

## Identification of Novel Long Noncoding RNAs Underlying Vertebrate Cardiovascular Development

Leo Kurian, PhD\*; Aitor Aguirre, PhD\*; Ignacio Sancho-Martinez, PhD\*; Christopher Benner, PhD; Tomoaki Hishida, PhD; Thai B. Nguyen, BSc; Pradeep Reddy, PhD; Emmanuel Nivet, PhD; Marie N. Krause, MD; David A. Nelles, BSc; Concepcion Rodriguez Esteban, PhD; Josep M. Campistol, MD; Gene W. Yeo, PhD; Juan Carlos Izpisua Belmonte, PhD

**Background**—Long noncoding RNAs (lncRNAs) have emerged as critical epigenetic regulators with important functions in development and disease. Here, we sought to identify and functionally characterize novel lncRNAs critical for vertebrate development.

**Methods and Results**—By relying on human pluripotent stem cell differentiation models, we investigated lncRNAs differentially regulated at key steps during human cardiovascular development with a special focus on vascular endothelial cells. RNA sequencing led to the generation of large data sets that serve as a gene expression roadmap highlighting gene expression changes during human pluripotent cell differentiation. Stage-specific analyses led to the identification of 3 previously uncharacterized lncRNAs, *TERMINATOR*, *ALIEN*, and *PUNISHER*, specifically expressed in undifferentiated pluripotent stem cells, cardiovascular progenitors, and differentiated endothelial cells, respectively. Functional characterization, including localization studies, dynamic expression analyses, epigenetic modification monitoring, and knockdown experiments in lower vertebrates, as well as murine embryos and human cells, confirmed a critical role for each lncRNA specific for each analyzed developmental stage.

**Conclusions**—We have identified and functionally characterized 3 novel lncRNAs involved in vertebrate and human cardiovascular development, and we provide a comprehensive transcriptomic roadmap that sheds new light on the molecular mechanisms underlying human embryonic development, mesodermal commitment, and cardiovascular specification. (*Circulation*. 2015;131:1278-1290. DOI: 10.1161/CIRCULATIONAHA.114.013303.)

**Key Words:** cardiovascular system ■ growth and development ■ RNA, long noncoding  
■ transcriptome ■ vertebrates

The advent of novel sequencing technologies has revealed that <2% of the human genome encodes for proteins.<sup>1-3</sup> Interestingly, despite not being translated into proteins, a large fraction of the mammalian genome is transcribed into what is known as noncoding RNAs (ncRNAs).<sup>1-3</sup> The consistent observation of pervasive transcription originating from these noncoding sites suggests that ncRNAs might play a key role in fundamental biological functions.<sup>1-3</sup> To date, thousands of ncRNAs have been putatively described. However, the precise functional roles of the vast majority remain unclear.<sup>4-6</sup>

---

### Editorial see p 1236 Clinical Perspective on p 1290

---

Among the different classes of ncRNAs, long ncRNAs (lncRNAs), broadly defined as those noncoding transcripts larger than 200 nucleotides,<sup>1</sup> represent one of the largest and least understood nucleic acid molecules in vertebrates. Overall, >9000 genomic loci are predicted to code for these transcript subclasses in the human genome.<sup>5,7,8</sup> Reports describing important developmental roles for certain lncRNAs in various vertebrates have appeared over the last few years.<sup>9-11</sup> Low expression

---

Received July 3, 2014; accepted January 29, 2015.

From Gene Expression Laboratory (L.K., A.A., I.S.-M., T.H., T.B.N., P.R., E.N., M.N.K., C.R.E., J.C.I.B.) and Integrative Genomics Core (C.B.), Salk Institute for Biological Studies, La Jolla, CA; University of California San Diego, Department of Cellular and Molecular Medicine, Stem Cell Program, Institute for Genomic Medicine, Sanford Consortium for Regenerative Medicine, La Jolla (L.K., T.B.N., D.A.N., G.W.Y.); and Hospital Clinic, University of Barcelona, IDIBAPS, Barcelona, Spain (J.M.C.).

Dr Kurian is currently at the Laboratory for Developmental and Regenerative RNA Biology, University of Cologne, Center for Molecular Medicine Cologne, CECAD, and Institute for Neurophysiology, Cologne, Germany. Dr Aguirre is currently at the University of California San Diego, School of Medicine, Department of Pediatrics, Genome Information Sciences Division, La Jolla, CA. Dr Nivet is currently at Aix Marseille Université, CNRS, NICN UMR 7259, Marseille, France.

\*Drs Kurian, Aguirre, and Sancho-Martinez contributed equally.

The online-only Data Supplement is available with this article at <http://circ.ahajournals.org/lookup/suppl/doi:10.1161/CIRCULATIONAHA.114.013303/-DC1>.

Correspondence to Juan Carlos Izpisua Belmonte, PhD, Salk Institute for Biological Studies, 10010 N Torrey Pines Rd, La Jolla, CA 92037. E-mail: belmonte@salk.edu

© 2015 American Heart Association, Inc.

*Circulation* is available at <http://circ.ahajournals.org>

DOI: 10.1161/CIRCULATIONAHA.114.013303

levels and poor sequence conservation have limited the identification and functional characterization of novel lncRNAs.<sup>12</sup> With the combination of novel sequencing and analytic technologies and the powerful platforms that pluripotent stem cell models provide for the study of human embryonic development, these limitations can now be circumvented.

Using a methodology recently developed in our laboratory that allows the highly efficient generation of human cardiovascular progenitor cells of mesodermal origin and terminally differentiated vascular endothelial cells,<sup>13</sup> we report here the identification and functional characterization of 3 novel human lncRNAs indispensable for the development of the cardiovascular system in different vertebrate species.

## Methods

### Cell Culture

H1 (WA1, WiCell; passage 25–40) human embryonic stem cells (hESCs) were used for differentiation studies. Terminally differentiated primary human umbilical vein endothelial cells (HUVECs) were used to investigate the transcriptome and functional roles of lncRNA candidates in unmodified vascular endothelial cells. Briefly, hESCs were cultured in chemically defined growth media, mTeSR (Stemcell Technologies), on growth factor-reduced Matrigel (BD Biosciences)-coated plates. After 70% to 80% confluent hESCs were treated with dispase (Invitrogen) for 7 minutes at 37°C, the colonies were dispersed to small clusters and lifted carefully with a 5-mL glass pipette at a ratio of ≈1:6. HUVECs, purchased from Promocell, were cultured in endothelial basal medium supplemented with endothelial growth medium SingleQuots (Lonza). hESC-derived endothelial cells were cultivated in endothelial growth medium-2 bullet kit media (Lonza). Endothelial cells were grown on collagen I-coated plates (BD Biosciences). All cell lines were maintained in an incubator (37°C and 5% CO<sub>2</sub>) with media changes every day (hESCs) or every second day (HUVECs).

### Cell Lineages Analyzed

Pluripotent stem cells were differentiated into early mesoderm-derived cells including c-Kit<sup>+</sup> and KDR<sup>+</sup> cardiovascular progenitors (day 2–4 of differentiation) and committed CD34<sup>+</sup>CD31<sup>+</sup> bipotent vascular progenitor cells (day 4–8 of differentiation) as previously described.<sup>13</sup> Kurian et al<sup>13</sup> provides detailed characterization of the differentiation methodologies and cell types generated. Terminally differentiated primary HUVECs were also included in our analyses as positive controls, and no significant differences between differentiated endothelial cells and HUVECs were found.<sup>13</sup> Briefly, for the generation of early cardiovascular progenitors (day 2–4) and committed vascular progenitor cells (day 4–8), undifferentiated hESCs were freshly split onto Matrigel-coated plates, making sure that the subcolonies were of small size (≈300–500 cells per colony). Cells were differentiated to different progenitor stages with the use of a chemically defined mesodermal induction media as previously described<sup>13</sup> (Dulbecco modified Eagle medium:F12, 15 mg/mL stem cell-grade BSA [MP Biomedicals], 17.5 µg/mL human insulin [SAFC Bioscience], 275 µg/mL human holo transferrin [Sigma Aldrich], 20 ng/mL basic fibroblast growth factor [Humanzyme], 50 ng/mL human vascular endothelial growth factor-165 aa [Humanzyme], 25 ng/mL human bone morphogenetic protein 4 [Stemgent], 450 µmol/L monothioglycerol [Sigma Aldrich], and 2.25 mmol/L each L-glutamine and nonessential amino acids [Invitrogen]). Switching day 8 mesodermal induction media–differentiated cells to endothelial growth medium-2 media as described<sup>13</sup> induced endothelial cell differentiation. HUVECs are of primary origin and were therefore not modified during our studies.

### RNA Sequencing

Total RNA from roughly 1×10<sup>7</sup> cells of each of the groups indicated above was isolated with TRIzol (Invitrogen). Intact total RNA samples were treated with DNase1 and sent to the Beijing Genomics Institute for sequencing. Before sequencing, all samples were subjected to

quality control processes to ensure the lack of contaminating DNA and the integrity of the RNA. All the RNA samples met the following RNA quality threshold: optical density (OD)260/280=2 to 2.2; OD260/230≥2.0; 28S:18S>1.0, RNA integrity number>7. Whole-transcriptome sequencing was then performed. Briefly, TruSeq Stranded Total RNA with a Ribo-Zero Human Kit (Illumina) was used to remove ribosomal RNA and to prepare strand-specific paired-end RNA sequence libraries. Ninety million 2× 100-bp paired end reads were sequenced on an Illumina HiSeq2000 instrument for each library. Reads were aligned to the hg19/GRCh37 version of the human genome by the use of STAR (version 2.1.4a). Only reads that aligned to a single unique location were kept for further analysis. Quantification of reads on each strand in 10-kb windows across the entire genome and comparisons with DNA sequencing data were used to further evaluate RNA purity and integrity (Figure I in the online-only Data Supplement).

### Primers Used in This Study

Table I in the online-only Data Supplement provides a list of primers and their respective sequences.

### Morpholino Embryo Injections

Morpholinos (Gene Tools) were designed against highly conserved regions in the lncRNAs or alternatively to block putative splice sites (Table II in the online-only Data Supplement). A pair of morpholinos was generated and tested per lncRNA. For zebrafish injection, morpholinos were dissolved in water at a 2-nmol/L stock concentration and diluted to a 2-nmol/L working concentration in PBS/phenol red solution. Embryo injections were performed by injecting ≈1 nL morpholino solution at the 1-cell stage with the use of a FemtoJet (Eppendorf). Morphants were evaluated at 24, 48, 72, and 96 hours in a StereoLumar stereoscope (Zeiss). For murine experiments, CD-1 female mice were superovulated by injecting pregnant mare's serum gonadotropin followed by human chorionic gonadotropin 48 hours later. The female mice were housed with males after human chorionic gonadotropin injection. One-cell-stage embryos were collected the next day after 20 hours and were injected with 10 pL of 2-nmol/L stock Terminator morpholino with a FemtoJet. The embryos were cultured until the blastocyst stage in K-modified simplex optimized medium, and RNA was extracted with the use of an RNeasy Mini Kit (Qiagen).

### Data Availability

All data sets in this study are available in the *National Center for Biotechnology Information* Gene Expression Omnibus under the accession number GSE54969.

### Statistical Evaluation

Statistical analyses of all end points were performed by statisticians at the Salk Institute and University of California San Diego using Excel, SPSS, or GraphPad software. All data presented a normal distribution. Statistical significance was evaluated with a standard unpaired Student *t* test (2 tailed; *P*<0.05) when appropriate. For multiple-comparison analysis, 1-way ANOVA with the Dunn posttest correction was applied when appropriate (*P*<0.05). Comparisons of groups with small sample sizes (*n*<6) were performed as follows: Mann-Whitney test (2 sided; 95% confidence level; *P*<0.05) was used, and a Kruskal-Wallis test with the Dunn posttest correction (*P*<0.05) was applied for multiple comparisons. All data are presented as mean±SD and represent a minimum of 3 independent experiments with at least 3 technical replicates unless otherwise stated.

## Results

### Noncoding RNAs Account for Half of the Transcriptome During Human Vascular Differentiation

With the goal of identifying and functionally validating novel genetic elements underlying cardiovascular vertebrate

development, we first decided to perform an RNA sequencing-based screening of transcripts differentially regulated during stem cell differentiation. We focused on our recently reported methodology for the efficient differentiation of human pluripotent cells to early c-Kit<sup>+</sup> cardiovascular progenitors, committed CD34<sup>+</sup>CD31<sup>+</sup> vascular progenitors, and terminally differentiated functional vascular cells.<sup>13</sup> We selected 5 different stages: undifferentiated hESCs, mesoderm committed cells (day 2; KDR<sup>+</sup>), early cardiovascular progenitor cells (day 4; c-Kit<sup>+</sup>), committed vascular progenitor cells (day 8; CD34<sup>+</sup>CD31<sup>+</sup>), and vascular endothelial cells (VE-cadherin<sup>+</sup>endoglin<sup>+</sup>). Kurian et al<sup>13</sup> provide detailed characterization of the generated cell lineages (Figure IA in the online-only Data Supplement). After rRNA depletion and size removal of species whose length was smaller than 150 nucleotides, we generated data sets with a sequencing depth of 90 million paired-end reads of 100 bp per sample (Figure 1A and 1B and Figure IIB and IIC and Table III in the online-only Data Supplement). To provide an overall estimation of the transcriptome changes occurring during the differentiation of pluripotent cells to vascular progenitor cells, we first decided to collectively assess global expression changes in all cell lineages, including early c-Kit<sup>+</sup> cardiovascular cells appearing from day 2 to 4 during the course of differentiation and committed vascular progenitor cells (day 4–8).<sup>13</sup> To describe the global transcriptomic changes across the genome, we investigated the genomic coverage of RNA sequencing data (at least 1 read). All different cell types indicated that ≈56% of the genome was transcriptionally activated during vascular progenitor cell differentiation<sup>14–17</sup> (Figure IIB in the online-only Data Supplement). More than 25 000 transcripts were expressed on differentiation (reads per kilobase per million >0.1), with 13 796 transcripts exhibiting at least 3-fold changes in expression. Among these 13 796 transcripts, a total of 44% represented ncRNAs (Figure 1A and Figure IIB in the online-only Data Supplement). After the exclusion of repetitive sequences, we identified a total of 406 novel transcripts that were not annotated previously in any database (83 were promoter antisense, 188 were intergenic, and 135 were anti-sense gene sequences).<sup>14–17</sup> By considering both annotated and nonannotated transcripts, we observed that one of the largest groups differentially regulated during differentiation was lncRNAs (≈1924). These observations therefore suggest that lncRNAs have a functional involvement during mesoderm development and cell fate specification.<sup>5,9,10,18</sup>

