (Novikova et al., 2013), with the goal of obtaining more detailed mechanistic insight into *Bvht* function. Here, we performed SHAPE probing on *in vitro* transcribed full-length *Bvht* (Figure S1). We also performed DMS (dimethylsulfate) probing of full-length *Bvht* which methylates unpaired adenosine and cytidine nucleotides (Tijerina et al., 2007) (Figure 1A, lower panel). We next repeated the SHAPE and DMS probing on shorter fragments to identify sub-domains of *Bvht*. When a region’s reactivity in shorter fragments shows similarity to the profile in the full-length RNA, it suggests that this region adopts a modular fold in the context of full-length RNA structure. As shown in Figure 1B, we generated overlapping fragments and performed SHAPE probing as above. Detailed comparisons between each fragment and the full-length transcript revealed several regions of similar reactivity (Figure 1B). For example, the ~55 nt stretch at the 3’-end of *Bvht* exhibited high reactivity using both SHAPE and DMS probing indicating a low probability of being structured and was left out of the analysis. We obtain the fold for *Bvht* by piecing together the modular sub-folds.

The overall secondary structure shown in Figures 1C is most consistent with all of our experimental data based on SHAPE and DMS analysis of full-length *Bvht* and of the shorter fragments. *Bvht* consists of 12 helices, 8 terminal loops, 5 sizeable (>5 nts) internal loops and a 5-way junction (5WJ). *Bvht* appears to be organized into three domains, roughly corresponding to its three exons: the 5’-domain (H1-H2), central domain (H3-H8) and 3’-domain (H9-H12) (Figure 1C). The 5’-domain contains an asymmetric G-rich internal loop structure (AGIL) between H1 and H2, consisting of a large single-stranded region (14 nts) on the 5’ side and very short single-stranded region (3 nts) on the 3’ side. The central

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

domain consists of a 5-way junction (H4, H5, H6, H7, H8) connected to the 5'-domain by H3. The 3'-domain contains 4 helices (H9, H10, H11, H12).

Braveheart AGIL motif is necessary for proper ESC differentiation

To date, lncRNA function has largely been determined by transcript knockdown or by genetic deletion of large regions that may encompass regulatory elements confounding phenotypic interpretation. We focused on dissecting the function of the AGIL region because it appeared to be less commonly represented in known RNA secondary structure databases and because G-rich regions often play regulatory roles in the genome (Aguilera and Garcia-Muse, 2012; Rhodes and Lipps, 2015). For example, after searching the Gutell database of secondary structures of ribosomal and RNase P RNAs (Cannone et al., 2002), we found that only 13 of >400,000 asymmetric 5' internal loops had similar size and asymmetry. The crystal structure of one such loop was recently solved, forming an intricate tightly packed configuration of purines (Ren, et al., 2016). Thus, using CRISPR/Cas9 homology directed repair (HDR), we generated an 11-nucleotide deletion in AGIL (denoted *bvht*^{dAGIL}) at the endogenous *Bvht* locus in mESCs to disrupt this loop (Figure 2A, B). We used a dual selection strategy to facilitate recovery of homozygous clones (~20–50% frequency) and expanded several clones for experimental evaluation.

SHAPE probing of *bvht*^{dAGIL} RNA shows deletion of the AGIL motif does not destabilize overall *Bvht* structure (Figures 2C and S2A). Thus, we next examined *Bvht* levels in ESCs and found that the mutant transcript was expressed at comparable levels to wild-type by Northern blot and qRT-PCR (Figure 2D, E). Similar to shRNA-mediated depletion of *Bvht* (Klattenhoff et al., 2013), *bvht*^{dAGIL} did not affect expression of pluripotency markers such as Oct4 and Nanog and mutant ESCs showed normal morphology and self-renewal properties as well as typical cell cycle kinetics (Figures 2E,F and S2B). We then tested whether *bvht*^{dAGIL} could form embryoid bodies (EBs), which give rise to derivatives of all three germ layers. Notably, cardiomyocytes can form in EBs and can be visualized as beating cell clusters. We allowed wild-type and mutant ESCs to aggregate in the absence of pluripotency growth factors and then measured the percentage of spontaneously beating EBs at different time points. We found that *bvht*^{dAGIL} EBs show significantly reduced beating (~5%) compared to wild-type cells (~25%) at Day 10, similar to our observations in *Bvht*-depleted EBs (Klattenhoff et al., 2013).

Helical junctions are often important for the structural and catalytic properties of RNAs (Bindewald et al., 2008). For example, a four-way junction promotes the functional folded state of the hairpin ribozyme (Tan et al., 2003). Thus, we also introduced mismatch mutations into the H4 region (*bvht*^{H4mis}) to destabilize the 5-way junction (5WJ) by CRISPR/Cas9-mediated HDR and selected clones (Figure S2C–E). In contrast to *bvht*^{dAGIL}, alteration of the H4 region did not significantly affect the percentage of beating EBs (Figure 2G). *Bvht*^{dAGIL} EBs also displayed a failure to activate genes associated with the cardiac contractile apparatus such as cardiac troponin T (cTnT) and myosin heavy chain genes whereas *bvht*^{H4mis} EBs showed normal expression comparable to wild-type controls (Figure 2H). In contrast, neuronal and endodermal genes were expressed normally in *Bvht*^{dAGIL} EBs in response to retinoic acid treatment similar to wild-type and *bvht*^{H4mis} EBs (Figure

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

S2F,G). These data suggest that the *Bvht* AGIL motif is specifically necessary for formation of spontaneously contracting EBs. The results do not preclude a secondary role for the 5WJ.

Braveheart AGIL motif is necessary for cardiovascular lineage commitment

To further dissect AGIL function in the cardiovascular lineage, we employed a directed *in vitro* cardiomyocyte (CM) differentiation assay that permits isolation of cell populations at well-defined stages (ESCs, precardiac mesoderm (MES), cardiac progenitors (CPs), and (CMs)) (Kattman et al., 2011; Wamstad et al., 2012) (Figure 3A). At each stage, cells are subject to FACS using antibodies against specific markers to quantify differentiation efficiency. Using this approach, we routinely isolate a high percentage of Pdgfra+, Flk1+ (MES), Nkx2.5-GFP+ (CP), and cTnT+ (CM) cell populations (Figure 3B). In contrast, FACS of *bvht*^{dAGIL} cells showed a striking reduction in the percentage of CP and CM marked cells during differentiation. We also demonstrate that although *bvht*^{dAGIL} and wild-type cells showed similar morphology at Day 4 (MES), immunofluorescence of the cultures at Day 5.3 (CP) and Day 10 (CM) using antibodies against Nkx2.5-GFP or cTnT respectively, showed no staining in the mutant cells (Figure 3C). These results are highly reproducible among multiple independent *bvht*^{dAGIL} ESC clones and similar to shRNA depletion of *Bvht* (Figure S3A,B,E), suggesting that the differentiation defects are not due to off-target effects.

We next analyzed the expression of a set of cardiac transcription factors (TFs) that failed to activate upon shRNA-mediated depletion of *Bvht* (Figure S3C,D) (Klattenhoff et al., 2013). The mesodermal marker Brachyury showed higher expression at Day 4 in *bvht*^{dAGIL} cells and sustained expression at Day 5.3 compared to wild-type controls (Figure 3D). MesP1 is one of the earliest known markers of a common multi-potent cardiovascular progenitor (Bondue et al., 2008; Lindsley et al., 2008) and showed decreased expression at Day 4 (MES) in *bvht*-shRNA depleted cells (Klattenhoff et al., 2013). Although MesP1 expression showed no change in the *bvht*^{dAGIL} mutant, we observed a failure to activate the cardiac transcription factors downstream of this factor including Nkx2.5, Gata4, Gata6, Hand1, Hand2, Tbx5, and Mef2c compared to wild-type cells suggesting that distinct regions of *Bvht* contribute to its total activity (Figure 3E). These data are highly reproducible using multiple independent ESC clones (Figure S3F). Moreover, expression of wild-type *Bvht* from the ROSA26 locus in the *bvht*^{dAGIL} background (Figure S3G) rescued the cardiomyocyte differentiation defect indicating that the phenotype is due to loss of AGIL function (Figure S3H–K). Together, our data point to a central role for the *Bvht* AGIL motif in specifying the cardiovascular lineage.

Braveheart AGIL interacts with factors that bind G-rich nucleic acids

A prevailing model suggests that lncRNAs act as molecular scaffolds, mediating interactions with proteins (Geisler and Coller, 2013; Quinn and Chang, 2015; Rinn and Chang, 2012). Although genome-wide studies support binding between lncRNAs and proteins, few studies have identified RNA structural motifs responsible for these interactions (Chu et al., 2015). To identify proteins that potentially interact with the *Bvht* AGIL motif, we used a human protein microarray platform that has successfully identified lncRNA-binding proteins (Kretz et al., 2013; Siprashvili et al., 2012). Full-length *Bvht* and *bvht*^{dAGIL} transcripts were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

generated by *in vitro* transcription and labeled with Cy5 (Figure S4A and Table S2). Equal concentrations of labeled transcript were then individually incubated with the protein microarray containing approximately ~9,400 recombinant human proteins (Human ProtoArray). Using a stringent cut-off (z-score >3), we identified 12 candidates that strongly interacted with the wild-type transcript (Figures 4A and S4B and Table S2). Notably, four of these candidates (CNBP, HNRNPF, SFRS9, KCNAB2) showed dramatically decreased binding when the array was probed with the $bvht^{dAGIL}$ transcript (Figure 4B, C). These proteins are conserved between mouse and human and are highly expressed across the differentiation time course except for KCNAB2 (Figure S4C). We previously showed that *Bvht* interacts with the Polycomb Repressive Complex (PRC2) (Klattenhoff et al., 2013), however, we did not detect any change in the interaction with PRC2 in the mutant ESCs by RIP (data not shown) suggesting that this motif is not required for this interaction and that cooperation between *Bvht* and PRC2 may be a later event in regulating CM differentiation.

We next validated the interaction between *Bvht* and mouse CNBP, HNRNPF, and SFRS9 by expressing mouse Flag-tagged versions of these factors in both wild-type and $bvht^{dAGIL}$ ESCs followed by immunoprecipitation using an anti-Flag antibody (Figure 4D). We found that all three candidates co-purified with wild-type *Bvht* but not $bvht^{dAGIL}$ as shown by qRT-PCR. Upon analysis of ProtoArray results available for ~20 noncoding RNAs (Kretz et al., 2013; Marques Howarth et al., 2014; Siprashvili et al., 2012), CNBP and HNRNPF binding appeared to be highly specific to *Bvht* whereas SFRS9 interacted broadly with other noncoding RNAs. HNRNPF, a member of ubiquitously expressed heterogeneous nuclear ribonucleoproteins family, is a RNA-binding protein with roles in mRNA splicing, mRNA metabolism and transport and can bind G-rich sequences (Matunis et al., 1994; Reznik et al., 2014; Wang et al., 2012). CNBP (ZNF9) is a zinc finger transcription factor containing seven CCHC-type zinc fingers and one RNA recognition motif (RGG) (Figure 5A), that also binds G-rich single stranded DNA and RNA (Armas et al., 2008; Calcaterra et al., 2010). CNBP has roles in neural crest cell expansion and null mice die around E10.5 (Chen et al., 2003; Weiner et al., 2007; Weiner et al., 2011), however, its overall function is poorly characterized. Notably, CNBP is highly expressed in heart and skeletal muscle, and animals exhibit severe dilated cardiomyopathy in heterozygous $cnbp^{+/-}$ mice (Chen et al., 2007). Moreover, CNBP is currently the only known gene linked to myotonic dystrophy type 2 in human, and patients often display severe heart defects (Jones et al., 2011; Lee et al., 2012; Liquori et al., 2001). Thus, given its binding preference for single-stranded G-rich nucleic acids and its understudied roles in the heart, we focused on further characterization of CNBP.

CNBP represses cardiomyocyte differentiation

To test the function of CNBP in our system, we introduced small indels using CRISPR/Cas9 genome-editing in both wild-type and $bvht^{dAGIL}$ ESCs, generating $cnbp^{KO}$ and $cnbp^{KO};bvht^{dAGIL}$ ESCs (Figures 5A and S5A). Clones were sequenced for the presence of the mutations and immunoblot confirmed loss of CNBP in both $cnbp^{KO}$ and $cnbp^{KO};bvht^{dAGIL}$ ESCs (Figure 5B). Importantly, neither disruption of the *Bvht* AGIL motif nor $cnbp^{KO}$ affected the expression of either CNBP or *Bvht,* respectively (Figures 5B

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and S5B). Moreover, loss of CNBP did not affect the expression of ESC pluripotency markers Oct4 and Nanog, similar to *bvht*^{dAGIL} (Figure S5B).

We next tested two independent *cnbp*^{KO} ESC clones for their ability to differentiate into cardiomyocytes. As shown in Figure S5C, *cnbp*^{KO} cells show similar morphologies to wild-type cells at both Day 2 and Day 4 of differentiation and are fully capable of differentiating into CPs at Day 5.3 and CMs at Day 10 as shown by immunofluorescence analysis of Nkx2.5-GFP and cTnT respectively. In fact, *cnbp*^{KO} cells generate significantly higher percentages of Nkx2.5-GPF+ cells (CP) at Day 5.3 and cTnT+ cells (CM) at Day 10 by FACS when compared to wild-type cells (Figure 5C). Moreover, qRT-PCR analysis showed that cardiac transcription factors (e.g. Nkx2.5, Gata4, Gata6, Hand2, Tbx5) at Day 5.3 and CM marker genes (e.g. cTnT, Myh6, Myh7) at Day 10 exhibit higher expression levels in *cnbp*^{KO} cells compared to wild-type cells (Figure 5D, E).

To further test CNBP function, we constitutively over-expressed Flag-tagged CNBP in wild-type ESCs, which did not affect the expression levels of *Bvht* and ESCs pluripotency markers Oct4 and Nanog (Figure S5D). In contrast to the *cnbp*^{KO}, cells expressing higher levels of CNBP produced significantly lower percentages of Nkx2.5-GPF+ cells (CP) at Day 5.3 and cTnT+ cells (CM) at Day 10 compared to control cells by FACS (Figures 5G and S5E). Consistent with these data, cardiac transcription factors and CM marker genes showed decreased expression levels upon CNBP overexpression (Figure 5H,I). Together, our data suggests that CNBP functions, in part, as a negative regulator of cardiovascular lineage commitment.

Loss of CNBP partially rescues the *bvht*^{dAGIL} phenotype

Based on the above results, we hypothesized that *Bvht* may functionally antagonize CNBP to promote cardiovascular lineage commitment predicting that loss of CNBP would rescue the *bvht*^{dAGIL} mutant phenotype. To test this idea, we first performed EB differentiation of *cnbp*^{KO}*;bvht*^{dAGIL} ESCs compared to wild-type ESCs. At Day 12 of EB differentiation, the expression levels of cardiomyocyte marker genes including cTnT, Myh6, and Myh7 were significantly restored in the *cnbp*^{KO}*;bvht*^{dAGIL} double mutant cells (Figure S6A). We then performed the cardiomyocyte differentiation assay and found that the *cnbp*^{KO}*;bvht*^{dAGIL} double mutants produced significantly increased percentages of CP and CM cells compared to the *bvht*^{dAGIL} mutant alone (Figure 6A). Nkx2.5 is expressed throughout the CP to CM stages (Ma et al., 2008; Wamstad et al., 2012). Markedly, *cnbp*^{KO}*;bvht*^{dRHT} mutant cells generated a comparable percentage of Nkx2.5-GFP+ cells at Day 10 (CM) to wild-type cells. Coinciding with FACS analysis, the *cnbp*^{KO}*;bvht*^{dAGIL} cells also showed significant levels of Nkx2.5 and cTnT by immunofluorescence staining (Figure 6B,E) whereas levels were undetectable in *bvht*^{dAGIL} mutant alone (Figure 3C).

We then analyzed the expression of the mesodermal marker Brachyury and key cardiac transcription factors by qRT-PCR (Figure 6C,D). Brachyury levels in *cnbp*^{KO}*;bvht*^{dAGIL} double mutants showed comparable expression levels to wild-type cells. Expression of Nkx2.5, Gata4, Gata6, Hand2, and Mef2c was also partially restored at both Day 5.3 and Day 10. Moreover, we observed that the cardiomyocyte-specific genes cTnT, Myh6, and Myh7 showed a significant increase in expression (50%~70% relative to wild-type cells) in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cnbp^{KO};*bvht*^{dAGIL} double mutants compared to the AGIL mutant alone (Figure 6E). Together, our data suggest that CNBP and *Bvht* function together to regulate cardiovascular lineage commitment.