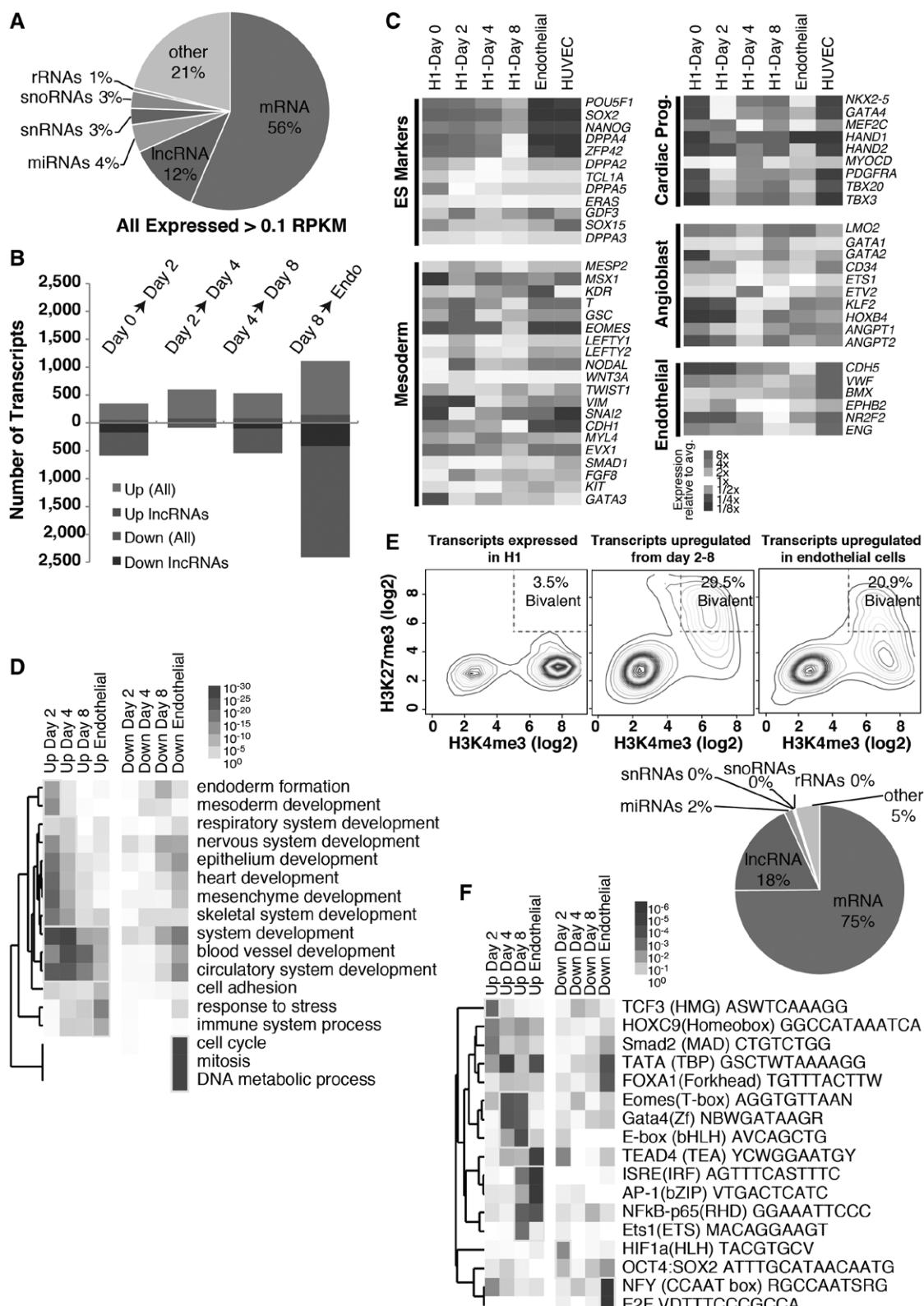
### Noncoding RNA Regulation Accounts for the Majority of Transcriptomic Changes Regulating Lineage Specification

Next, we focused our attention on the analyses of each respective cell lineage generated during differentiation. Similar to recent reports,<sup>19,20</sup> we validated the robustness of our system by focusing our attention on genes characteristic of each of the different stages analyzed: pluripotent cells (day 0; *POU5F1* [*Oct4*], *NANOG*, *SOX2*, *ZFP42*),<sup>13,19</sup> mesoderm-committed cardiovascular progenitor cells (day 2–4; *T* [*Brachyury*], *MSX1*, *GSC*, *EOMES*, *WNT3A*, *SNAI2*, *EVXI*),<sup>9,10,13</sup> committed vascular progenitor cells (day 4–8; *GATA1*, *GATA2*, *LMO2*, *ETSI*, *HOXB4*),<sup>13,21</sup> and endothelial cells (*CDH5*,

*VWF*, *EPHB2*, *NRF2F2*, *ENG*)<sup>13</sup> (Figure 1C). As expected, all genes appeared upregulated at the corresponding developmental stages (Figure 1C). Gene ontology analysis highlighted the expression of key genetic circuitries driving cardiovascular commitment (upregulated at day 2) and blood vessel and heart development (upregulated from day 2–8; Figure 1D). These results are in agreement with previous reports and further demonstrate that early cardiac and vascular developmental programs share common genetic pathways before further cell-type specification in humans.<sup>21,22</sup> Chromatin modification analysis identified an increasing number of bivalent sites as differentiation progressed, for both protein-coding and lncRNA sites (Figure 1E and Figure IID in the online-only Data Supplement). Differentiation to endothelial cells resulted in the upregulation of >1000 genes and the downregulation of ≈2500 genes (Figure 1B). Genome-wide methylation profiling highlighted significant hypomethylation at lncRNA sites (19%; Figure IIE in the online-only Data Supplement). Together, these results suggested an important regulatory role for lncRNAs during lineage specification (eg, compared with 5% for microRNAs; Figure IIE in the online-only Data Supplement).

### Identification of ncRNAs Differentially Regulated During Human Pluripotent Stem Cell Differentiation to Vascular Endothelial Cells

To identify enriched DNA motifs acting as regulatory elements, we used HOMER, an analytic tool suitable for promoter enrichment analyses.<sup>15</sup> Promoter motif enrichment (−300, +50 bp from the transcriptional start site) revealed that transcripts differentially expressed during differentiation harbor binding sites for the major transcriptional networks regulating vascular development compared with nonregulated promoters (Figure 1F). Key developmental drivers differentially regulated in a time-dependent manner included members of the homeobox gene family, mesodermal regulators (*EOMES*, *GATA*, *SMAD*), and vascular transcription factors (*ETS*, *HIF*; Figure 1F). Interestingly, lncRNAs made up the largest class of noncoding transcripts subjected to strict temporal regulation patterns during differentiation. These results are in line with the notion that lncRNAs are more tightly controlled in terms of timing and cell-type specificity than protein-coding transcripts.<sup>18,23</sup> We then selected 4 distinct groups of lncRNAs, corresponding to the 4 top-level clusters, highlighted by hierarchical clustering (Figure IIIA in the online-only Data Supplement). These 4 distinct groups had the highest average dissimilarity between one another and correlated with the different developmental stages analyzed, with ≈300 lncRNAs expressed specifically in pluripotent stem cells, ≈100 in mesodermal progenitors (day 0–2), ≈250 in early cardiovascular progenitor cells (day 2–4), ≈550 in late vascular progenitors (day 4–8), and finally, ≈200 in differentiated endothelial cells (Figure IIIA in the online-only Data Supplement). Furthermore, antisense lncRNA expression demonstrated a positive correlation with the respective sense protein-coding counterparts in many cases (Figure IIIB in the online-only Data Supplement). We next used a stringent 5-step filtering process at each indicated differentiation stage to identify novel lncRNAs, the sequence and functionality of



**Figure 1.** Transcriptome kinetics during human embryonic stem cell differentiation to endothelial cells. **A**, Annotation of transcripts expressed  $>0.1$  reads per kilobase per million (RPKM) during differentiation of H1 human embryonic stem cells into endothelial cells. **B**, Total genes regulated  $>3$ -fold between key stages of vascular differentiation. Fractions attributed to IncRNA regulation are shown in dark gray and black. **C**, Dynamics of key lineage-restricted markers during vascular differentiation from human embryonic stem (ES) cells to endothelial cells. **D**, Gene ontology enrichment for biological processes regulated at each specific stage of differentiation. **E**, Contour plots depicting the relative fraction of genes presenting bivalent methylation marks (H3K4me3 and H3K27me3) at promoters of each gene (ChIP-Seq reads [ $\log_2$ ] from  $-2\text{kb}$ ,  $+2\text{kb}$  from the GENCODE TSS). Bivalent genes are defined as those with  $>5$  ( $\log_2$ ) normalized ChIP-Seq reads. Pie chart represents the summary of genes activated during vascular differentiation that are driven by bivalent promoters, including IncRNAs. **F**, Regulatory motif enrichment in the promoters ( $-300$  bp,  $+50$  bp) of critical genes at each stage of differentiation. HUVEC indicates human umbilical vein endothelial cell.

which might be conserved across different vertebrate species (Figure IIIA in the online-only Data Supplement). To do this, we focused on the following parameters: (1) cell stage–specific patterns of expression distinguishing pluripotent cells, cardiovascular progenitor cells, and terminally differentiated endothelial cells in multiple biological replicates (at least 3); (2) significant sequence conservation across different vertebrate species; (3) exon-intron structure data; (4) availability of expressed sequenced tags in human, mouse, and zebrafish to evaluate coding potential and facilitate expression analyses; and (5) physical location in close proximity to known cardiovascular regulatory elements. Our analyses highlighted a total of 75 lncRNAs that successfully fulfilled the first 3 criteria (ie, specific expression, exon-intron structure, partial sequence conservation across vertebrates) and at least 1 of the latter (eg, expressed sequenced tag data in lower vertebrates). Next, we randomly selected 3 previously uncharacterized lncRNAs specifically expressed in pluripotent stem cells (*TERMINATOR*), vascular progenitors (*ALIEN*), and mature endothelial cells (*PUNISHER*) for further characterization (names were retrospectively assigned on the basis of the phenotypes elicited in zebrafish). ncRNA identification relies, to a great extent, on known sequence-defining gene characteristics (eg, well-defined promoters, exon-intron structure, codon conservation, and ribosomal footprints), as well as a lack of coding potential.<sup>6,24</sup> None of these transcripts possessed coding potential, as determined by GeneID analysis<sup>24</sup> (Figure IIIC in the online-only Data Supplement), as well as a lack of identifiable amino acid domains and lack of codon conservation across evolution or ribosomal footprints (data not shown).

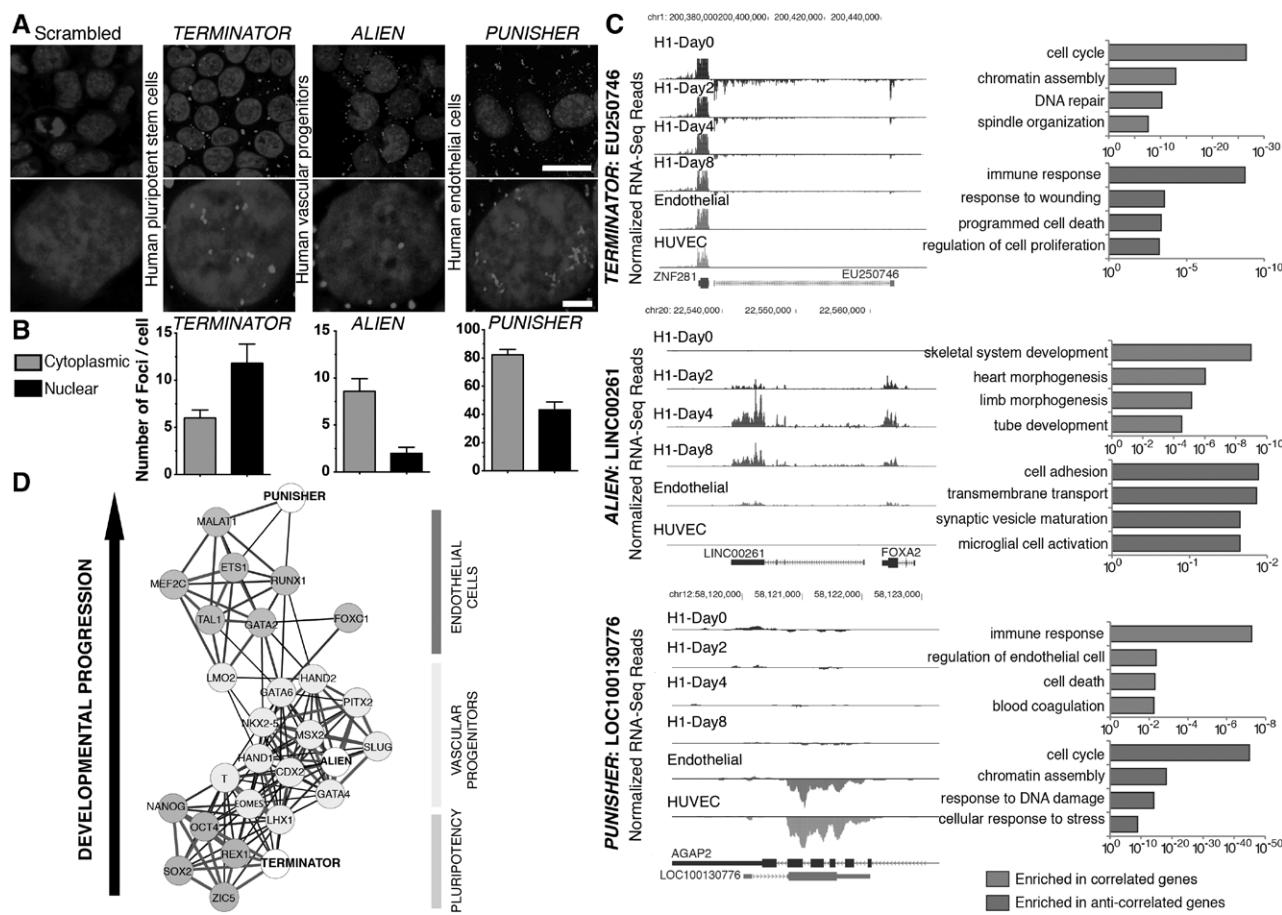
### Characterization of 3 Novel lncRNAs Differentially Expressed During Endothelial Cell Differentiation

Next, we sought to determine the functionality of the 3 different lncRNAs selected. To do so, we first focused on localization studies in the different cell lineages in which the 3 lncRNAs were identified as differentially regulated (pluripotent cells, cardiovascular progenitors, and endothelial cells). Fluorescence *in situ* hybridization experiments<sup>18</sup> revealed preferential expression of *TERMINATOR* in the nucleus of pluripotent stem cells, whereas nuclear, perinuclear, and cytosolic localization of *ALIEN* and *PUNISHER* was found in cardiovascular progenitors and endothelial cells, respectively (Figure 2A and 2B). lncRNAs have been demonstrated to play important roles in gene regulation during cell fate specification and development.<sup>3,9,10,25,26</sup> To comprehensively characterize the gene networks associated with the different lncRNAs, we performed Pearson correlation analyses in which protein-coding mRNA expression was systematically evaluated and associated with each of the 3 different lncRNAs. We next focused exclusively on those transcripts with a correlation coefficient >0.85 for further gene ontology analyses. Expression of *TERMINATOR* correlated positively with genes involved in cell cycle, DNA repair, and chromatin assembly and negatively with genes involved in cell death and regulation of proliferation (Figure 2C and Figure IVA and IVB in the online-only Data Supplement). *ALIEN* expression correlated positively with transcripts involved in skeletal muscle development, heart morphogenesis, and tube formation and

correlated inversely with cell adhesion, membrane transport, and neural function related genes (Figure 2C and Figure IVA and IVB in the online-only Data Supplement), suggesting that *ALIEN* might possess a functional role during early cardiovascular development before vascular specialization. Finally, *PUNISHER* demonstrated a positive correlation with genes participating in definitive vascular development while negatively correlating with cell-cycle regulators, chromatin modifiers, and DNA damage response genes (Figure 2C and Figure IVA and IVB in the online-only Data Supplement). Next, we performed RNA immunoprecipitation coupled to mass spectrometry analysis (Table IV in the online-only Data Supplement). Protein complex analysis on pulldown led to the identification of proteins involved in RNA binding,<sup>27</sup> posttranscriptional control, and epigenetic remodeling<sup>28</sup> (Figure 2C and Table IV in the online-only Data Supplement). Spring-embedded algorithms (Figure 2D) revealed hits closely correlating with each lncRNA. *TERMINATOR* expression was associated with *POU5F1* (*Oct4*), *SOX2*, *ZIC5*, and *REX1*, all of them known regulators of pluripotency. *ALIEN* demonstrated a high degree of correlation with pivotal drivers of mesoderm and cardiovascular commitment, including *T* (*Brachyury*), *EOMES*, *MIXL1*, and *GATA4* (Figure 2D). Finally, expression of transcription factors essential for endothelial cells such as *TAL1* and *FOXC1* correlated with *PUNISHER* (Figure 2D). Together, these results indicate a stage-specific function for each of the different lncRNAs in regulating gene expression. Additionally, none of the identified lncRNAs were physically associated with polypeptides of the ribosome translational machinery. Thus, this confirms the noncoding nature of the selected transcripts.