DISCUSSION

Our work establishes that RNA secondary structure determination coupled with genetic studies can reveal important functional motifs required for lncRNA mechanisms of action. Our study revealed several important findings regarding the role of *Bvht* in cardiovascular lineage commitment. First, we show that *Bvht* adopts a modular secondary structure *in vitro* that harbors a 5' asymmetric G-rich internal loop, termed AGIL. Remarkably, a small 11 nt deletion in the AGIL motif (*bvht*^{dAGIL}) within the ~590 nt non-coding transcript prevents the transition from nascent mesoderm to the cardiac progenitor state in our *in vitro* differentiation assay. Second, we found that the zinc finger transcription factor CNBP specifically interacts with *Bvht*, precisely at the AGIL motif. We also show that CNBP acts as a negative regulator of the cardiac developmental program and that genetic ablation of CNBP partially rescues the differentiation defect of *bvht*^{dAGIL} mutant cells. Collectively, these data suggest that *Bvht* functionally antagonizes CNBP to promote cardiovascular lineage commitment (Figure 6F).

In some cases, lncRNAs such as GAS5, PANDA, NF-YA, NORAD have been reported to function as molecular decoys to titrate interacting proteins away from their regulatory targets through competitive binding (Hung et al., 2011; Kino et al., 2010; Lee et al., 2016). However, the low abundance of *Bvht* transcript makes the molecular decoy model unlikely to explain the mode of action of *Bvht*. Expression of *Bvht* from the ROSA26 locus using its endogenous promoter largely rescues the AGIL mutant phenotype suggesting that low copy number is sufficient to mediate its function *in trans* in a locus specific manner (Figure S3G–K). Recently, lncRNAs including *Fendrr, PRNA,* and *PARTICL* were found to target specific genomic loci through directly hybridizing to nascent DNA via sequence complementarity or DNA:DNA:RNA, (Grote et al., 2013; O'Leary et al., 2015; Schmitz et al., 2010). In addition, it has been proposed that low abundance RNAs such as the RNA component of telomerase (TERC), which can perform multiple turnover reactions, could accomplish super-stoichometric functionalities (Goff and Rinn, 2015; Mozdy and Cech, 2006; Zappulla and Cech, 2004), providing another potential model for studying the molecular mechanisms of low abundance lncRNAs such as *Bvht* in future studies.

Our results suggest CNBP is a critical component of *Bvht*'s mode of action in cardiovascular lineage commitment. CNBP is highly conserved among vertebrates and can bind single-stranded G-rich DNA or RNA (Calcaterra et al., 2010). It has been proposed that CNBP acts as a nucleic acid chaperone and can promote the formation of G-quadruplex (G4) structures in which four guanines are assembled in a planar arrangement by Hoogsteen hydrogen bonding followed by intra- or inter- molecular folding of the tetramers (Armas et al., 2008; Borgognone et al., 2010; Rhodes and Lipps, 2015). For example, CNBP represses the expression of heterogeneous ribonucleoprotein K (hnRNPK) in fibrosarcoma cells and c-Myc in human HeLa cells through its conversion of promoter G-rich sequences into G4 DNA (Chen et al., 2013; Qiu et al., 2014). We found that different algorithms including

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

QGRS Mapper, QGRS-H Predictor and TetraplexFinder all predict *Bvht* AGIL motif forms a G-quadruplex (Figure S6B) (Kikin et al., 2006; Menendez et al., 2012; Yadav et al., 2008). Notably, G4 motifs have been identified in the promoters or UTRs of cardiac genes such Nkx2.5, Gata4, Mef2d (Nie et al., 2015; Zhang et al., 2008). Moreover, the specific inactivation of the G4 resolving RNA helicase RHAU in either cardiac mesoderm or progenitors leads to abnormal heart development (Nie et al., 2015). Thus, It is possible that *Bvht* and CNBP function together to regulate cardiac gene expression through control of G4 structures. Our probing studies indicate that the stems that flank the AGIL motif may be important for maintaining the G-rich loop in a single-stranded conformation which could be important for facilitating CNBP binding to this region. Thus, detailed mechanistic follow up of this and other models as well as dissecting the function of additional *Bvht* AGIL-interacting proteins will be a focus of future investigation.

Our work suggests that the identification of specific motifs by carefully dissecting individual lncRNAs is critical for understanding overall lncRNA function and can explain why these transcripts are overall lowly conserved at the sequence level. Recently, *in vivo* DMS and SHAPE methods were developed to directly probe RNA structure in living cells (Ding et al., 2014; Rouskin et al., 2014; Spitale et al., 2013), however, the low abundance of many lncRNAs makes it difficult probe their structures *in vivo* and the complex binding of proteins to RNAs can make interpretation of probing signals more complicated. Thus, secondary structure maps of free RNA molecules are necessary to facilitate a complete understanding of how these structures contribute to lncRNA modes of action under complex conditions. Together, determination of lncRNA motifs using both *in vitro* and *in vivo* probing results could be used to identify motif sequence fingerprints and homologues across species through phylogenetic sequence alignments and covariance analysis (Sanbonmatsu, 2016). Recent studies show that lncRNAs make undertake multiple secondary structure conformations in vivo (Lu, et al., 2016). While our combined 3S and CRISPR analyses give us high confidence, especially in the AGIL motif, we have not eliminated the possibility of alternative folds. In some cases, RNA also forms higher-order structures composed of tightly packed secondary structure elements (Leontis et al., 2006; Weeks, 2010). Thus, dissecting tertiary structures of lncRNAs under physiological conditions also represents an important area for future investigation. *In vitro* secondary structures of lncRNAs are also an important first step towards crystallographic and cryo-EM 3-D structures. Ultimately, studies aimed at mechanistic dissection of lncRNA structures are expected to facilitate a detailed understanding of how these transcripts contribute to fundamental biological processes and open the door to exploiting RNA motifs as biological and therapeutic tools.

EXPERIMENTAL PROCEDURES

Detailed experimental and analysis methods can be found in the Supplemental Materials.

Chemical Probing

SHAPE probing was performed using fast-acting 1M7 reagent (Deigan et al., 2009), and DMS probing was performed as described (Tijerina et al., 2007).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ESC Lines and Growth Conditions

Mouse ESCs were cultured on irradiated MEFs using standard conditions as previously described (Wamstad et al., 2012). NKX2.5-GFP mouse ESCs (Hsiao et al., 2008) were used as wild-type ESC (WT) in this study.

Generation of ESC Lines with CRISPR/Cas9

CRISPR/Cas9-mediated homology directed repair or non-homologous end joining was performed as described (Ran et al., 2013) using a bicistronic expression vector expressing Cas9 and sgRNA (px330, Addgene #42230).

ESC Differentiation

Embryoid body formation and directed differentiation were performed as described (Klattenhoff et al., 2013; Wamstad et al., 2012).

Immunostaining ESCs and Differentiated Cell Types

Cells were fixed and stained according to our previous studies (Klattenhoff et al., 2013; Wamstad et al., 2012).

Flag-tagged CNBP, HNRNPF, SFSR9

Flag-tagged CNBP/HNRNPF/SFRS9 cassette was cloned into pEGIP (Addgene #26777). Lentiviral production and ESC infection were performed using protocols from the RNAi Consortium (Broad Institute).

RNA Immunoprecipitation

Cells were UV cross-linked and RNA immunoprecipitation was performed as described (Jeon and Lee, 2011; Lai et al., 2013).

ProtoArray Processing and Analysis

In vitro RNA production and labeling followed by probing the ProtoArray Human Protein Microarray v5.0 (Life Technologies cat# PAH0525101) were performed as described (Siprashvili et al., 2012).

ACCESSION NUMBERS

ProtoArray raw data have been deposited to ArrayExpress with accession numbers: E-MTAB-4995.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank members of the Boyer lab, David Bartel, Igor Ulitski, and Matt Taliaferro for insightful discussions and for critical evaluation of the manuscript. We also thank Chikdu Shivalila for technical support in CRIPSR/Cas9 experiments, Lionel Lam for help scanning the protein microarray, and the flow cytometry core for technical

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

support. K.Y.S., S.P.H., I.V.N. were supported by LANL-LDRD. This work was also supported in part by the Koch Institute Core grant (P30CA014051) and the NHLBI Bench to Bassinet Program (U01HL098179) to L.A.B.