### Novel lncRNAs Functionally Control Pluripotency, Cardiovascular Commitment, and Endothelial Cell Identity

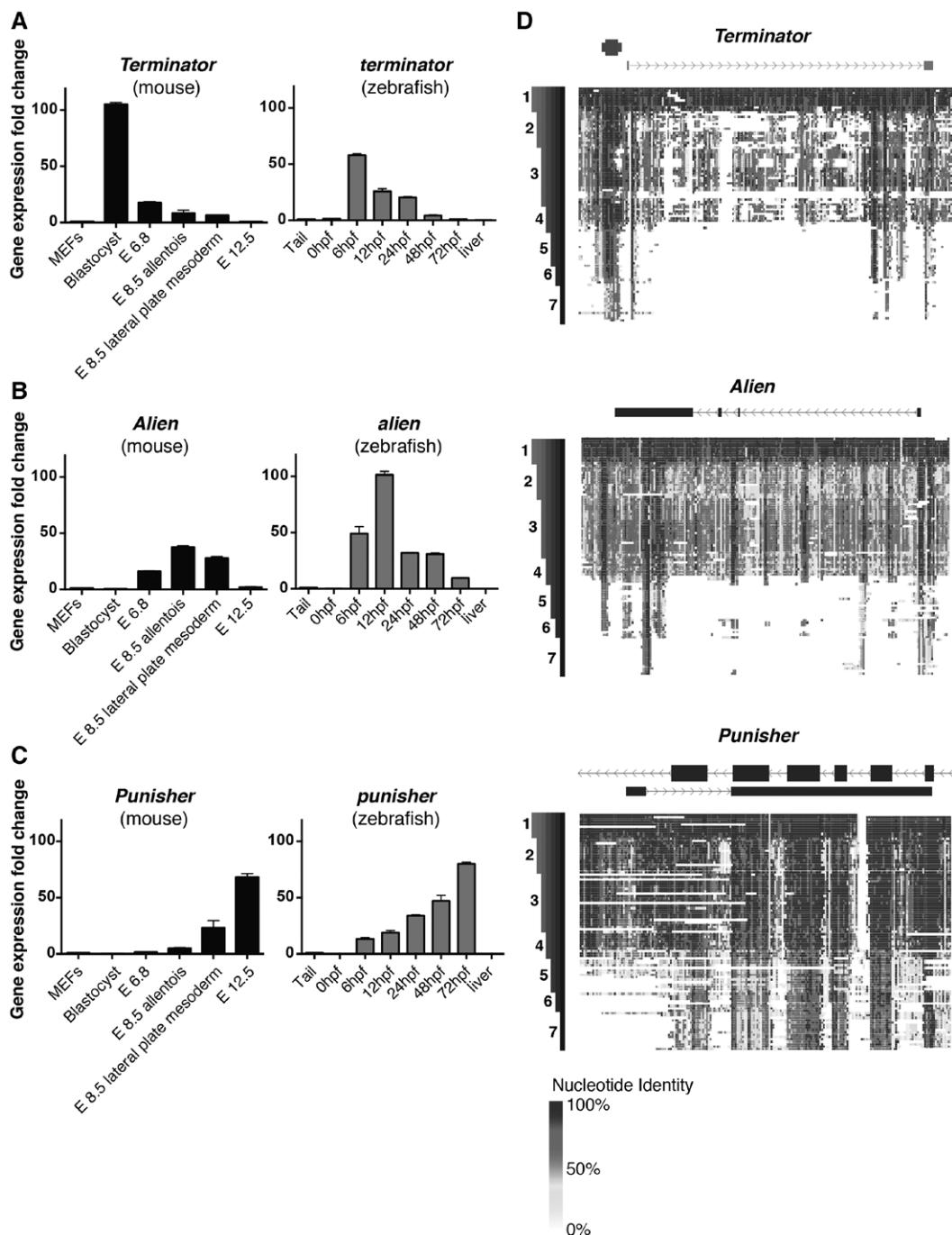
To gain insights into the physiological relevance of the identified lncRNAs, we first evaluated their expression pattern during mouse and zebrafish embryonic development. In accordance with their expression during human vascular differentiation, *Terminator* expression was maximal at the blastocyst stage in mouse and at 6 hours after fertilization in zebrafish, closely correlating with the expression of *Pou5f1* (*Oct4*) and *Nanog* (Figure 3A and Figures IIIC and V in the online-only Data Supplement). *Terminator* levels were undetectable in 1-cell-stage zebrafish embryos immediately after fertilization, suggesting that *Terminator* was not already present in the oocyte before fertilization (Figure 3A). *Alien* was expressed in the allantois and lateral plate mesoderm of E8.5 mouse embryos and zebrafish embryos 12 hours after fertilization (Figure 3B), correlating with the expression of *T* (*Brachyury*), *Meox1*, and *Mixl1* (Figure V in the online-only Data Supplement). The endothelial cell-specific lncRNA *Punisher* showed the highest levels of expression in mouse embryos at embryonic day 12.5 and in zebrafish at 72 hours after fertilization, once the vasculature was formed (Figure 3C and Figure V in the online-only Data Supplement). PhyloP analysis for detailed evolutionary conservation revealed relatively short, highly conserved sequence regions of 250 to 500 bp (>90%–95%) across



**Figure 2.** Characterization of *TERMINATOR*, *ALIEN*, and *PUNISHER*, 3 novel developmentally regulated lncRNAs. **A**, Representative images of subcellular localization of *TERMINATOR* in human embryonic stem (ES) cells, *ALIEN* in vascular progenitors, and *PUNISHER* in primary endothelial cells as determined by RNA in situ hybridization using specific locked nucleic acid (LNA) probes. A scrambled control LNA probe has been tested in all 3 different cell types with similar results (representative pictures on the left). **B**, Quantification of nuclear and cytosolic lncRNA foci ( $n \geq 5$ ). **C**, Left, RNA-Seq read density coverage along lncRNA loci. Right, Gene ontology functional enrichment analysis from all genes exhibiting expression profiles similar to *TERMINATOR* (top), *ALIEN* (middle), or *PUNISHER* (bottom). **D**, Network depicting correlated gene expression profiles of the uncharacterized lncRNAs and key developmental transcription factors. A Pearson correlation threshold of 0.85 was used to define edges in the network. Thick gray lines indicate higher correlation relative to black lines. Data are represented as mean $\pm$ SD. Scale bars: 25  $\mu$ m (A, top) and 5  $\mu$ m (A, bottom). HUVEC indicates human umbilical vein endothelial cell.

vertebrates<sup>29</sup> (Figure 3D and Table V in the online-only Data Supplement). Additionally, as previously mentioned, there were described expressed sequenced tags transcribed from the conserved loci in mouse and in zebrafish. *TERMINATOR* was found to be an intergenic lncRNA located next to *ZNF281* in human and mouse (in both cases in chromosome 1). It presents H3K27Ac sites in start and end regions (marking actively transcribed chromatin), and its genomic location is conserved in sense. *ALIEN* represents a subclass of lncRNAs, a lincRNA (long interspersed noncoding RNA). It was found located in an intergenic region with low gene density. It is encoded in sense with exons, and its location is conserved next to *FOXA2* in mouse, humans, and zebrafish. Finally, *PUNISHER* was found to be an antisense lncRNA covering exonic and intronic sequences of the *AGAP2* gene. Its location is conserved in mouse and humans. It is transcribed as an antisense to *AGAP2*, and it presents H3K27Ac sites. Taken together, our results indicate that these 3 lncRNAs are conserved from zebrafish to humans and show similar stage-specific expression during development.

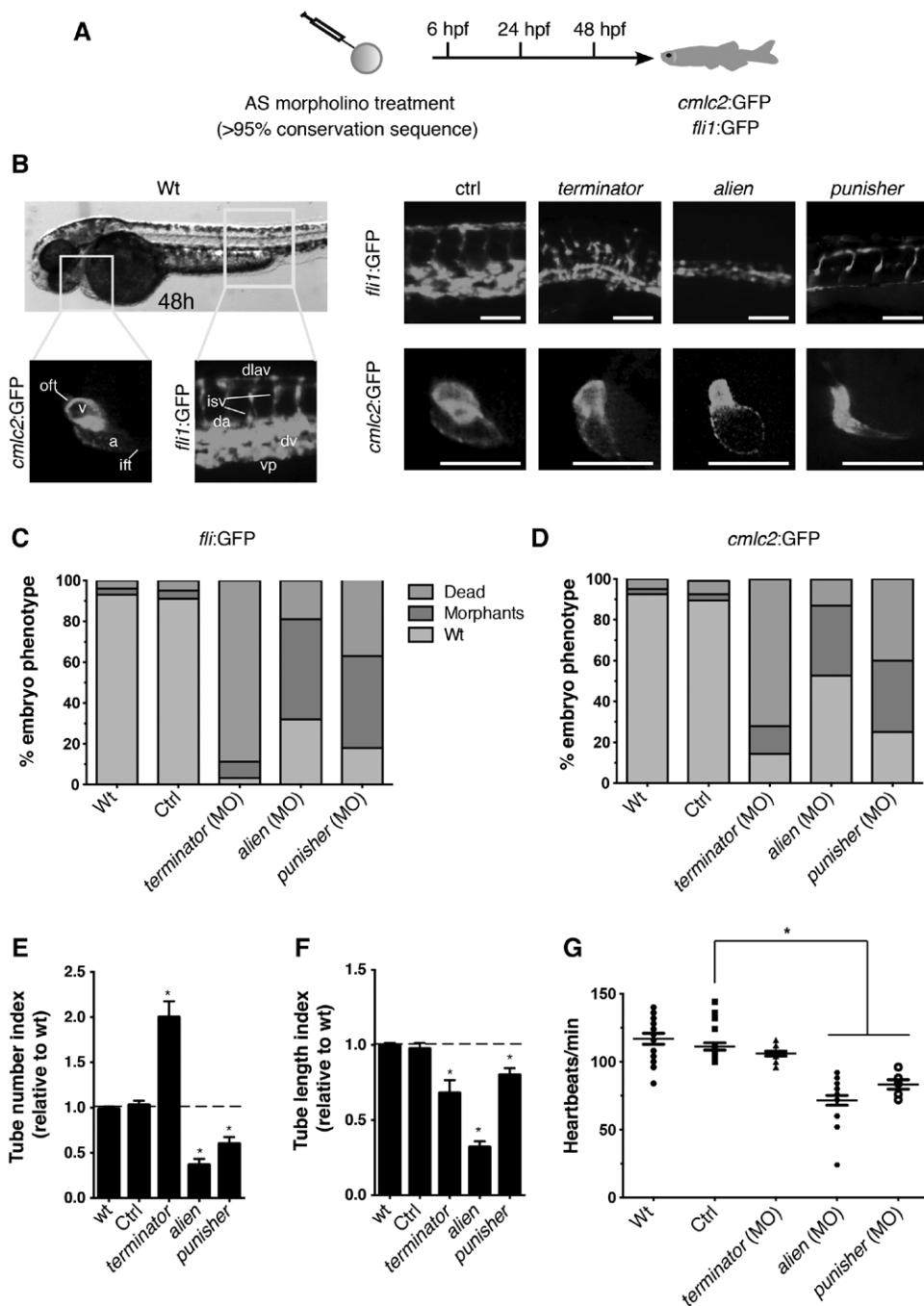
To further characterize the functional roles of the different lncRNAs during vertebrate development, we designed a series of experiments in different vertebrate systems, including zebrafish and mammalian murine embryos. Antisense morpholino oligonucleotides, synthetic molecules used to sterically block RNA binding sites in the absence of degradation, targeting highly conserved regions or putative splice sites in the identified lncRNAs were designed to induce specific knockdown or functional blockade.<sup>30,31</sup> Loss-of-function experiments were first performed in the transgenic zebrafish strains *fli1*:GFP (vascular reporter) and *cmlc2*:GFP (cardiac reporter; Figure 4A) with specific antisense morpholino and compared with the corresponding nontargeting sequences used as negative controls. Morpholino injections targeting *terminator* compromised development at the gastrulation stage and resulted in >70% lethality, whereas the surviving embryos exhibited developmental arrest and severe cardiovascular defects (Figure 4B–4D and Tables VI and VII in the online-only Data Supplement). In addition, delayed epiboly stages and detachment of the cell mass from



**Figure 3.** IncRNA sequence conservation and expression profiles through vertebrate evolution and development. **A** through **C**, Dynamic expression pattern in developing mouse and zebrafish embryos for *Terminator* (**A**), *Alien* (**B**), and *Punisher* (**C**) as determined by quantitative reverse transcription–polymerase chain reaction ( $n=5$  per group). **D**, Vertebrate genomic alignments and PhyloP conservation scores for each uncharacterized IncRNA across 44 species. See Table V in the online-only Data Supplement for the full list of animals used for PhyloP conservation analysis.<sup>1–7</sup> Data are represented as mean $\pm$ SD. E indicates embryonic day; hpf, hours postfertilization; and MEF, mouse embryonic fibroblast.

the yolk at 5 to 7 hours after fertilization could be observed early during development (Figure VI in the online-only Data Supplement), in line with peak expression of *terminator* at 6 hours after fertilization. Morpholino-mediated loss of function of *alien* led to severe impairment in multiple anatomic structures, highlighting the specific role of *alien* in mesodermal specification. Among other mesoderm-related defects, *alien* inhibition resulted in defective vascular patterning, with

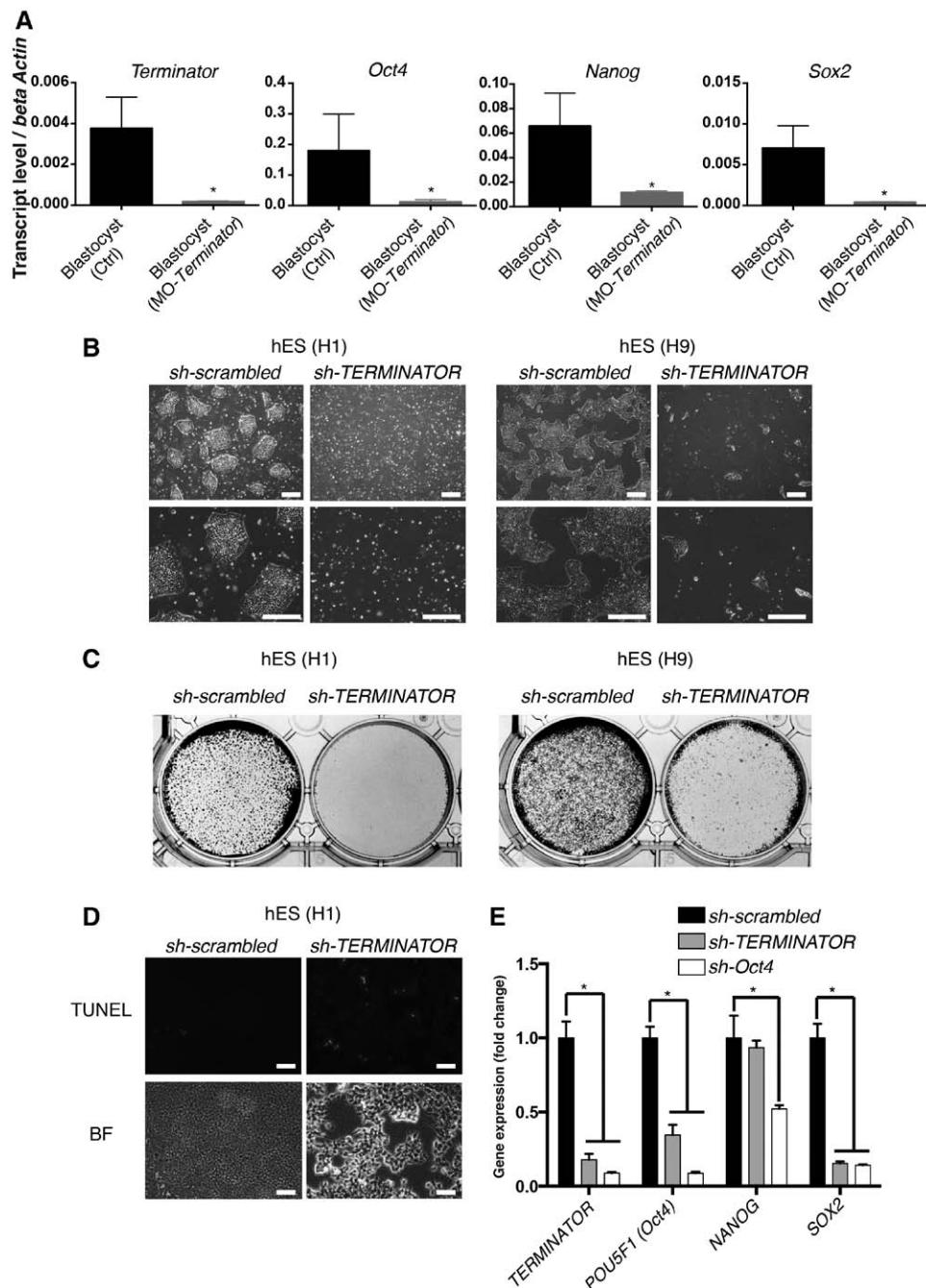
pronounced branching defects abrogating the correct formation of dorsal and intersegmental blood vessels and defective cardiac chamber formation, demonstrating that *alien* specifically functions in an early developmental progenitor stage common to both the vascular and cardiac lineages (Figure 4B–4D and Tables VI and VII in the online-only Data Supplement). Inhibition of *punisher* also resulted in severe defects in the vasculature, including defective branching and



**Figure 4.** In vivo functional evaluation of conserved lncRNAs during zebrafish development by morpholino (MO)-mediated loss of function. **A**, Morpholinos were designed to block specific highly conserved regions or putative splice sites and were injected into cardiovascular reporter zebrafish embryos. **B**, Representative phenotypes observed for the different knockdowns during vascular (*fli1:GFP*) and cardiac (*cmhc2:GFP*) development at 48 hours postfertilization (hpf). **C** and **D**, Quantification of vascular (**C**) and cardiac (**D**) phenotypes obtained for the different morpholinos ( $n > 70$ ). **E** and **F**, Analysis of defects in blood vessel formation by quantification of tube number (**E**) and tube length (**F**;  $n > 70$ ). **G**, Heart function measured as heartbeats per minute for the different morphants ( $n > 10$ ). Data are represented as mean  $\pm$  SD. Wt indicates wild-type. \* $P < 0.05$ .

compromised vessel formation (Figure 4B–4D and Tables VI and VII in the online-only Data Supplement). Abrogation of Punisher activities demonstrated significant changes in vessel tube number and length and severely impaired cardiac development and function (Figure 4E–4G and Tables VI and VII in the online-only Data Supplement). To confirm the in vivo role of terminator and punisher, we performed

rescue experiments. Coinjection of mature human lncRNA sequences alongside the respective morpholinos sufficed for rescuing the observed phenotypes with a significant effect in embryo survival, animal morphology, and branching and development of the cardiovascular system (Figure VII and Table VIII in the online-only Data Supplement). It should be noted that zebrafish morpholinos could not efficiently target



**Figure 5.** Terminator is essential for pluripotent stem cells survival. **A**, Morpholino (MO)-mediated loss of function of Terminator in mouse blastocyst-stage embryos led to significant downregulation of pluripotency factors *Oct4*, *Nanog*, and *Sox2*, indicating a fundamental role during very early development ( $n > 8$ ). **B** and **C**, Massive cell death in human embryonic stem (hES) cell lines (H1 and H9) was induced by loss of TERMINATOR, as indicated by representative bright-field pictures of sh-scrambled and sh-TERMINATOR-treated cell culture (**B**). In **C**, treated cells were stained by crystal violet. **D**, TERMINATOR knockdown induced apoptosis in human pluripotent stem cells as indicated by terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) staining. **E**, Downregulation of essential pluripotency transcription factors *Oct4* and *Sox2* on TERMINATOR knockdown ( $n=5$ ). Data are represented as mean $\pm$ SD. Scale bars, 200  $\mu$ m (**B**) and 75  $\mu$ m (**D**). \* $P < 0.05$ . BF indicates bright-field.

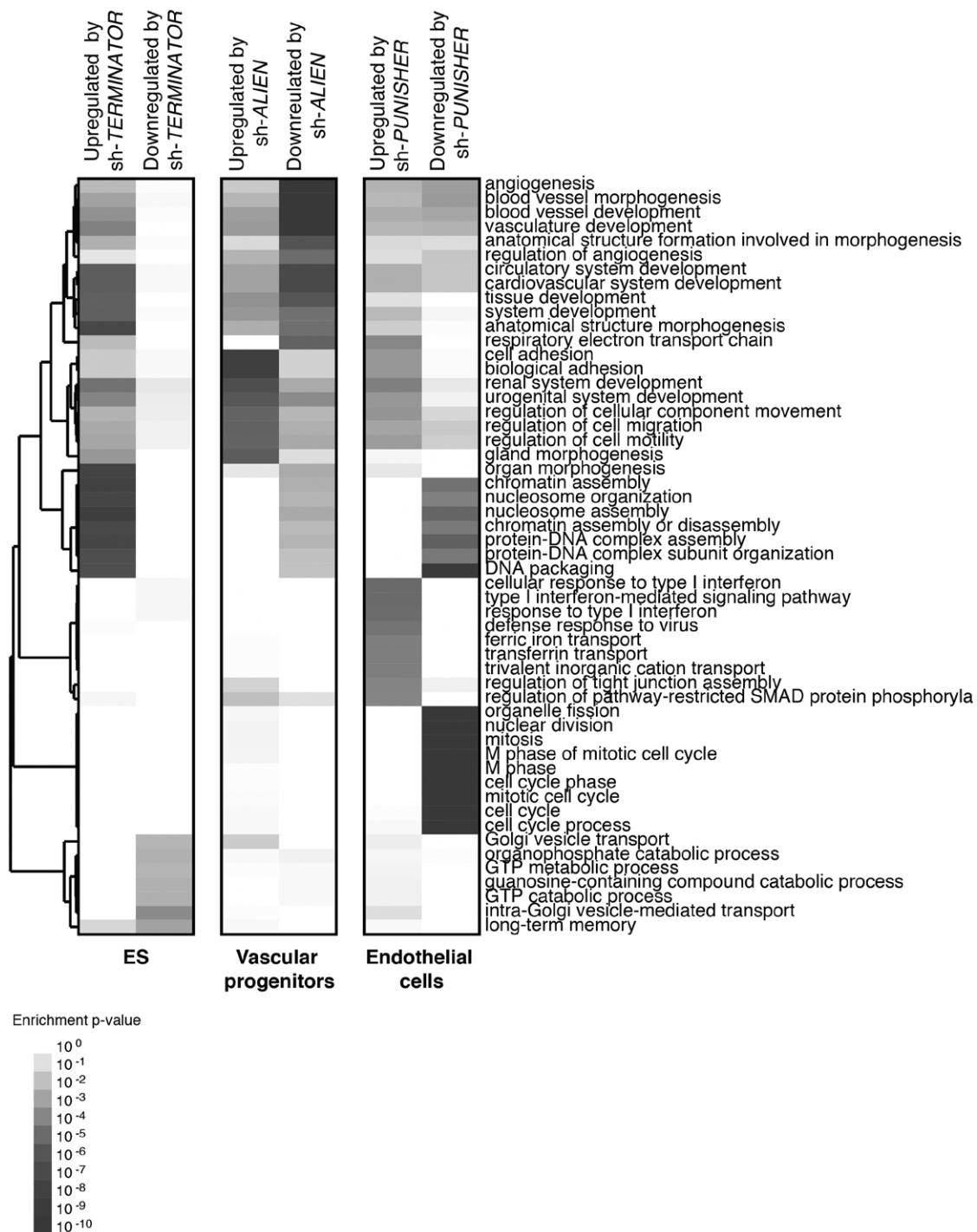
the human lncRNA counterparts because of sequence divergence. Therefore, these rescue experiments further highlighted a qualitatively specific functional role for each of the different lncRNAs. Together, our results demonstrated a specific function for the identified lncRNAs in the differentiation of pluripotent cells to mesoderm and further specification of cardiovascular followed by vascular and endothelial cell specification in vertebrates.

### Identified lncRNAs Demonstrate a Critical Role at Specific Stages During Mammalian Development

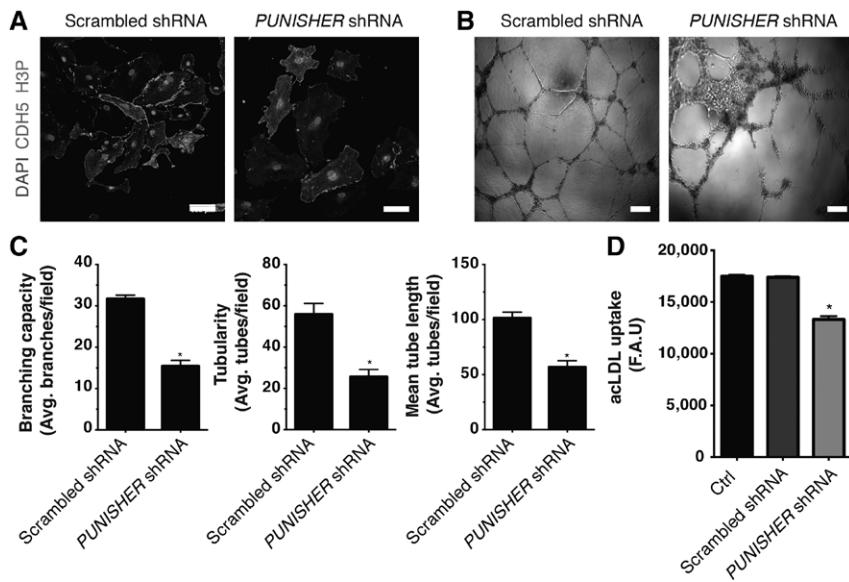
Finally, to provide an overview of the role that the identified lncRNAs play in mammalian development, we evaluated the role of all 3 lncRNAs in mammals by loss-of-function experiments in mouse embryos and human cells. To do so, 3 independent biological replicates were pooled prior to microarray hybridization. Further analysis indicated that TERMINATOR

blockade in mouse embryos and human ESCs resulted in downregulation of the pluripotency factors *POU5F1* (*Oct4*), *SOX2*, and *NANOG* at the blastocyst stage and led to significant cell death (Figure 5A–5E). Downregulation of *TERMINATOR* resulted in the upregulation of 506 genes and the downregulation of 185 different genes involved in cell-cell interactions and chromatin remodeling necessary for the maintenance of a pluripotent state (Figure 6). *ALIEN* knockdown resulted in the

significant upregulation of 738 genes involved in cell adhesion and extracellular matrix remodeling, whereas downregulation of 503 genes related to angiogenesis and blood vessel development could be observed (Figure 6). Finally, *PUNISHER* knockdown resulted in profound gene expression changes in endothelial cells, with 802 genes involved in mitosis and cell division being downregulated and 831 genes involved in cell adhesion and extracellular interactions being upregulated



**Figure 6.** Gene expression profiling of key molecular networks influenced by lncRNA knockdown during human vascular differentiation. Heat map depicting genetic networks differentially regulated on shRNA-mediated knockdown of *TERMINATOR* (human embryonic stem [ES] cells), *ALIEN* (vascular progenitors), and *PUNISHER* (human umbilical vein endothelial cells), respectively. Enriched biological processes are indicated on the right.



**Figure 7.** PUNISHER knockdown results in severe functional aberrations in human primary endothelial cells. **A**, Loss of PUNISHER in endothelial cells led to significant reduction in cell proliferation indicated by loss of phosphorylated histone H3 48 hours after knockdown. **B**, Representative bright-field pictures showing well-organized capillary-like structures formed by scramble-treated endothelial cells (**left**) as opposed to PUNISHER knockdown endothelial cells (**right**). **C**, Quantitative analysis after Matrigel tube formation assay on knockdown of PUNISHER indicated severe defects in branching (**left**), tubularity (**middle**), and mean tube length (**right**;  $n \geq 5$ ). **D**, Acetylated low-density lipoprotein (acLDL) functional assay showed significantly reduced uptake on PUNISHER knockdown ( $n=5$ ). Note that the control group represents primary endothelial cells. Data are represented as mean $\pm$ SD. Scale bars, 100  $\mu$ m (**A**) and 1 mm (**B**). \* $P < 0.05$ . F.A.U. indicates fluorescence arbitrary units.

(Figure 6). PUNISHER knockdown resulted in decreased histone H3 phosphorylation (Figure 7A), a marker indicative of mitosis, and impaired human vessel maturation (Figure 7B and 7C). PUNISHER knockdown also impaired acetyl-low-density lipoprotein uptake, a hallmark of endothelial cell functionality, therefore indicating severely impaired endothelial cell function compared with scrambled shRNA controls and unmodified endothelial cells (Figure 7D).

## Discussion

Recent discoveries indicating that the mammalian transcriptome is comprised of a large majority of noncoding transcripts ( $\approx 60\%$ ) as opposed to coding RNAs ( $\approx 7\%$ ) have brought about the question of whether noncoding sequences can play a role in controlling cellular fate and functionality, ultimately affecting biological diversity.<sup>25,32–34</sup> Despite major efforts, little information is available on the actual cellular functions of ncRNAs in the context of human development, physiology, and disease.<sup>9,10,19,35,36</sup> Among the existing information, lncRNAs identified in pluripotent cells have been ascribed mainly to gene-network interactions, the regulation of chromatin, and the control of the pluripotency network.<sup>19,33,35</sup> Similarly, other recent reports have focused on the identification of lncRNAs during cardiomyocyte generation and identified *Braveheart* and *Fenderr* as critical players during the development of the early cardiovascular system and ultimately the murine heart.<sup>9,10</sup>

Pluripotent cell differentiation has been demonstrated to be a suitable platform for recapitulating key developmental stages in a dish and for the establishment of cellular disease models.<sup>37,38</sup> Therefore, with the goal of comprehensively investigating human cardiovascular development, we analyzed human ESCs during the course of differentiation to mesoderm, cardiovascular progenitor cells, and differentiated vascular endothelial cells. By relying on our recently reported methodologies for the differentiation of human stem cells to cardiovascular progenitors and terminally differentiated vascular endothelial cells, we report here the identification and characterization of

3 novel human lncRNAs (see Table IX in the online-only Data Supplement for a summary of findings for each lncRNA). First, we found a total of 75 novel lncRNAs with different expression patterns depending on the differentiation stage. Among those lncRNAs fulfilling our preselection criteria, we next selected 3 novel noncoding transcripts (*TERMINATOR*, *ALIEN*, and *PUNISHER*) for further characterization in different vertebrate models. We found that *TERMINATOR* specifically controls pluripotent stem cell identity, *ALIEN* impairs cardiovascular development, and *PUNISHER* compromises endothelial cell function. The differentiation platform used, however, presents several limitations. As with most protocols on pluripotent stem cell differentiation, the generated cells generally present a fetal-like signature, thus potentially obviating lncRNAs that are specifically expressed in adult cells. In addition, in vitro differentiation protocols and genetic profiling suffer from the inherent disadvantage of providing snapshots of what is otherwise a continuous process. Thus, we cannot rule out that, despite collecting samples every second day, other more tightly regulated lncRNAs are underrepresented in our data sets. Despite these limitations and considering the role that the identified lncRNAs play during cardiovascular development, it is tempting to speculate that deregulation of lncRNAs affecting early embryonic development might be one of the causes of congenital cardiovascular diseases and malformations. Indeed, our work indicates that deregulation of lncRNAs and the resulting changes in gene expression have profound implication in embryonic development and cardiovascular system formation and function across different vertebrate species. Although the implications of lncRNAs in human disease remain obscure and we are only now starting to unravel the complex role that non-coding genetic elements play in the context of disease, it seems clear that leveraging next-generation sequencing technologies with patient data collection might shed new light on how aberrant lncRNA expression functionality correlates with disease. Understanding the precise role of ncRNAs during development and disease might ultimately open new avenues for the development of human therapeutics.