REFERENCES

Aguilera A, and Garcia-Muse T (2012). R loops: from transcription byproducts to threats to genome stability. Mol Cell 46, 115–124. [PubMed: 22541554]

Armas P, Nasif S, and Calcaterra NB (2008). Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. J Cell Biochem 103, 1013–1036. [PubMed: 17661353]

Bindewald E, Hayes R, Yingling YG, Kasprzak W, and Shapiro BA (2008). RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. Nucleic Acids Res 36, D392–397. [PubMed: 17947325]

Bondue A, Lapouge G, Paulissen C, Semeraro C, Iacovino M, Kyba M, and Blanpain C (2008). Mesp1 Acts as a Master Regulator of Multipotent Cardiovascular Progenitor Specification. Cell Stem Cell 3, 69–84. [PubMed: 18593560]

Borgognone M, Armas P, and Calcaterra NB (2010). Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. Biochem J 428, 491–498. [PubMed: 20394585]

Calcaterra NB, Armas P, Weiner AM, and Borgognone M (2010). CNBP: a multifunctional nucleic acid chaperone involved in cell death and proliferation control. IUBMB life 62, 707–714. [PubMed: 20960530]

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC bioinformatics 3, 2. [PubMed: 11869452]

Chen S, Su L, Qiu J, Xiao N, Lin J, Tan JH, Ou TM, Gu LQ, Huang ZS, and Li D (2013). Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. Biochimica et biophysica acta 1830, 4769–4777. [PubMed: 23774591]

Chen W, Liang YQ, Deng WJ, Shimizu K, Ashique AM, Li E, and Li YP (2003). The zinc-finger protein CNBP is required for forebrain formation in the mouse. Development (Cambridge, England) 130, 1367–1379.

Chen W, Wang Y, Abe Y, Cheney L, Udd B, and Li YP (2007). Haploinsuffciency for Znf9 in Znf9+/− mice is associated with multiorgan abnormalities resembling myotonic dystrophy. J Mol Biol 368, 8–17. [PubMed: 17335846]

Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, and Chang HY (2015). Systematic discovery of Xist RNA binding proteins. Cell 161, 404–416. [PubMed: 25843628]

Deigan KE, Li TW, Mathews DH, and Weeks KM (2009). Accurate SHAPE-directed RNA structure determination. Proceedings of the National Academy of Sciences of the United States of America 106, 97–102. [PubMed: 19109441]

Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, and Assmann SM (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505, 696–700. [PubMed: 24270811]

Fatica A, and Bozzoni I (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 15, 7–21. [PubMed: 24296535]

Flynn Ryan A., and Chang Howard Y. (2014). Long Noncoding RNAs in Cell-Fate Programming and Reprogramming. Cell Stem Cell 14, 752–761. [PubMed: 24905165]

Geisler S, and Coller J (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 14, 699–712. [PubMed: 24105322]

Goff LA, and Rinn JL (2015). Linking RNA biology to lncRNAs. Genome research 25, 1456–1465. [PubMed: 26430155]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Grote P, Wittler L, Hendrix D, Koch F, Wahrisch S, Beisaw A, Macura K, Blass G, Kellis M, Werber M, et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Developmental cell 24, 206–214. [PubMed: 23369715]

Gutell RR, Lee JC, and Cannone JJ (2002). The accuracy of ribosomal RNA comparative structure models. Current opinion in structural biology 12, 301–310. [PubMed: 12127448]

Hsiao EC, Yoshinaga Y, Nguyen TD, Musone SL, Kim JE, Swinton P, Espineda I, Manalac C, deJong PJ, and Conklin BR (2008). Marking Embryonic Stem Cells with a 2A Self-Cleaving Peptide: A NKX2–5 Emerald GFP BAC Reporter. PLoS ONE 3, e2532. [PubMed: 18596956]

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet 43, 621–629. [PubMed: 21642992]

Jeon Y, and Lee JT (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell 146, 119–133. [PubMed: 21729784]

Johnsson P, Lipovich L, Grander D, and Morris KV (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochimica et biophysica acta 1840, 1063–1071. [PubMed: 24184936]

Jones K, Jin B, Iakova P, Huichalaf C, Sarkar P, Schneider-Gold C, Schoser B, Meola G, Shyu AB, Timchenko N, et al. (2011). RNA Foci, CUGBP1, and ZNF9 are the primary targets of the mutant CUG and CCUG repeats expanded in myotonic dystrophies type 1 and type 2. Am J Pathol 179, 2475–2489. [PubMed: 21889481]

Kattman SJ, Witty AD, Gagliardi M, Dubois NC, Niapour M, Hotta A, Ellis J, and Keller G (2011). Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell 8, 228–240. [PubMed: 21295278]

Kikin O, D'Antonio L, and Bagga PS (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res 34, W676–682. [PubMed: 16845096]

Kino T, Hurt DE, Ichijo T, Nader N, and Chrousos GP (2010). Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor. Sci. Signal 3, ra8. [PubMed: 20124551]

Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, et al. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 152, 570–583. [PubMed: 23352431]

Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, Lee CS, Flockhart RJ, Groff AF, Chow J, et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 493, 231–235. [PubMed: 23201690]

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, and Shiekhattar R (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494, 497–501. [PubMed: 23417068]

Lee JT, and Bartolomei MS (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 152, 1308–1323. [PubMed: 23498939]

Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y, and Mendell JT (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell 164, 69–80. [PubMed: 26724866]

Lee TM, Maurer MS, Karbassi I, Braastad C, Batish SD, and Chung WK (2012). Severe dilated cardiomyopathy in a patient with myotonic dystrophy type 2 and homozygous repeat expansion in ZNF9. Congestive heart failure 18, 183–186. [PubMed: 22587749]

Leontis NB, Lescoute A, and Westhof E (2006). The building blocks and motifs of RNA architecture. Current opinion in structural biology 16, 279–287. [PubMed: 16713707]

Lindsley RC, Gill JG, Murphy TL, Langer EM, Cai M, Mashayekhi M, Wang W, Niwa N, Nerbonne JM, Kyba M, et al. (2008). Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. Cell Stem Cell 3, 55–68. [PubMed: 18593559]

Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, and Ranum LP (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science 293, 864–867. [PubMed: 11486088]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Ma Q, Zhou B, and Pu WT (2008). Reassessment of Isl1 and Nkx2–5 cardiac fate maps using a Gata4-based reporter of Cre activity. Dev Biol 323, 98–104. [PubMed: 18775691]

Marques Howarth M, Simpson D, Ngok SP, Nieves B, Chen R, Siprashvili Z, Vaka D, Breese MR, Crompton BD, Alexe G, et al. (2014). Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. J Clin Invest 124, 5275–5290. [PubMed: 25401475]

Matunis MJ, Xing J, and Dreyfuss G (1994). The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. Nucleic Acids Res 22, 1059–1067. [PubMed: 7512260]

Menendez C, Frees S, and Bagga PS (2012). QGRS-H Predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. Nucleic Acids Research 40, W96–W103. [PubMed: 22576365]

Mian IS (1997). Comparative sequence analysis of ribonucleases HII, III, II PH and D. Nucleic Acids Res 25, 3187–3195. [PubMed: 9241229]

Mondragon A (2013). Structural studies of RNase P. Annu Rev Biophys 42, 537–557. [PubMed: 23654306]

Mozdy AD, and Cech TR (2006). Low abundance of telomerase in yeast: implications for telomerase haploinsufficiency. RNA 12, 1721–1737. [PubMed: 16894218]

Nie J, Jiang M, Zhang X, Tang H, Jin H, Huang X, Yuan B, Zhang C, Lai JC, Nagamine Y, et al. (2015). Post-transcriptional Regulation of Nkx2–5 by RHAU in Heart Development. Cell Rep 13, 723–732. [PubMed: 26489465]

Noller HF (1984). Structure of Ribosomal RNA. Annual Review of Biochemistry 53, 119–162.

Novikova IV, Hennelly SP, and Sanbonmatsu KY (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Res 40, 5034–5051. [PubMed: 22362738]

Novikova IV, Hennelly SP, and Sanbonmatsu KY (2013). 3S: Shotgun Secondary Structure determination for long non-coding RNAs. Methods 63, 170–177. [PubMed: 23927838]

O'Leary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, Moertl S, Trau M, Atkinson MJ, and Anastasov N (2015). PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation. Cell Rep 11, 474–485. [PubMed: 25900080]

Parsch J, Braverman JM, and Stephan W (2000). Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics 154, 909–921. [PubMed: 10655240]

Ponting CP, Oliver PL, and Reik W (2009). Evolution and functions of long noncoding RNAs. Cell 136, 629–641. [PubMed: 19239885]

Qiu J, Chen S, Su L, Liu J, Xiao N, Ou TM, Tan JH, Gu LQ, Huang ZS, and Li D (2014). Cellular nucleic acid binding protein suppresses tumor cell metastasis and induces tumor cell death by downregulating heterogeneous ribonucleoprotein K in fibrosarcoma cells. Biochimica et biophysica acta 1840, 2244–2252. [PubMed: 24594223]

Quinn JJ, and Chang HY (2015). Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17, 47–62.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, and Zhang F (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protocols 8, 2281–2308. [PubMed: 24157548]

Reznik B, Clement SL, and Lykke-Andersen J (2014). hnRNP F complexes with tristetraprolin and stimulates ARE-mRNA decay. PLoS One 9, e100992. [PubMed: 24978456]

Rhodes D, and Lipps HJ (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Research.