Our work establishes an analytic pipeline for the systematic study of lncRNAs in cardiovascular development and demonstrates that appropriate in vitro systems can be used to identify novel players controlling lineage commitment and human development. Ultimately, broad-scale application of genomic strategies based on the use of pluripotent cells might shed new light on the fundamental mechanisms underlying human development and disease.

### Acknowledgments

We thank M. Schwarz, P. Schwarz, I. Dubova, M. Marti, C. Fabregat, and D. Okamura for administrative and technical support. We also thank Ruoping Chen and the Beijing Genomics Institute for sequencing services and Nasun Hah for valuable advice.

### Sources of Funding

Work in Dr Yeo's laboratory was supported by grants from the National Institutes of Health (U54 HG007005, R01 HG004659, and R01 NS075449) and the California Institute for Regenerative Medicine (RB3-05009). Work in Dr Izpisua Belmonte's laboratory was supported by grants from the National Institutes of Health (5U01HL107442), CIRM, the G. Harold and Leila Y. Mathers Charitable Foundation, and the Leona M. and Harry B. Helmsley Charitable Trust (2012-PG-MED002). Dr Kurian was partially supported by the California Institute for Regenerative Medicine (CIRM). Dr Nivet was partially supported by an F.M. Kirby Foundation postdoctoral fellowship. Dr Sancho-Martinez was partially supported by a Nomis Foundation postdoctoral fellowship.

### Disclosures

None.

### References

- Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, Dunham I, Elnitski LL, Farnham PJ, Feingold EA, Gerstein M, Giddings MC, Gilbert DM, Gingeras TR, Green ED, Guigo R, Hubbard T, Kent J, Lieb JD, Myers RM, Pazin MJ, Ren B, Stamatoyannopoulos JA, Weng Z, White KP, Hardison RC. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA*. 2014;111:6131–6138. doi: 10.1073/pnas.1318948111.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
- Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011;12:861–874. doi: 10.1038/nrg3074.
- Atkinson SR, Marguerat S, Bähler J. Exploring long non-coding RNAs through sequencing. *Semin Cell Dev Biol*. 2012;23:200–205. doi: 10.1016/j.semcdb.2011.12.003.
- Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell*. 2013;152:1298–1307. doi: 10.1016/j.cell.2013.02.012.
- Derrien T, Guigó R, Johnson R. The long non-coding RNAs: a new (p)layer in the “dark matter.” *Front Genet*. 2011;2:107. doi: 10.3389/fgene.2011.00107.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915–1927. doi: 10.1101/gad.17446611.
- Yang L, Froberg JE, Lee JT. Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem Sci*. 2014;39:35–43. doi: 10.1016/j.tibs.2013.10.002.
- Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, Herrmann BG. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell*. 2013;24:206–214. doi: 10.1016/j.devcel.2012.12.012.
- Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 2013;152:570–583. doi: 10.1016/j.cell.2013.01.003.
- Paralkar VR, Mishra T, Luan J, Yao Y, Kossenkov AV, Anderson SM, Dunagin M, Pimkin M, Gore M, Sun D, Konuthula N, Raj A, An X, Mohandas N, Bodine DM, Hardison RC, Weiss MJ. Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood*. 2014;123:1927–1937. doi: 10.1182/blood-2013-12-544494.
- Marques AC, Ponting CP. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol*. 2009;10:R124. doi: 10.1186/gb-2009-10-11-r124.
- Kurian L, Sancho-Martinez I, Nivet E, Aguirre A, Moon K, Pendares C, Volle-Challier C, Bono F, Herbert JM, Pulecio J, Xia Y, Li M, Montserrat N, Ruiz S, Dubova I, Rodriguez C, Denli AM, Boscolo FS, Thiagarajan RD, Gage FH, Loring JF, Laurent LC, Izpisua Belmonte JC. Conversion of human fibroblasts to angioblast-like progenitor cells. *Nat Methods*. 2013;10:77–83. doi: 10.1038/nmeth.2255.
- Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:i1–i13. doi: 10.1093/nar/gkn923.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murru C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–589. doi: 10.1016/j.molcel.2010.05.004.
- Saldanha AJ. Java Treeview: extensible visualization of microarray data. *Bioinformatics*. 2004;20:3246–3248. doi: 10.1093/bioinformatics/bth349.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–2504. doi: 10.1101/gr.1239303.
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10:155–159. doi: 10.1038/nrg2521.
- Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, Zhang X, Coyne M, Fostel JL, Holmes L, Meldrim J, Guttman M, Epstein C, Park H, Kohlbacher O, Rinn J, Gnirke A, Lander ES, Bernstein BE, Meissner A. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell*. 2013;153:1149–1163. doi: 10.1016/j.cell.2013.04.037.
- Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, Yang H, Wang T, Lee AY, Swanson SA, Zhang J, Zhu Y, Kim A, Nery JR, Urich MA, Kuan S, Yen CA, Klugman S, Yu P, Suknuntha K, Propson NE, Chen H, Edsall LE, Wagner U, Li Y, Ye Z, Kulkarni A, Xuan Z, Chung WY, Chi NC, Antosiewicz-Bourget JE, Slukvin I, Stewart R, Zhang MQ, Wang W, Thomson JA, Ecker JR, Ren B. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*. 2013;153:1134–1148. doi: 10.1016/j.cell.2013.04.022.
- Bondue A, Lapouge G, Paulissen C, Semeraro C, Iacovino M, Kyba M, Blanpain C. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell*. 2008;3:69–84. doi: 10.1016/j.stem.2008.06.009.
- Bondue A, Tänner S, Chiapparo G, Chabab S, Ramialison M, Paulissen C, Beck B, Harvey R, Blanpain C. Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation. *J Cell Biol*. 2011;192:751–765. doi: 10.1083/jcb.201007063.
- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Soldà G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*. 2008;18:1433–1445. doi: 10.1101/gr.078378.108.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22:1775–1789. doi: 10.1101/gr.132159.111.
- Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol*. 2013;20:300–307. doi: 10.1038/nsmb.2480.
- Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet*. 2011;12:136–149. doi: 10.1038/nrg2904.
- Gong C, Popp MW, Maquat LE. Biochemical analysis of long non-coding RNA-containing ribonucleoprotein complexes. *Methods*. 2012;58:88–93. doi: 10.1016/j.yjmeth.2012.06.020.

28. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007;129:1311–1323. doi: 10.1016/j.cell.2007.05.022.
29. Rosenblom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* 2013;41(Database issue):D56–D63. doi: 10.1093/nar/gks1172.
30. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell.* 2011;147:1537–1550. doi: 10.1016/j.cell.2011.11.055.
31. Sarma K, Levasseur P, Aristarkhov A, Lee JT. Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proc Natl Acad Sci USA.* 2010;107:22196–22201. doi: 10.1073/pnas.1009785107.
32. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA.* 2009;106:11667–11672. doi: 10.1073/pnas.0904715106.
33. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature.* 2011;477:295–300. doi: 10.1038/nature10398.
34. Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R. Long noncoding RNAs with enhancer-like function in human cells. *Cell.* 2010;143:46–58. doi: 10.1016/j.cell.2010.09.001.
35. Loewer S, Cabilio MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, Manos PD, Datta S, Lander ES, Schlaeger TM, Daley GQ, Rinn JL. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet.* 2010;42:1113–1117. doi: 10.1038/ng.710.
36. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbrecht C, Wang P, Wang Y, Kong B, Langerød A, Børresen-Dale AL, Kim SK, van de Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ, Chang HY. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet.* 2011;43:621–629. doi: 10.1038/ng.848.
37. Daley GQ. The promise and perils of stem cell therapeutics. *Cell Stem Cell.* 2012;10:740–749. doi: 10.1016/j.stem.2012.05.010.
38. Robinton DA, Daley GQ. The promise of induced pluripotent stem cells in research and therapy. *Nature.* 2012;481:295–305. doi: 10.1038/nature10761.

## CLINICAL PERSPECTIVE

Long noncoding RNAs (lncRNAs) are a new class of regulatory RNA molecules able to modulate diverse processes such as development, pluripotency, and disease by altering posttranscriptional and posttranslational regulation, recombination, protein complex formation, and cell-cell signaling. Accordingly, it is clear that identification and characterization of novel lncRNAs regulating human development are of major interest. Human pluripotent stem cells have the capacity to generate any given somatic cell lineage on differentiation. *In vitro* differentiation recapitulates key developmental stages, thus allowing comprehensive studies on early human development and disease. Leveraging methodologies for the differentiation of cardiovascular lineages from human pluripotent stem cells in combination with next-generation sequencing, we provide here the transcriptomic changes underlying vertebrate cardiovascular development. In addition, we provide an analytic framework for the identification of novel genetic elements underlying the development of the cardiovascular system in multiple vertebrate species. Our work indicates that deregulation of specific lncRNAs and the resulting changes in gene expression may have profound implications in embryonic development and cardiovascular system formation and function across different vertebrate species. The combination of next-generation sequencing technologies with patient data collection might shed new light on how aberrant lncRNA expression correlates with disease. A better understanding of the precise role of noncoding RNAs during development and disease might ultimately open new avenues for the development of human therapeutics specifically targeting this previously overlooked family of genes.

## **SUPPLEMENTAL MATERIAL**

Expanded Methods, 9 Supplemental Tables and Supplemental Table Legends, 7  
Supplemental Figures and Figure Legends.

## Expanded Methods

*Animals.* A total of 8 different developing murine embryos per experimental group were used in our experiments. Zebrafish experiments included a minimum of 20 different animals per experimental group. Mice and zebrafish were housed in an AAALAC (Association for Assessment and Accreditation of Laboratory Animal Care) -accredited facility and in compliance with all directives related to laboratory animal protection. Mice were purchased from Charles River Laboratories, housed in air-flow racks in a restricted-access area and maintained on a 12-h light/dark cycle at a constant temperature ( $22 \pm 1^{\circ}\text{C}$ ). AB (wild-type) zebrafish, *cmlc2:GFP* and *fli1:GFP* zebrafish were maintained at  $28.5^{\circ}\text{C}$  and 70% humidity in the Salk fish facilities. Experiments were conducted with the approval of the Salk Institute and the Institutional Animal Care and Use Committee (IACUC).

*DNA methylation analysis.* DNA was extracted from human undifferentiated ESCs (ESCs-shRNA control, ESC-shRNA *TERMINATOR*), differentiated vascular endothelial progenitor cells (D8-ESCs-shRNA control, D8-ESCs-shRNA *ALIEN*) and primary endothelial cells (HUVEC-shRNA control and HUVECs-shRNA *PUNISHER*) as previously described<sup>1</sup> and methylation studies conducted identically as reported<sup>1</sup>. Briefly, Illumina 450K Infinium Methylation Arrays were normalized and preprocessed with Genome Studio, a proprietary Illumina software, for analysis of data generated on Illumina platforms. Probes with missing values were removed. The groups compared included human

undifferentiated ESCs (ESCs-shRNA control, ESC-shRNA *TERMINATOR*), differentiated vascular endothelial progenitor cells (D8-ESCs-shRNA control, D8-ESCs-shRNA *ALIEN*) and primary endothelial cells (HUVEC-shRNA control and HUVECs-shRNA *PUNISHER*).

*Data analysis.* Principal Component Analysis (PCA) was performed using R function “prcomp” ([www.r-project.org](http://www.r-project.org)). PCA is a method for identifying the primary sources of variation between gene expression profiles within a dataset. The original data can then be re-plotted using coordinates derived from the PCA analysis to help display the relative similarity between each of the samples. Gene Ontology (GO) enrichment analysis was performed using DAVID (Database for Annotation, Visualization and Integrated Discovery), a web-based platform for gene set enrichment testing<sup>2</sup>. DAVID assesses the enrichment between a set of co-regulated genes and other sets of genes grouped by common functions. In this study we used the biological process gene sets defined by the Gene Ontology Consortium ([geneontology.org](http://geneontology.org)). Gene set enrichment is scored using the Fisher Exact test, which incorporates the number of genes that are common between regulated and functional groups, the total number of genes in each group, and the total number of genes considered in the analysis. Promoter motif enrichment was performed using HOMER (Hypergeometric Optimization of Motif EnRichment), a de novo motif discovery algorithm that uses hypergeometric scoring to identify enriched DNA motifs in DNA sequences<sup>3</sup>. Promoters from co-regulated genes (-300 to +50 bp from the TSS) are differentially compared to

non-regulated promoters to identify significant regulatory elements. Java TreeView<sup>4</sup> (<http://jtreeview.sourceforge.net/>) and Cytoscape<sup>5</sup> (<http://www.cytoscape.org/>) are free open source academic tools. Clustering of GO terms and motif enrichment results was performed with Cluster 3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>) and visualized with Java TreeView. Correlations between lncRNAs and transcription factor expression profiles were visualized using Cytoscape.

Similar to previous publications, gene expression data from the RNA-seq experiments was normalized by the total number of uniquely alignable reads as FPKM (fragments per kilobase per million reads mapped)<sup>6,7</sup>.

To analyze the number of reads and evaluate differential expression, samples were randomized and differentially expressed genes defined by transcript changes of more than 3-fold (FPKM) with a minimum level of expression of 0.1 FPKM. FPKM reads were compared by using DEseq<sup>8</sup> which allows for statistical analysis of small sample volumes. Briefly, DEseq addresses the restrictive assumptions of Poisson distributions by modeling count data (FPKM) with negative binomial (NB) distributions. Of note, all different technical replicates were pooled prior further analyses as described<sup>9,10</sup>. Independent biological replicate FPKM values were modeled by DEseq as indicated above and expressed as Mean with standard deviation or SEM as indicated. Hierarchical clustering was performed using average linkage using uncentered correlation as the distance metric.

*Short Hairpin RNA (shRNA) knockdowns and lentiviral transductions.* Briefly, shRNAs (for seed sequences please refer to Table 2 in the online-only Data Supplement) where cloned into the PLKO1 vector (Sigma Aldrich). The indicated cell types where then infected with the corresponding lentiviral particles for 12 hours, allowed to rest for 24 hours and puromycin-selected with 2 $\mu$ g/ml puromycin for 36 hours. In the case of *TERMINATOR*, results were evaluated after 15 hours of selection. For *ALIEN*, the knockdown lines where further differentiated to angioblasts and evaluated on day 0, day 2, day 4 and day 8 of differentiation. For *PUNISHER*, the knockdown lines where generated directly in HUVECs.

*Endothelial cell function assays.* For Acetylated Low-Density-Lipoprotein (Ac-LDL) uptake, 80% confluent endothelial cells were incubated with 10 $\mu$ g/ml Ac-LDL (L23380, Molecular Probes) for 3 hours in DMEM:F12. The cells were washed three times with PBS, dissociated using TrypLE and analyzed by flow cytometry. To assess the formation of capillary-like structures, a suspension of 4x10<sup>5</sup> endothelial cells/ml in the presence EBM-2/EGM-2 was prepared. Subsequently, 100 $\mu$ l per well was dispensed on flat-bottom 96-well plates coated with Matrigel (BD). Tube formation was observed after 24 hours of incubation, and a minimum of three replicates per experiment were analyzed.