Rinn JL, and Chang HY (2012). Genome Regulation by Long Noncoding RNAs. Annual Review of Biochemistry 81, 145–166.

Rizki G, and Boyer LA (2015). Lncing epigenetic control of transcription to cardiovascular development and disease. Circ Res 117, 192–206. [PubMed: 26139858]

Rouskin S, Zubradt M, Washietl S, Kellis M, and Weissman JS (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505, 701–705. [PubMed: 24336214]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sanbonmatsu KY (2016). Towards structural classification of long non-coding RNAs. Biochimica et biophysica acta 1859, 41–45. [PubMed: 26537437]

Schmitz KM, Mayer C, Postepska A, and Grummt I (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 24, 2264–2269. [PubMed: 20952535]

Serganov A, and Patel DJ (2007). Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet 8, 776–790. [PubMed: 17846637]

Siprashvili Z, Webster D, Kretz M, Johnston D, Rinn J, Chang H, and Khavari P (2012). Identification of proteins binding coding and non-coding human RNAs using protein microarrays. BMC Genomics 13, 633. [PubMed: 23157412]

Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, and Pyle AM (2015). HOTAIR forms an intricate and modular secondary structure. Mol Cell 58, 353–361. [PubMed: 25866246]

Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, and Chang HY (2013). RNA SHAPE analysis in living cells. Nature chemical biology 9, 18–20. [PubMed: 23178934]

Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature 519, 486–490. [PubMed: 25799993]

Tan E, Wilson TJ, Nahas MK, Clegg RM, Lilley DM, and Ha T (2003). A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. Proc Natl Acad Sci U S A 100, 9308–9313. [PubMed: 12883002]

Tijerina P, Mohr S, and Russell R (2007). DMS footprinting of structured RNAs and RNA-protein complexes. Nature protocols 2, 2608–2623. [PubMed: 17948004]

Ulitsky I, and Bartel DP (2013). lincRNAs: genomics, evolution, and mechanisms. Cell 154, 26–46. [PubMed: 23827673]

Wamstad Joseph A., Alexander Jeffrey M., Truty Rebecca M., Shrikumar A, Li F, Eilertson Kirsten E., Ding H, Wylie John N., Pico Alexander R., Capra John A., et al. (2012). Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. Cell 151, 206–220. [PubMed: 22981692]

Wang Y, Ma M, Xiao X, and Wang Z (2012). Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. Nat Struct Mol Biol 19, 1044–1052. [PubMed: 22983564]

Weeks KM (2010). Advances in RNA structure analysis by chemical probing. Current opinion in structural biology 20, 295–304. [PubMed: 20447823]

Weiner AM, Allende ML, Becker TS, and Calcaterra NB (2007). CNBP mediates neural crest cell expansion by controlling cell proliferation and cell survival during rostral head development. J Cell Biochem 102, 1553–1570. [PubMed: 17471504]

Weiner AM, Sdrigotti MA, Kelsh RN, and Calcaterra NB (2011). Deciphering the cellular and molecular roles of cellular nucleic acid binding protein during cranial neural crest development. Development, growth & differentiation 53, 934–947.

Yadav VK, Abraham JK, Mani P, Kulshrestha R, and Chowdhury S (2008). QuadBase: genome-wide database of G4 DNA--occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. Nucleic Acids Res 36, D381–385. [PubMed: 17962308]

Zappulla DC, and Cech TR (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. Proc Natl Acad Sci U S A 101, 10024–10029. [PubMed: 15226497]

Zhang R, Lin Y, and Zhang CT (2008). Greglist: a database listing potential G-quadruplex regulated genes. Nucleic Acids Res 36, D372–376. [PubMed: 17916572]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

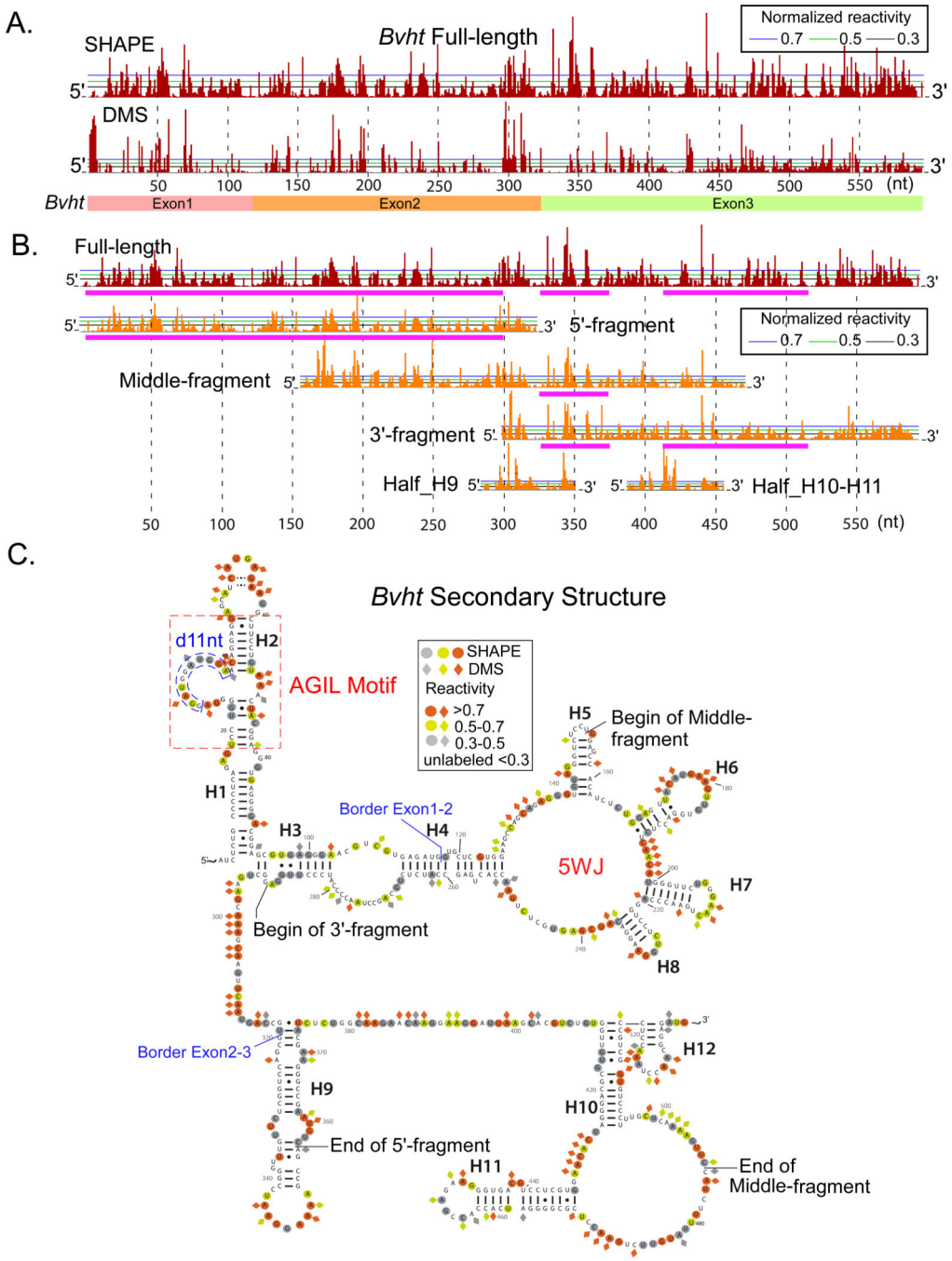

Figure 1. *Bvht* secondary structure determination by chemical probing.

(A) Normalized SHAPE (Top) and DMS (Bottom) probing reactivity profiles of full-length *Bvht*. Horizontal lines indicate normalized dimensionless reactivity. Both traces were normalized by the reactivities for highly reactive nucleotides. Nucleotides that have a normalized reactivity >0.5 are considered as highly flexible and likely represent single-stranded regions. Positions of *Bvht* exons are labeled below the reactivity profile.