*RNA immunoprecipitation (RIP)/mass spectrometry.* Generation of biotinylated RNA for the lncRNAs (*TERMINATOR*, *ALIEN* and *PUNISHER*, identified by their

specific expression in undifferentiated pluripotent cells, cardiovascular progenitors and committed vascular endothelial cells) was carried out with amplified cDNAs prepared from the respective cell types and cloned into pENTR/D-TOPO. RNAs were transcribed *in vitro* with the Megascript *in vitro* transcription kit (Ambion) along with control mRNA XEF. *In vitro* transcribed RNA was DNase1 treated and purified using TRIzol (Invitrogen). Purified RNA was ligated at the 3' end with a single biotinylated cytidine residue (3'End biotinylation kit BK, Pierce). *In vitro* transcribed 3' biotinylated lncRNAs (50pmols) were denatured and re-folded by heating to 65°C for 5 minutes and cooled down at room temperature in 10mM HEPES and 10mM MgCl<sub>2</sub>. Lysates were prepared from the respective cell types (10<sup>7</sup> ES cells for TERMINATOR, mesoderm progenitors for ALIEN and endothelial cells for PUNISHER) by re-suspending the cells in 1ml RIP buffer (150mM KCL, 25mM Tris pH 7.4, 0.5mM DTT, 0.5% NP40, 1mM Phenylmethylsulfonyl fluoride (PMSF), EDTA minus protease inhibitor cocktail (Roche) and 20U/ml RNaseOUT (Invitrogen)) followed by homogenization by 30 strokes using a dounce homogenizer. The lysates were centrifuged at 13,000rpm for 10 minutes at 4°C and pre-cleared using 40µl of pre-equilibrated Streptavidin-coupled Dynabeads (Invitrogen) for 20 minutes at 4°C. After pre-clearing, the lysates were mixed with 20µg/ml yeast tRNA and incubated for 20 minutes at 4°C, and then mixed with their respective transcripts and incubated for 1 hour at room temperature. 60µl of the pre-equilibrated streptavidin-coupled Dynabeads were mixed with the lysate-transcript mix and incubated for another hour at room temperature followed by 4 washes using RIP

buffer. The proteins bound to the RNA were directly trypsinized on the beads and submitted for mass spectrometry.

*Gene expression microarray analysis.* The following groups of samples were analyzed: human undifferentiated ESCs (ESCs-shRNA control, ESC-shRNA *TERMINATOR*), differentiated vascular endothelial progenitor cells (D8-ESCs-shRNA control, D8-ESCs-shRNA *ALIEN*) and primary endothelial cells (HUVEC-shRNA control and HUVECs-shRNA *PUNISHER*). Total RNA from roughly  $1 \times 10^7$  cells of each of the above-indicated groups was isolated with TRIzol (Invitrogen). All samples were subjected to quality control processes to ensure the lack of contaminating DNA and integrity of the RNA. All the RNA samples met the following RNA quality threshold- OD<sub>260/280</sub> = 2-2.2; OD<sub>260/230</sub>  $\geq$  2.0; 28S:18S > 1.0, RIN > 7. Microarray expression profiles were obtained using the AffymetrixGeneChip Human Gene 1.0 ST Array (Affymetrix). Amplification, labeling and hybridizations were performed according to protocols from Ambion and Affymetrix. Briefly, 200ng of total RNA were used to generate cDNAs with the Ambion® WT Expression Kit (Ambion/Applied Biosystems), labeled using the WT Terminal Labeling Kit (Affymetrix), and then hybridized to Human Gene 1.0 ST for 16 h at 45°C and 60rpm in a GeneChip® Hybridization Oven 640. Following hybridization, the array was washed and stained in the AffymetrixGeneChip® Fluidics Station 450. The stained array was scanned using an AffymetrixGeneChip® Scanner 3000 7G, generating CEL files for each array. The following groups of samples were analyzed: human undifferentiated ESCs

(ESCs-shRNA control, ESC-shRNA *TERMINATOR*), differentiated vascular endothelial progenitor cells (D8-ESCs-shRNA control, D8-ESCs-shRNA *ALIEN*) and primary endothelial cells (HUVEC-shRNA control and HUVECs-shRNA *PUNISHER*). After quality control of raw data, it was background corrected, quantile-normalized and summarized to a gene-level using the robust multi-chip average (RMA)<sup>11</sup> obtaining a total of 28832 transcript clusters, excluding controls, which roughly correspond to genes. NetAffx 32 annotations, human genome 19 built, were used to summarize data into transcript clusters and to annotate analyzed data. Linear Models for Microarray (LIMMA)<sup>12</sup>, a moderated t-statistics model, was used for detecting differentially expressed genes between the conditions. Genes with a minimum fold change of 1.5 and p- value less than 0.05 were selected as significant.

*Real time-PCR analysis.* Total cellular RNA was isolated using TRIzol Reagent (Invitrogen) according to the manufacturer's recommendations. 2µg of DNase1 (Invitrogen) treated total RNA was used for cDNA synthesis using the SuperScript II Reverse Transcriptase kit for RT-PCR (Invitrogen). Real-time PCR was performed using the SYBR-Green PCR Master mix (Applied Biosystems) using 150 nM primer concentrations, 40 cycles, at the indicated annealing temperatures (Table 2 in the online-only Data Supplement) in a CX96 Real Time PCR detection system (Bio-Rad). The levels of expression of the genes were normalized to corresponding *GAPDH* values (for the respective species) and are shown as fold change relative to the value of the control sample. All the samples

were done in triplicate. The primers used for real time-PCR experiments are listed in Table 1 in the online-only Data Supplement.

*In situ hybridization.* Cells were fixed in 4% Paraformaldehyde (PFA) for 10 minutes at room temperature, washed, subjected to an acetylation step and afterwards treated briefly with proteinase K (Life Technologies). Samples were then prehybridized for 4 hours and hybridized overnight with digoxigenin (DIG)-labeled locked nucleic acids (LNA) probes for the corresponding lncRNAs (Exiqon). The next day slides were washed and immunolabeled with anti-DIG-alkaline phosphatase antibodies (1:2,000) overnight at 4°C. Alkaline phosphatase activity was detected by incubating samples in Fast Red solution (Dako) for 2 hours. Samples were then washed, counterstained with DAPI (Life Technologies) and mounted in Vecta-shield (Vector Labs) for imaging in a confocal microscope (Zeiss LSM710).

*Terminal deoxynucleotidyltransferase-mediated dUTP-biotin nick end labeling (TUNEL) assay.* Cells were seeded in 8-well chamber slides (Millipore), cultured for 48 h, and treated with the respective shRNA for 48 h. The cells were fixed in 0.14 M phosphate-buffered saline (PBS; pH 7.4) containing 4% PFA at 4°C for 20 min, washed in PBS three times for 5 min, and permeabilized in PBS containing 0.1% Triton X-100–0.1% sodium citrate at 4°C for 2 min. The cells were washed and stained, using fluorescein isothiocyanate (FITC)-labeled dUTP and terminal deoxynucleotide transferase at 37°C for 1 hour using the *in situ* cell death detection kit (Roche). The TUNEL reaction was terminated by the addition

of 2x SSC (0.3 M NaCl plus 0.03 M Na citrate) for 10 min, and the cells were counterstained with propidium iodide-RNase A solution (Roche) for 10 min at room temperature. Samples were washed with distilled water and mounted using Vectashield mounting solution (Vector Labs).

*Immunofluorescence microscopy.* Briefly, cells were washed twice with PBS and fixed using 4% PFA in PBS. After fixation, cells were blocked and permeabilized for 1 hour at 37°C with 5% BSA/5% appropriate serum/ in PBS in the presence of 0.1% Triton X-100. Subsequently, cells were incubated with the indicated primary antibody either for 1 hour at room temperature or overnight at 4°C. The cells were then washed twice with PBS and incubated for 1 hour at 37°C with the respective secondary antibodies and DAPI for counterstaining, followed by mounting in Vecta-shield. Sections were analyzed by using a confocal microscope (Zeiss LSM710).

## Supplementary References

1. Kurian L, Sancho-Martinez I, Nivet E, Aguirre A, Moon K, Pendaries C, Volle-Challier C, Bono F, Herbert J-M, Pulecio J, Xia Y, Li M, Montserrat N, Ruiz S, Dubova I, Rodriguez C, Denli AM, Boscolo FS, Thiagarajan RD, Gage FH, Loring JF, Laurent LC, Izpisua Belmonte JC. Conversion of human fibroblasts to angioblast-like progenitor cells. *Nat Methods*. 2013;10:77–83.
2. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1–13.
3. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
4. Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics*. 2004;20:3246–8.
5. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504.
6. Wang Z, Gerstein M, Snyder M. Rna-seq: A revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57–63.
7. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nat Methods*. 2008;5:621–628.
8. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
9. Peng X, Wood CL, Blalock EM, Chen KC, Landfield PW, Stromberg AJ. Statistical implications of pooling rna samples for microarray experiments. *BMC Bioinformatics*. 2003;4:26.
10. Liu Y, Koyuturk M, Maxwell S, Xiang M, Veigl M, Cooper RS, Tayo BO, Li L, LaFramboise T, Wang Z, Zhu X, Chance MR. Discovery of common sequences absent in the human reference genome using pooled samples from next generation sequencing. *BMC Genomics*. 2014;15:685.

11. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249-264.
12. Singh RP, Liu D, Chaudhari A, Cherry SR, Leahy RM, Smith DJ. Investigation of different transcript quantitation tools for high-throughput mapping of brain gene expression using voxelation. *Journal of molecular histology*. 2004;35:397-402.

**Supplemental Table 1.** List of the primers used for real time-PCR experiments.

Human	Forward	Reverse	Optimal Annealing Temperature	Final Concentration
TERMINATOR	CGGAAAACAGTGGAGGAAGA	CGTGTGAAAGACAAGGCTCA	59	150nM
ALIEN	CAATCCCCTCTGAGCATTA	GCATTTTATTTCGGCTCCA	60	150nM
PUNISHER	GTCCTCCACTCCACCTCAAA	TGAGTTCCCTGATCGTGTCCA	61	150nM
NANOG	ACAACTGGCCGAAGAACAGCA	GGTTCCCAGTCGGGTTCAC	59	150nM
<i>POU5F1 (Oct4)</i>	GGGTTTTGGGATTAAGTCTCA	GCCCCCACCCCTTGTGTT	50	150nM
SOX2	CAAAATGGCCATGCAGGTT	AGTTGGGATCGAACAAAGCTATT	61	150nM
GAPDH	GGACTCATGACCACAGTCATGCC	TCAGGGATGACCTTGCCCCACAG	60	150nM
Mouse	Forward	Reverse		
<i>Terminator</i>	GCATGAGACTAGCCGAGAGG	AGCGTTACTGCCGAAGCTA	60	150nM
<i>Alien</i>	GCTCCAAGCTGTCAAAGACC	GAGGGTGCACAGCAGGAAGAG	59	150nM
<i>Punisher</i>	CCCCCTCCCTTCTACTGTCC	CTGGCCAGGATCTGACTCTC	59	150nM
<i>Nanog</i>	CAGGTGTTGAGGGTAGCT	CGGTTCATCATGGTACAGTC	60	150nM
<i>Pou5f1 (Oct4)</i>	TCTTCCACCAAGGCCCCGGCT	TGCGGGCGGACATGGGAGATCC	61	150nM
<i>Sox2</i>	TAGAGCTAGACTCCGGCGATG	TTGCCCTAA ACAAGACACGAAA	60	150nM
<i>T (Brachyury)</i>	ACAACCACCGCTGAAATATG	CTCTCACGATGTGAATCCGAG	60	150nM
<i>Meox1</i>	GAAACCCCCACTCAGAAGATAGC	TCGTTGAAGATTGCTCAGTC	58	150nM
<i>Mixl1</i>	ACGCAGTGCTTCCAAACC	CCCGCAAGTGGATGTCTGG	60	150nM
<i>Fli1</i>	ATGGACGGGACTATTAAGGAGG	GAAGCAGTCATATCTGCCTTGG	60	150nM
<i>Cdh5</i>	CACTGTTGGGAGGCCITC	GGGGCAGCGATTCACTTTCT	61	150nM
<i>Gapdh</i>	AACTTGGCATTGTGGAAGG	ACACATTGGGGTAGGAACA	60	150nM
Zebrafish	Forward	Reverse		
<i>terminator</i>	TTGTTGGGAGAAAAACTTGC	TGCGTCTGATTGGCTGTAG	59	150nM
<i>alien</i>	CGGCTATCTCTGCTGTTCC	GGCACCCCTGACATCAACTCT	60	150nM
<i>punisher</i>	CCGGTTTGCCATCACAT	CATTCTGCAAATGCACCAG	59	150nM
<i>nanog</i>	CAAAGATGCAGAGCAGACCA	GGTCAGAGGAACCCCTTCTC	58	150nM
<i>pou5f1 (Oct4)</i>	GAGGAGCCCCCTGCCCTTATA	AAGCATGGCTGATTGATCC	60	150nM
<i>t (Brachyury)</i>	TGGAAATACGTGAACGGTGA	ATACGGGTGCTTCATCCAG	60	150nM
<i>meox1</i>	ACCAAGAGCAACTCCGAGA	TCATCTGCTCAAGTCGTG	59	150nM
<i>mixl1</i>	CCACACATCAGAACACAGC	CGTGTCCCTTGGGACTGC	60	150nM
<i>fli1</i>	GGCTCTCAACAGTGGTCTC	CACAGCTGGATCTGACCTGA	60	150nM
<i>cdh5</i>	CGACCTAAAAACCCACCTGA	CAGCTTGCAAGGACAACAA	58	150nM
<i>gapdh</i>	GATACACGGAGCACCAGGTT	GCCATCAGGTACATACACG	57	150nM

**Supplemental Table 2.** List of sequences used for the generation of morpholinos and shRNA targeting the indicated lncRNAs.

MORPHOLINOS SEQUENCES	lncRNA targeted
5'-CGCTTGTGTTCCCTCTGGCTGCTT-3'	Terminator MO1
5'-TATATGCATCCAGGCCACCATTAA-3'	Terminator MO2
5'-GGGTGCCTGAATTAAATTGCTTAAG-3'	Alien MO1
5'-GGTATTGTTGTCTCTCCGCTGCC-3'	Alien MO2
5'-ACCTGTTTGTTAATGTTCTCAAG-3'	Punisher MO1
5'-TTCGTATTTGGCACGGATCCACGAC-3'	Punisher MO2
shRNA seed SEQUENCES	lncRNA targeted
CCTAAGGTTAACGTCGCCCTCG	scrambled
GGTTTATCTACCCAGTCTTAC	TERMINATOR
GCCAGGCTTCAATGTTAATT	ALIEN
CCACTCCACCTCAAACCTTTA	PUNISHER

**Supplemental Table 3.** Number of reads and alignable reads per sample obtained in the RNA-seq analysis.

Sample	Reads Sequenced	Uniquely Alignable Reads
H1-Day0	90366017	77338128
H1-Day2	86752934	68137698
H1-Day4	94594307	78960565
H1-Day8	96559954	80564506
Endothelial	82822721	65378322
HUVEC	97606136	81544588

**Supplemental Table 4.** RIP-mass spectrometry data for proteins interacting with each of the long non-coding RNAs identified in the study.

Terminator-interacting partners	Description	Role
<i>Transcriptional regulation</i>		
DDX39B	HLA-B-Associated Transcript 1 Protein	RNA splicing
DDX5	DEAD (Asp-Glu-Ala-Asp) Box Helicase 5	RNA-dependent ATPase
DHX9	DEAH (Asp-Glu-Ala-His) Box Helicase 9	RNA binding
FBL	RRNA 2'-O-Methyltransferase Fibrillarin	Nucleolar assembly
FXR1	Fragile X Mental Retardation, Autosomal Homolog 1	Ribosome assembly, translational control
PCBP1	Poly(RC) Binding Protein 1	RNA stability
RUVBL1	RuvB-Like AAA ATPase 1	DNA helicase
SYNCRIP	Synaptotagmin Binding, Cytoplasmic RNA Interacting Protein	mRNA editing and splicing
HNRNPA0*	heterogeneous nuclear ribonucleoprotein A0	RNA binding
DDX39	DEAD (Asp-Glu-Ala-Asp) Box Polypeptide 39A	ATP-dependent RNA helicase
DHX15	DEAH (Asp-Glu-Ala-His) box helicase 15	ATP-dependent RNA helicase
SFRS1*	serine/arginine-rich splicing factor 1	RNA binding protein
U2AF1	U2 small nuclear RNA auxiliary factor 1	splicing
YBX2	Y box binding protein 2	RNA-binding protein
PCBP1	poly(rC) binding protein 1	RNA binding protein
PCBP3	poly(rC) binding protein 3	RNA binding protein
<i>Epigenetic regulation</i>		
RUVBL1	RuvB-like AAA ATPase 1	DNA binding, NuA4 histone acetyltransferase complex
RUVBL2	RuvB-like AAA ATPase 2	DNA binding, NuA4 histone acetyltransferase complex
<i>Others</i>		
DAPK1	Death-Associated Protein Kinase 1	Apoptosis
RAB1*	RAB1A, Member RAS Oncogene Family	signaling
RAN	RAN, Member RAS Oncogene Family	signaling
EFCAB4B	EF-hand calcium binding domain 4B	cytoplasmic calcium release-activated calcium Channel
PHGDH	phosphoglycerate dehydrogenase	L-serine synthesis
RASEF	RAS and EF-hand domain containing	signaling
WDR1	WD repeat domain 1	disassembly of actin filaments
YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	signaling
HSPE1	heat shock 10kDa protein 1 (chaperonin 10)	protein folding
GNB2L1	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	signaling
NME1	NME/NM23 nucleoside diphosphate kinase 1	synthesis of nucleoside triphosphates
MSH6	mutS homolog 6	DNA mismatch repair system
<i>Alien interacting partners</i>		
<i>Transcriptional regulation</i>		
DDX39B	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39B	Spliceosome RNA Helicase
YBX3	Y box binding protein 3	DNA-Binding Protein
DDX39	DEAD (Asp-Glu-Ala-Asp) Box Polypeptide 39A	ATP-Dependent RNA Helicase
HNRNPC	heterogeneous nuclear ribonucleoprotein C (C1/C2)	RNA binding protein
HNRNPD	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)	RNA binding protein
HNRNPF	heterogeneous nuclear ribonucleoprotein F	RNA binding protein
HNRNPH1	heterogeneous nuclear ribonucleoprotein H1 (H)	RNA binding protein
HNRNPU	heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	RNA binding protein
SF3B3	splicing factor 3b, subunit 3, 130kDa	RNA binding protein
SFRS1	serine/arginine-rich splicing factor 1	RNA binding protein
SFRS2	serine/arginine-rich splicing factor 2	RNA binding protein
SFRS2B	serine/arginine-rich splicing factor 8	RNA binding protein
YBX1	Y box binding protein 1	RNA-Binding Protein
YBX2	Y box binding protein 2	RNA-Binding Protein
HNRNPA1L2	heterogeneous nuclear ribonucleoprotein A1-like 2	RNA binding protein
<i>Punisher-interacting partners</i>		
<i>Transcriptional regulation</i>		
DHX9	DEAH (Asp-Glu-Ala-His) Box Helicase 9	RNA binding
FXR1	fragile X mental retardation, autosomal homolog 1	Ribosome assembly, translational control
HNRNPA3	heterogeneous nuclear ribonucleoprotein A3	RNA binding protein
HNRNPD	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)	RNA binding protein
HNRNPH1	heterogeneous nuclear ribonucleoprotein H1 (H)	RNA binding protein
HNRNPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)	RNA binding protein
HNRNPU	heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	RNA binding protein
MYO1C	myosin IC	unconventional myosin, transcription initiation
PABPC4	poly(A) binding protein, cytoplasmic 4 (inducible form)	RNA binding
RBMX	RNA binding motif protein, X-linked	RNA binding, transcriptional regulation
SFPQ	splicing factor proline/glutamine-rich	DNA and RNA binding protein
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining)	DNA binding
<i>Epigenetic regulation</i>		
NAP1L1	nucleosome assembly protein 1-like 1	chromatin organization, cell proliferation
NAP1L4	nucleosome assembly protein 1-like 4	chromatin organization, cell proliferation
NASP	nuclear autoantigenic sperm protein (histone-binding)	histone binding
NCL	nucleolin	synthesis and maturation fo ribosomes
H2AFJ*	Histone 2A family, member J	Chromatin organization
HIST1H2AA*	histone cluster 1, H2aa	chromatin organization
<i>Others</i>		
RAB15*	RAB15, member RAS oncogene family	signaling

**Supplemental Table 5.** Key listing of species groups as shown in figure 3.

1	2	3	4	5	6	7
Chimp	Chinese_tree_shrew	Pig	Elephant	Saker_falcon	American_alligator	Tetraodon
Gorilla	Squirrel	Alpaca	Cape_elephant_shrew	Peregrine_falcon	Green_seaturtle	Fugu
Orangutan	Lesser_Egyptian_jerboa	Bactrian_camel	Manatee	Collared_flycatcher	Painted_turtle	Yellowbelly_pufferfish
Gibbon	Prairie_vole	Dolphin	Cape_golden_mole	White-throated_sparrow	Chinese_softshell_turtle	Nile_tilapia
Rhesus	Chinese_hamster	Killer_whale	Tenrec	Medium_ground Finch	Spiny_softshell_turtle	Princess_of_Burundi
Crab-eating_macaque	Golden_hamster	Tibetan_antelope	Aardvark	Zebra Finch	Lizard	Burton's_mouthbreeder
Baboon	Mouse	Cow	Armadillo	Tibetan_ground_jay	X_tropicalis	Zebra_mbuna
Green_monkey	Rat	Sheep	Opossum	Budgerigar	Coelacanth	Pundamilia_nyererei
Marmoset	Naked_mole-rat	Domestic goat	Tasmanian devil	Parrot		Medaka
Squirrel_monkey	Guinea_pig	Horse	Wallaby	Scarlet_macaw		Southern_platypish
Bushbaby	Chinchilla	White_rhinoceros	Platypus	Rock_pigeon		Stickleback
	Brush-tailed_rat	Cat		Mallard_duck		Atlantic_cod
	Rabbit	Dog		Chicken		Zebrafish
	Pika	Ferret		Turkey		Mexican_tetra_(cavefish)
		Panda				Spotted_gar
		Pacific_walrus				Lamprey
		Weddell_seal				
		Black_flying-fox				
		Megabat				
		David's_myotis_(bat)				
		Microbat				
		Big_brown_bat				
		Hedgehog				
		Shrew				
		Star-nosed_mole				

**Supplemental Table 6.** Main phenotypic defects observed after injection of morpholinos (MO), driven against a conserved-binding site, in one-cell stage *fli1*:GFP zebrafish embryos. Morpholinos were designed to target *terminator*, *alien* or *punisher*. ISV: intersegmental vessels; DLAV: dorsal longitudinal anastomotic vessel; CV: caudal vein; VP: vascular plexus.

Conditions	Injection type	Main defects observed
<i>Wild-type</i>	Uninjected (n=110)	None
<i>Control</i>	MO (n=207)	None
<i>terminator</i>	MO (n=178)	Developmental arrest ,defective vessel branching
<i>alien</i>	MO (n=212)	Grave vessel branching defects, loss of ISV, DLAV, CV, VP
<i>punisher</i>	MO (n=202)	Defective branching, reduced endothelial cell number, loss of ISV, DLAV, VP, CV

**Supplemental Table 7.** Main phenotypic defects observed after injection of morpholinos (MO), driven against a conserved-binding site, in one-cell stage *cmlc2*:GFPzebrafish embryos. Morpholinos were designed to target *terminator*, *alien* or *punisher*.

Condition	Injection type	Main defect observed
<i>Control</i>	Uninjected (n=165)	None
<i>Standard</i>	MO (n=153)	None
<i>Terminator</i>	MO (n=132)	Developmental arrest, enlarged chambers, defective looping
<i>Alien</i>	MO (n=70)	Enlarged atrium, reduced ventricle, defective looping
<i>Punisher</i>	MO (n=73)	Defective looping, small chambers, cardiac arrest

**Supplemental Table 8.** Zebrafish rescue experiments for *terminator* and *punisher*. Antisense morpholinos (MO) for *terminator* or *punisher* were injected at the once-cell stage in zebrafish embryos. Morphant versus wild type-like (Wt-like) phenotypes as well as survival rates were analyzed at 24 and 72 hours post-fertilization (hpf). Rescue experiments were conducted by co-injecting the human transcript of the lncRNA with its corresponding MO.

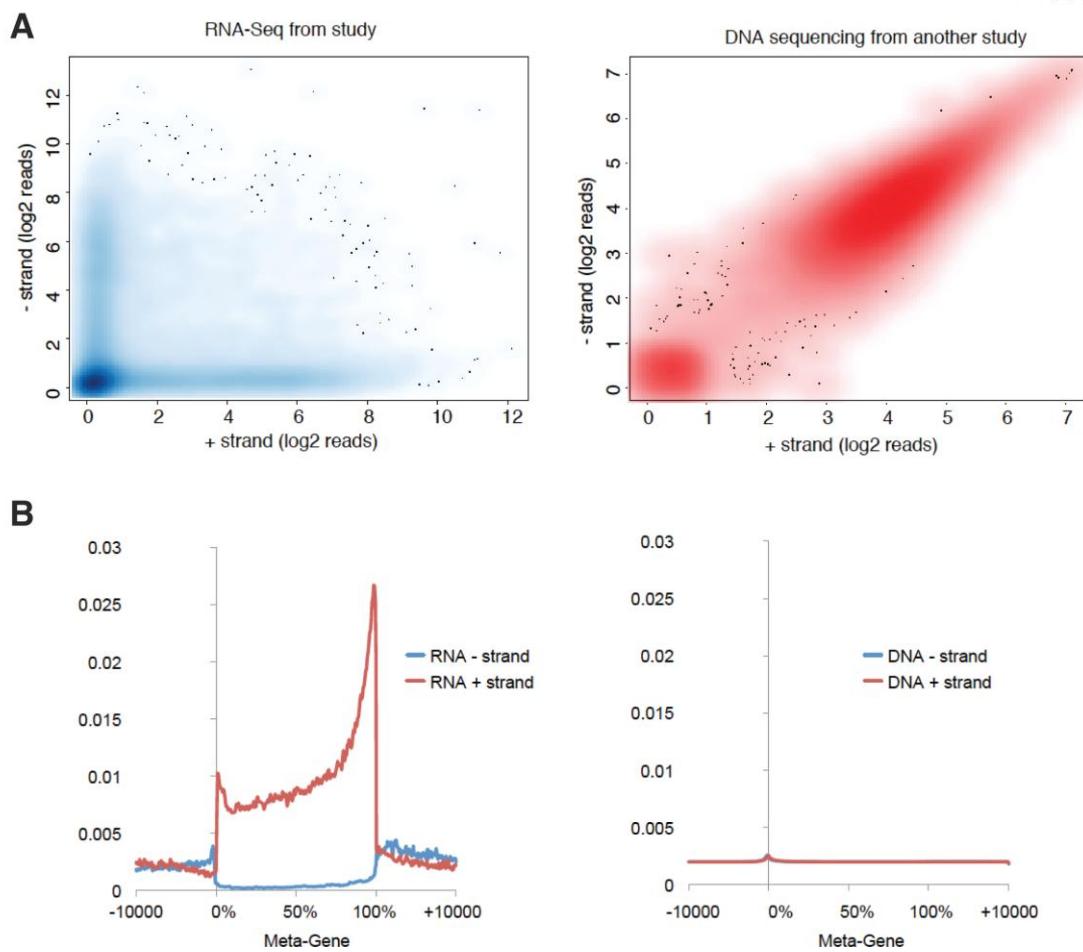
	Condition	Phenotype (%)		Survival (%)
		morphant	Wt-like	
Ctrl (24 hpf)	uninjected (n=109)	4	96	95
<i>terminator</i> (24 hpf)	MO (n=249)	81	19	55
	MO+hIncRNA (n=337)	65	35	79
Ctrl (72 hpf)	uninjected (n=157)	0	98	98
<i>punisher</i> (72 hpf)	MO (n=397)	33	37	70
	MO+hIncRNA (n=512)	17	55	72

**Supplemental Table 9.** Summary of the cardiovascular-related lncRNAs found in this study, together with their characteristic profiles of expression and phenotypic consequences upon shRNA/morpholino-induced knockdown. N/D: not determined; EC: endothelial cell; HUVEC: human umbilical vein endothelial cells; D0-D2, D2-D4: refer to days after differentiation starts, please refer to first section of results or alternatively to Kurian et al, Nature Methods, 2013.

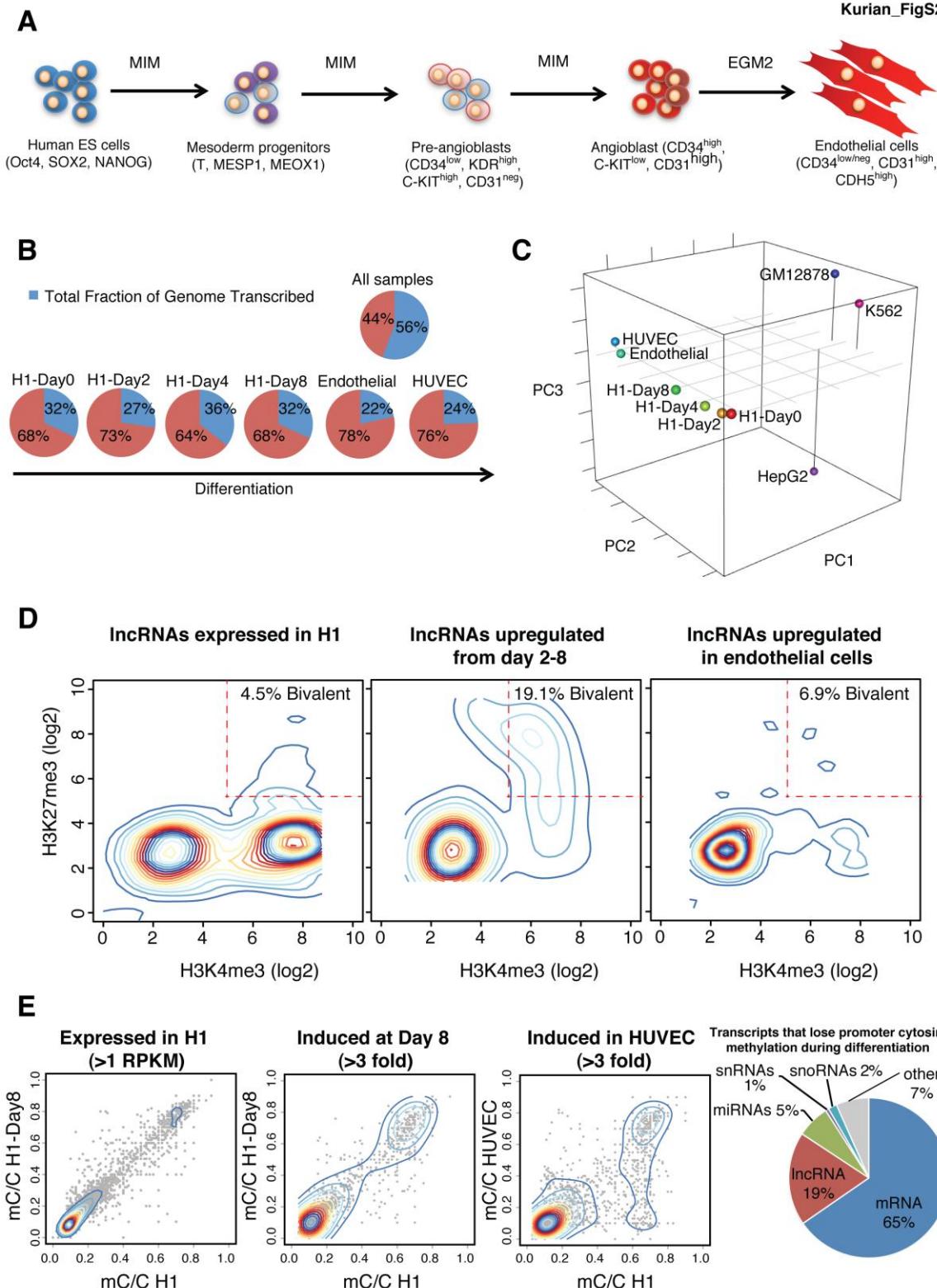
lncRNA		zebrafish			mouse			human		
Name	Human locus	Expression pattern	Cellular localization	Phenotype upon KD	Expression pattern	Cellular localization	Phenotype upon KD	Expression pattern	Cellular localization	Phenotype upon KD
<i>TERMINATOR (TMN)</i>	EU250746 (chr1:200,380,928-200,444,641)	Embryo up to 24 hpf	N/D	Early embryonic lethality (< 24h)	Blastocyst	N/D	N/D	Pluripotent stem cell restricted (D0-D2)	Nuclear	Loss of pluripotency (OCT4, SOX2), apoptosis
<i>ALIEN (ALN)</i>	LINC00261 (chr20:22,541,192-22,559,280)	Embryo 6-48 hpf	N/D	vessel branching defects, heart defects	E6.8-E8.5	N/D	N/D	Mesodermal restricted (D4-D8)	Cytoplasmic	Mesodermal arrest
<i>PUNISHER (PNS)</i>	LOC100130776 (AGAP2-AS1, chr12:58,120,023-58,122,139)	24-72 hpf	N/D	vessel defects, reduced EC number, heart defects	E8.5-E12.5	N/D	N/D	Endothelial cell restricted (EC/HUVEC)	Nuclear /cytoplasmic	Cell cycle arrest, defective blood vessel formation