(B) Shotgun secondary structure (3S) analysis of *Bvht*. Normalized SHAPE probing reactivity of indicated *Bvht* fragments is compared to full-length transcript. Full-length (1–

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

590), 5’-fragment (1–325), Middle-fragment (155–475), 3’-fragment (300–590), Half_H9 (282–349), and Half_H10-H11 (380–457). The sub-regions with highly similar reactivity patterns to full-length transcript are highlighted in purple under the reactivity profile.
(C) Secondary structure of *Bvht* was derived with 3S via SHAPE and DMS chemical probing experiments. The normalized SHAPE or DMS reactivity is represented by indicated colors. Circle, SHAPE; Diamond, DMS. The AGIL motif is highlighted by red dashed lines. H1 to H12 indicates the helices.
See also Figure S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

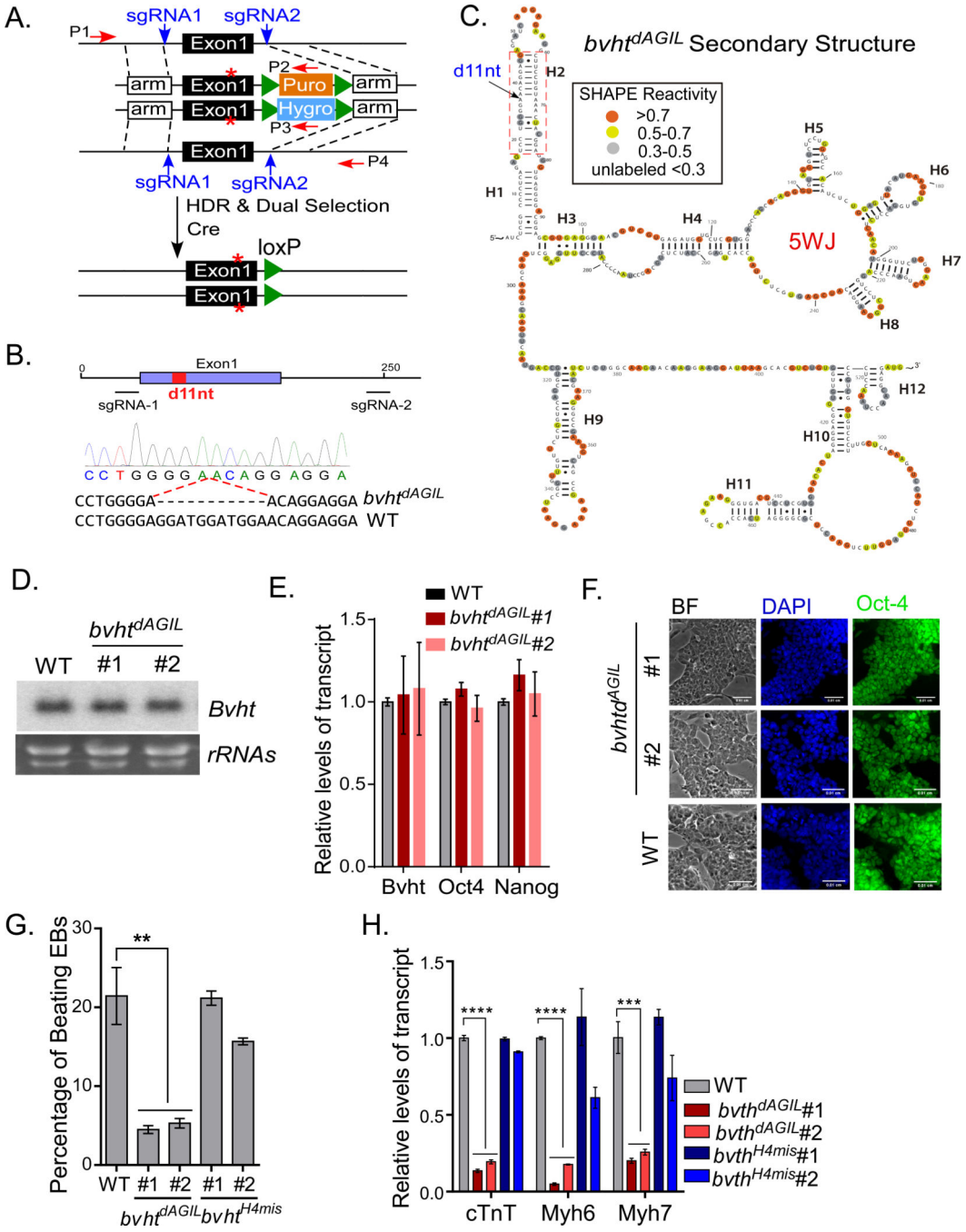

Figure 2. *Bvht* AGIL motif is necessary for formation of contracting embryoid bodies.

(A) Schematic showing the strategy of introducing mutations in *Bvht* endogenous locus. Two small guide RNAs (sgRNAs) and two repair templates including different selection cassettes (Puro or Hygro) as indicated are applied for CRISPR/Cas9-mediated HDR. After dual selection, both alleles will be mutated at designated loci. The selection cassettes are then removed by Cre recombinase-mediated recombination. Asterisk, mutations; triangle, loxP site; P1, P2, P3, P4 are primers for PCR-based screening.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(B) Diagram showing the positions of sgRNAs and d11nt in *Bvht* endogenous locus. Partial DNA sequencing trace of the PCR product of *bvht*^{dAGIL} ESC genomic DNA.
(C) Secondary structure of *bvht*^{dAGIL} was derived from SHAPE probing experiment. d11nt indicates the deleted 11nt sequences from AGIL motif.
(D) Northern Blot analysis showing the levels of *Bvht* transcripts in indicated ESC lines. Ribosomal RNAs (rRNAs) are used for loading control.
(E) qRT-PCR analysis showing the levels of *Bvht* and ESCs pluripotency markers Oct4 and Nanog in indicated ESC lines. Experiments were performed in triplicate and data are represented as mean values ± SD.
(F) Immunofluorescence staining of indicated ESCs using Oct4 antibody. Nuclei were stained with DAPI. BF, bright field. Scale bar, 100μm.
(G) Percentage of spontaneously contracting embryoid bodies (EBs) at Day 12 of differentiation (n>200) from indicated ESCs. Experiments were performed in triplicate and data are represented as mean values ± SD. ** $p<0.01$(two-tailed Student's t-test).
(H) qRT-PCR analysis of EBs at day 12 showing the relative levels of cardiomyocyte markers from indicated ESC lines.
All experiments were performed in triplicate and data are represented as mean values ± SD. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$ (two-tailed Student's t-test).
See also Figure S2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

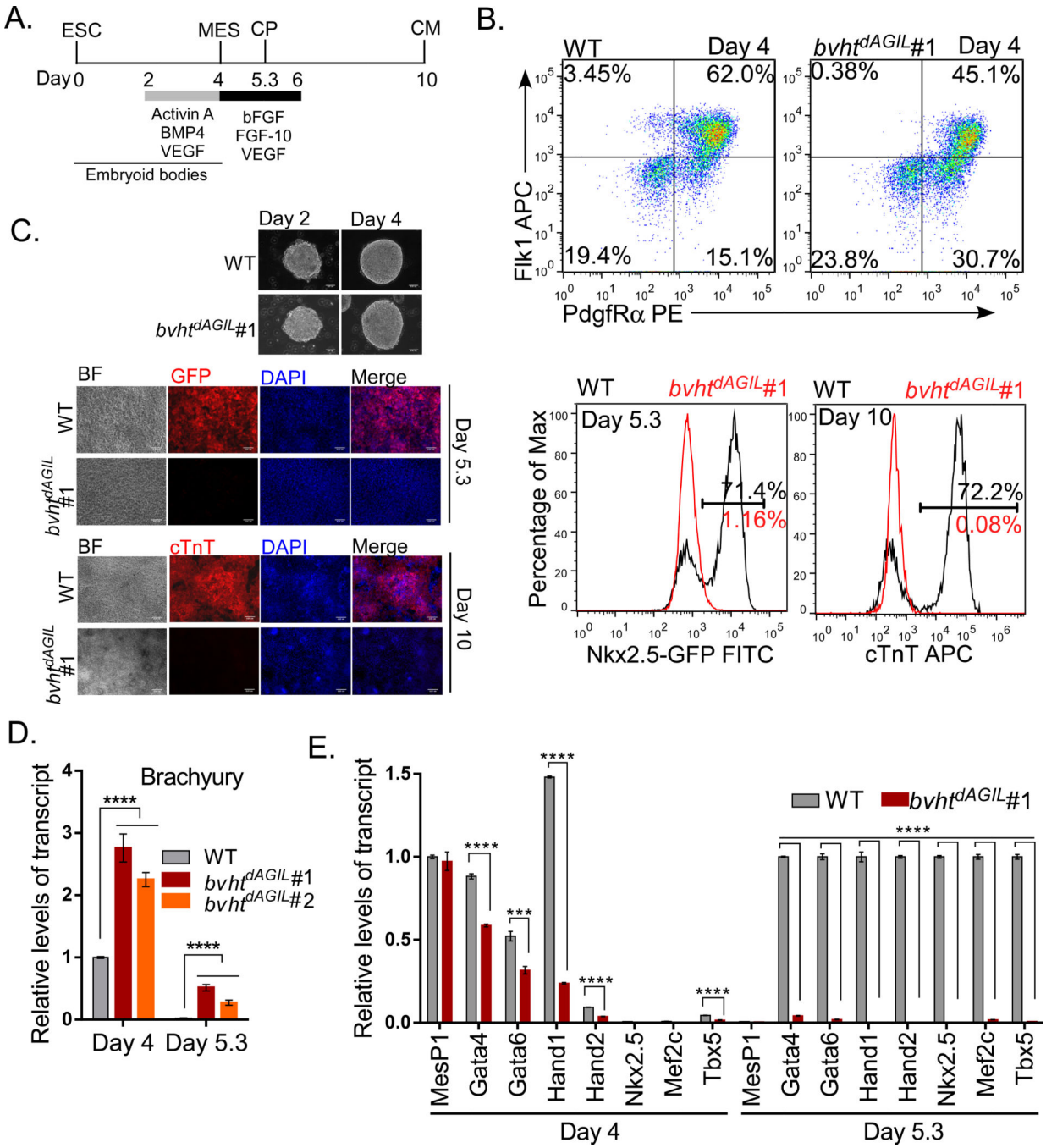

Figure 3. *Bvht* AGIL motif is necessary for cardiovascular lineage commitment

(A) Timeline of CM differentiation protocol. Black and gray bars represent the time period where differentiating cultures were treated with the growth factors listed below each respective bar.

(B) Cells at indicated time points were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations.

(C) Immunofluorescence staining of indicated cells using anti-GFP (Day 5.3) and anti-cTnT (Day 10) antibodies. Nuclei were stained with DAPI. BF, bright field. Scale bar, 100μm.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(D, E) qRT-PCR analysis showing the relative levels of Brachyury (D) and cardiac transcription factors (E) at Day 4 and Day 5.3 from indicated ESC lines. WT value at Day 4 (D) and Day 5.3 (E) is set to 1 for each gene.

Experiments were performed in triplicate and data are represented as mean values +/− SD. ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$ (two-tailed Student's t-test).

See also Figure S3.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

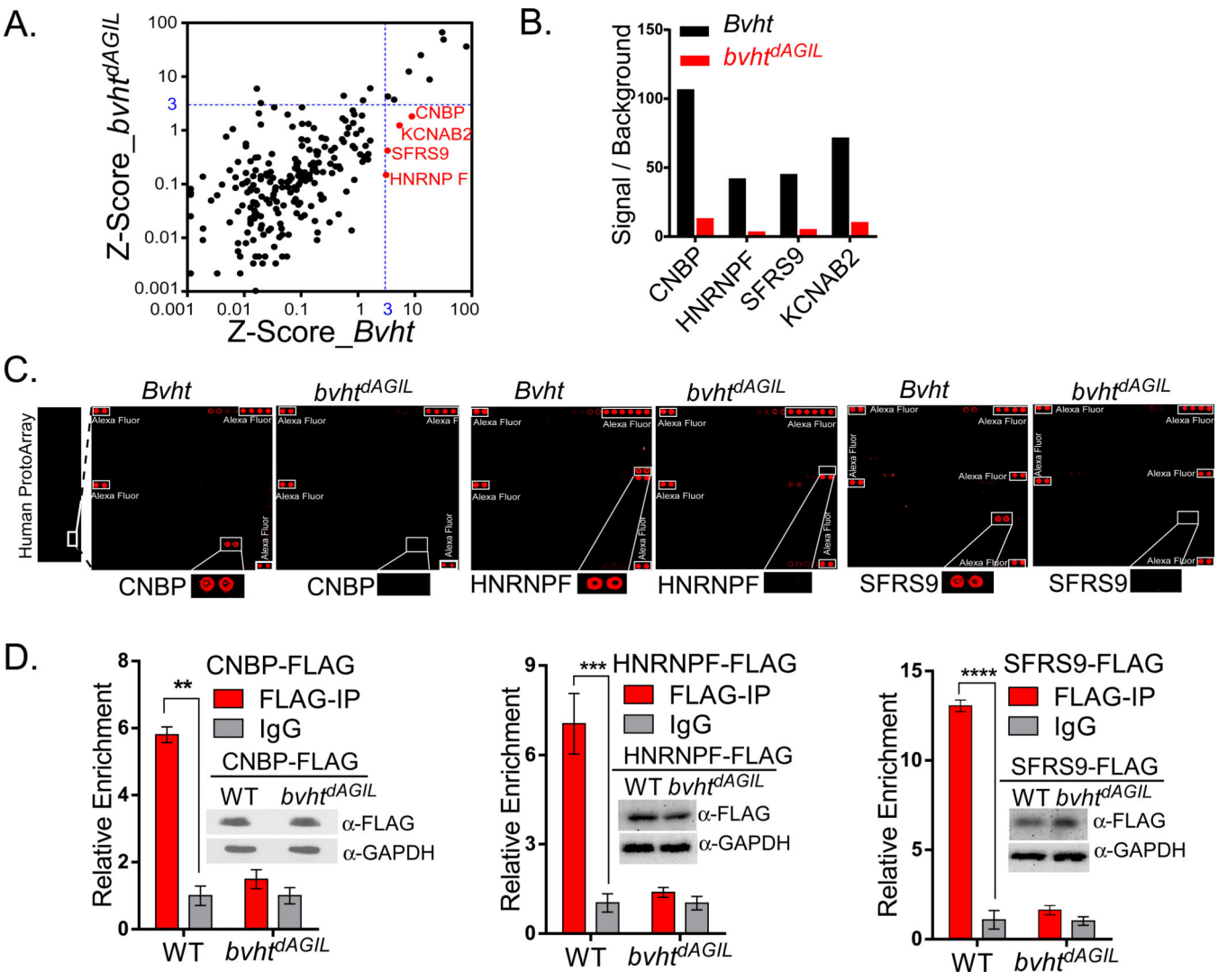

Figure 4. *Bvht* interacts with CNBP, a zinc finger transcription factor

(A) Protein microarray analysis detecting *Bvht*-interacting proteins. Cy5-labeled *Bvht* and *bvht*^{dAGIL} transcripts were incubated with a human recombinant protein microarray. Z-Scores of fold changes signal intensity over background are depicted in the scatter plot. The dashed blue line represents the Z-Score cutoff used to select significant RNA-protein binding events. The significant *Bvht*-binding proteins are colored in red. The horizontal axis is *Bvht*, and the vertical axis for *bvht*^{dAGIL}. A logarithmic scale was used to display both axes.

(B) Quantification of human protein microarray showing fold changes of signal intensity over background for indicated proteins. Values are the average of duplicate protein spots.

(C) Image of human protein microarray (left) and enlarged subarray (right) showing that mutation of AGIL motif dramatically reduces the interaction between *Bvht* and CNBP, HNRNPF, SFRS9. Alexa Fluor 647-labelled rabbit anti-mouse IgG or anti-human IgA2 in corners and middle edge of each subarray are used for reference.

(D) RNA immunoprecipitation showing the interaction between *Bvht* and CNBP, HNRNPF and SFRS9 in ESCs. Flag-tagged CNBP, HNRNPF, or SFRS9 was constitutively expressed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in both WT and *bvht*^{dAGIL} ESCs. Immunoblot analysis using anti-Flag antibody shows equal expression levels of Flag-tagged CNBP, HNRNPF, SFRS9 in indicated ESCs. Mouse IgG was used for negative control.

Experiments were performed in triplicate and data are represented as mean values +/− SD. ** p<0.01; *** p<0.001; **** p<0.0001 (two-tailed Student's t-test).

See also Figure S4 and Table S2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

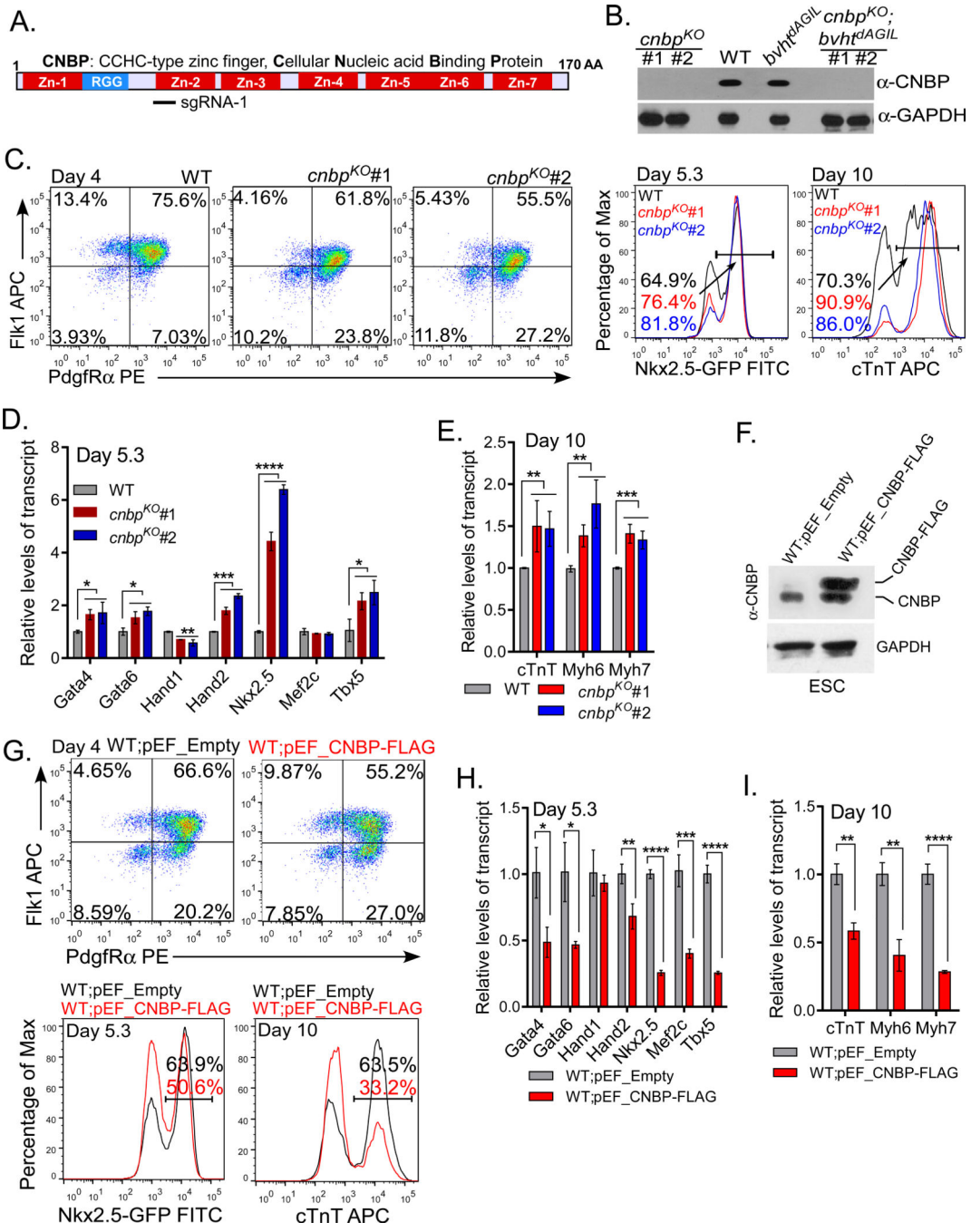

Figure 5. CNBP represses cardiomyocyte differentiation

(A) Diagram of CNBP (Uniprot P53996–2) functional domains, including seven CCHC zinc fingers (aa 4–21, 45–62, 65–82, 89–106, 110–127, 128–145,149–166) and RGG box of RNA-binding (aa 22–35). The target sequence of CNBP_sgRNA-1 is labeled on the bottom.

(B) Immunoblot analysis with anti-CNBP antibody showing the protein levels of CNBP in indicated ESC lines. GAPDH was used as loading control.

(C) Cells at indicated time points were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(D, E) qRT-PCR analysis showing the relative levels of cardiac marker genes at Day 5.3 and Day 10 of CM differentiation.

(F) Immunoblot analysis with anti-CNBP antibody showing the protein levels of endogenous CNBP and recombinant CNBP-FLAG in ESCs. GAPDH was used as loading control.

(G) Cells at indicated time points were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations.

(H, I) qRT-PCR analysis showing the relative levels of cardiac marker genes at Day 5.3 and Day 10 of CM differentiation.

Experiments were performed in triplicate and data are represented as mean values +/− SD. ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$ (two-tailed Student's t-test).

See also Figure S5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

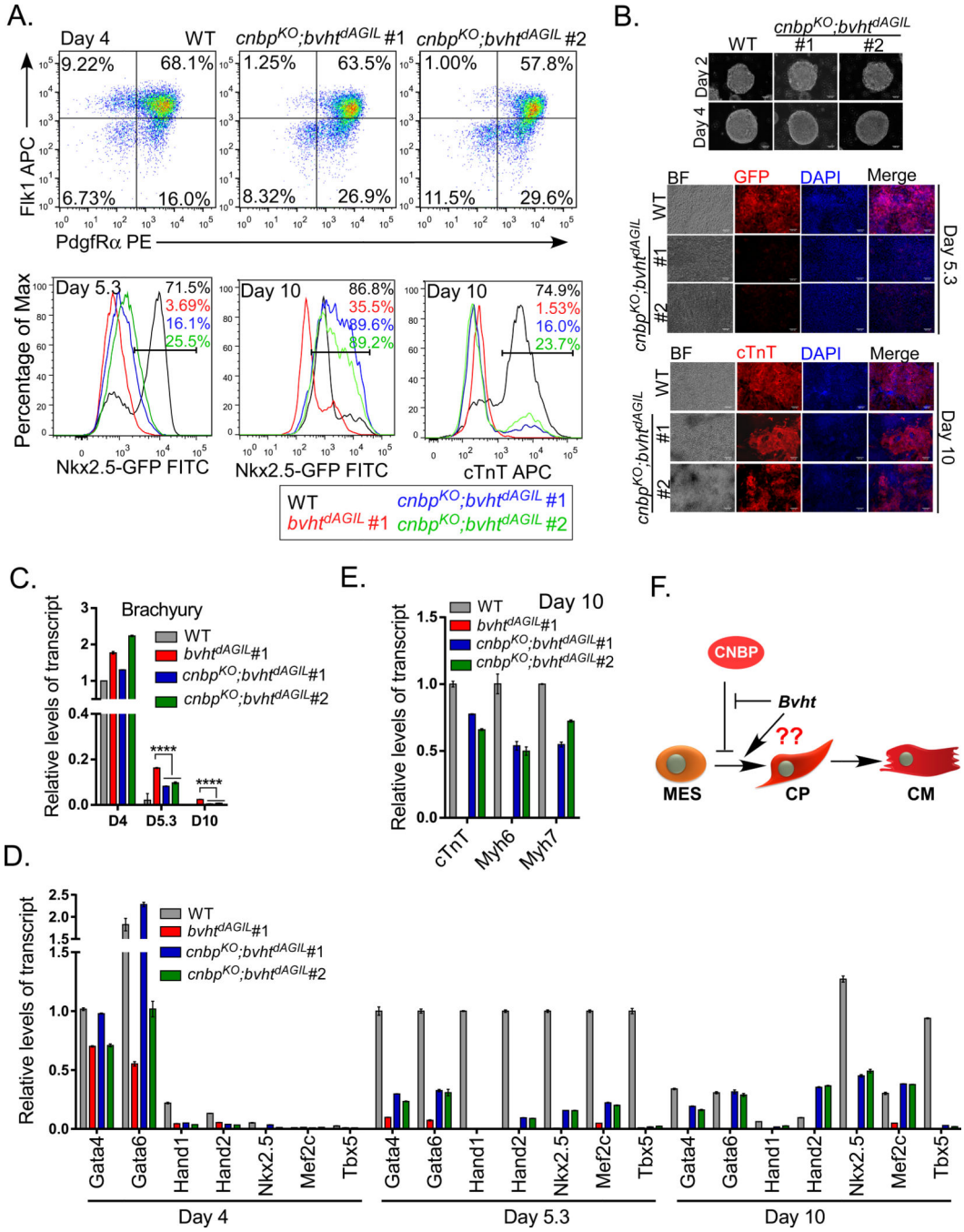

Figure 6. Loss of CNBP partially rescues the *bvht*^{dAGIL} phenotype

(A) Cells at indicated time points during CM differentiation were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations. (B) Immunofluorescence staining of indicated cells using anti-GFP (Day 5.3) and anti-cTnT (Day 10) antibodies. Nuclei were stained with DAPI. BF, bright field. Scale bar is 100μm. (C, D) qRT-PCR analysis showing the relative levels of Brachyury (C) and core cardiac transcription factors (D). WT value at Day 4 (C) or at Day 5.3 (D) was set to 1 for each gene.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(E) qRT-PCR analysis showing the relative levels of cardiomyocyte marker genes at Day 10. (F) Model of *Bvht* and CNBP regulating cardiovascular lineage commitment. *Bvht* functionally antagonizes the repression of CNBP on the transition from cardiac mesoderm to progenitors. Potential additional factors working together with *Bvht* remained to be elucidated.

Experiments were performed in triplicate and data are represented as mean values +/− SD. ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$ (two-tailed Student's t-test).

See also Figure S6.