□

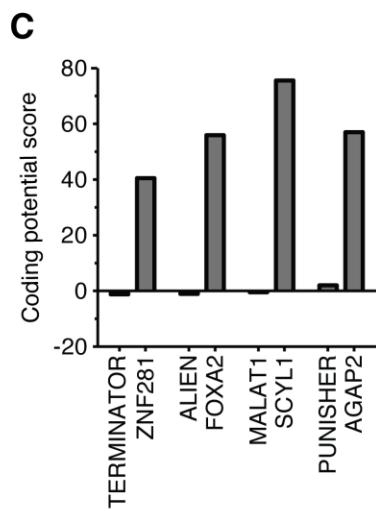
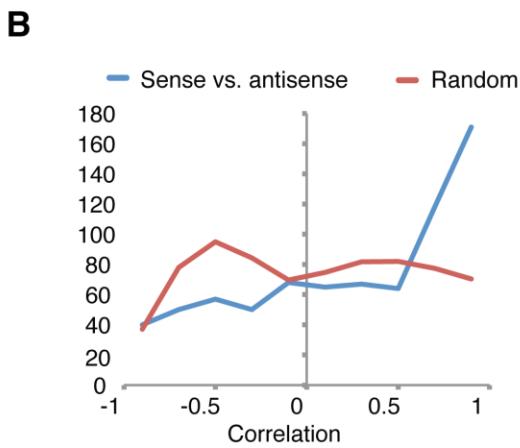
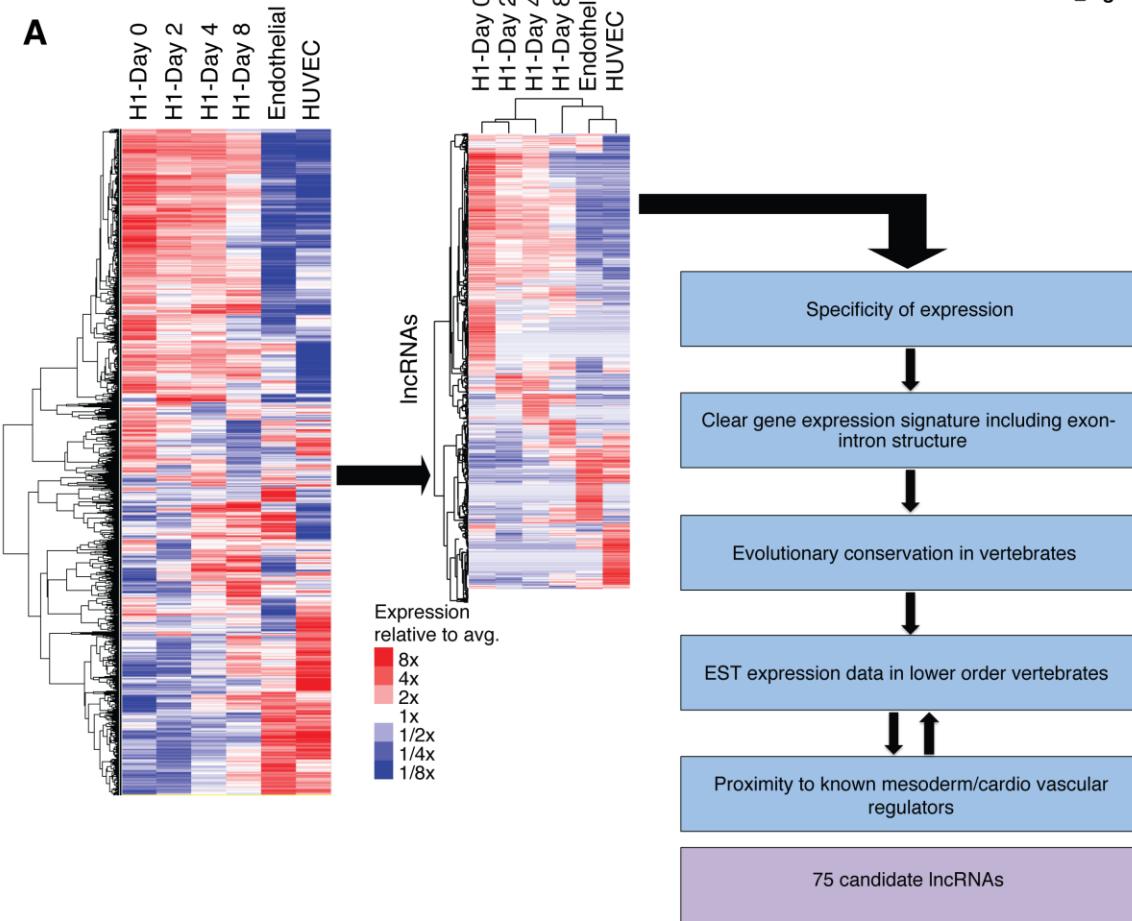
Kurian\_FigS1



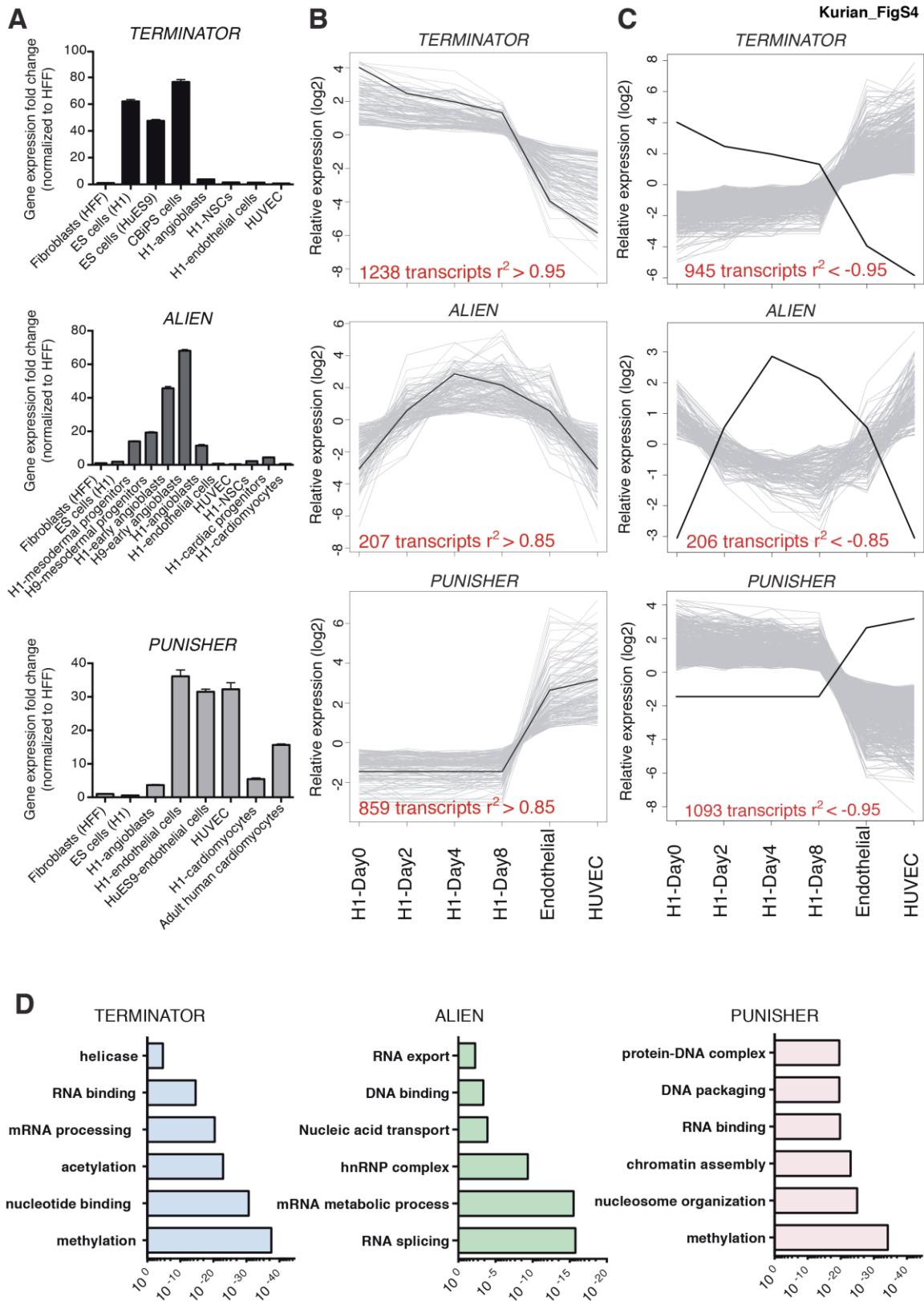
**Supplemental Figure 1. Analyses on RNA integrity and specificity of the sequencing reads.** **A**, To demonstrate that the RNA-seq experiment has minimal DNA contamination, we quantified reads on each strand in 10kb windows across the entire genome. Unlike DNA sequencing (right panel in red), our strand specific RNA sequencing data (left panel in blue) displays signals in a strand specific manner [>91% of regions with more than 16 normalized reads (per 10 million) compared to 0% for DNA sequencing]. **B**, The profile of reads across all genes additionally rules out DNA contamination as indicated by highly specific reads depending on the strand of origin as well as specific localization of signal within gene boundaries. Note that DNA sequencing data does not discriminate between different strands (right panel).



**Supplemental Figure 2. Transcriptional roadmap of human vascular differentiation.** **A**, Schematic representation showing the key stages of embryonic vascular differentiation presented in this study (hESCs, mesodermal progenitors, endothelial cells). **B**, Representation of the fraction of the human genome transcribed during the process of embryonic vascular differentiation. **C**, Principle component analysis of the transcriptome dynamics observed during differentiation. **D**, Bivalent histone methylation (H2K4me3 or H3K27me3) dynamics in lncRNAs during vascular differentiation. **E**, Genome-wide methylation profiling during differentiation from pluripotency to endothelial cells. Pie chart represents the summary of transcripts that have lost promoter cytosine methylation during vascular differentiation, including lncRNAs.

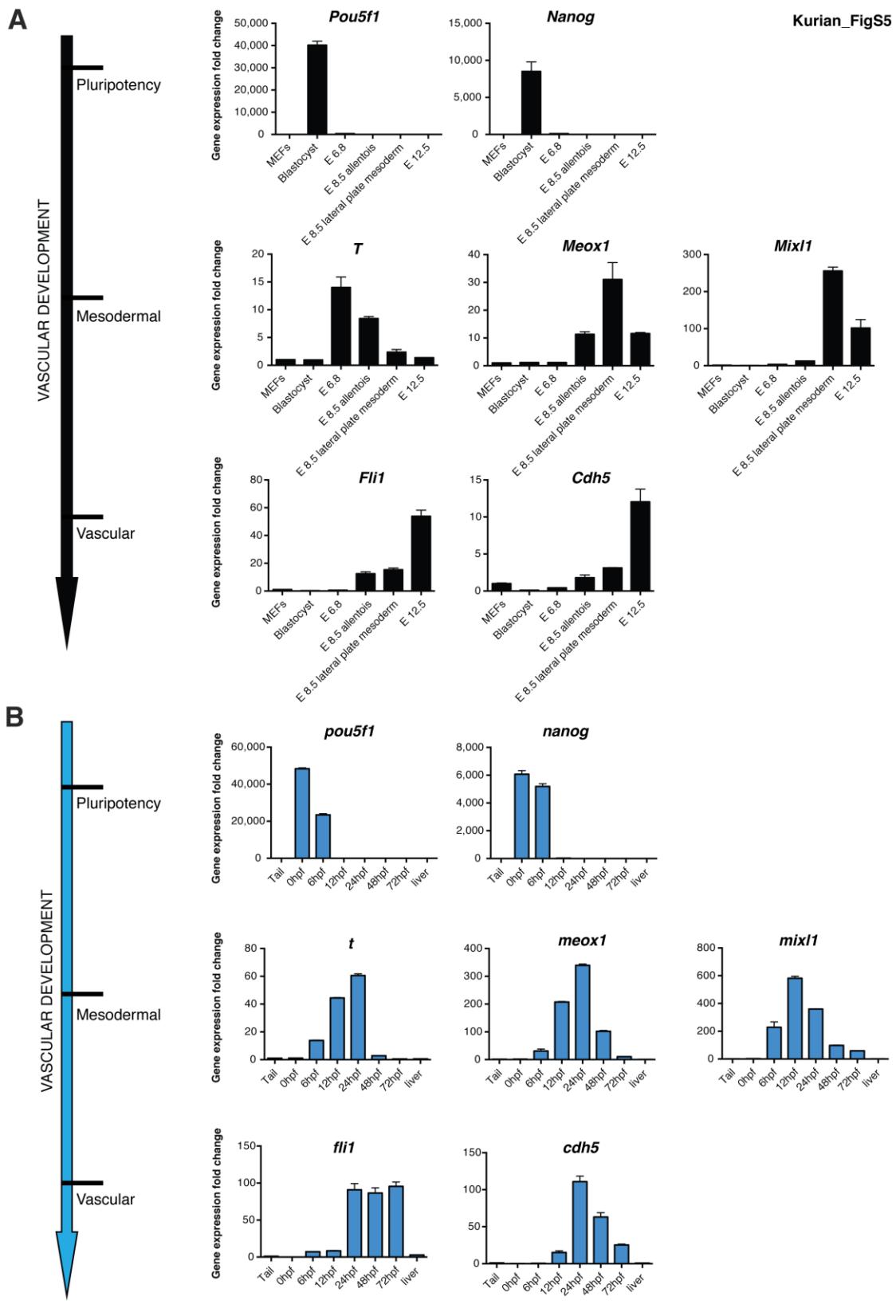


**Supplemental Figure 3. Identification of cell type/lineage specific lncRNAs during human vascular differentiation.** **A**, Hierarchical clustering of expressed genes from RNA-seq data at respective stages of vascular differentiation (left), and filtering strategy used to identify functional vascular lncRNAs (right). **B**, Gene expression correlation analysis of lncRNAs to their neighbouring protein coding counterparts. **C**, Codon substitution frequency analysis showing non-coding characteristics of identified lncRNAs compared to their neighbouring protein-coding genes.

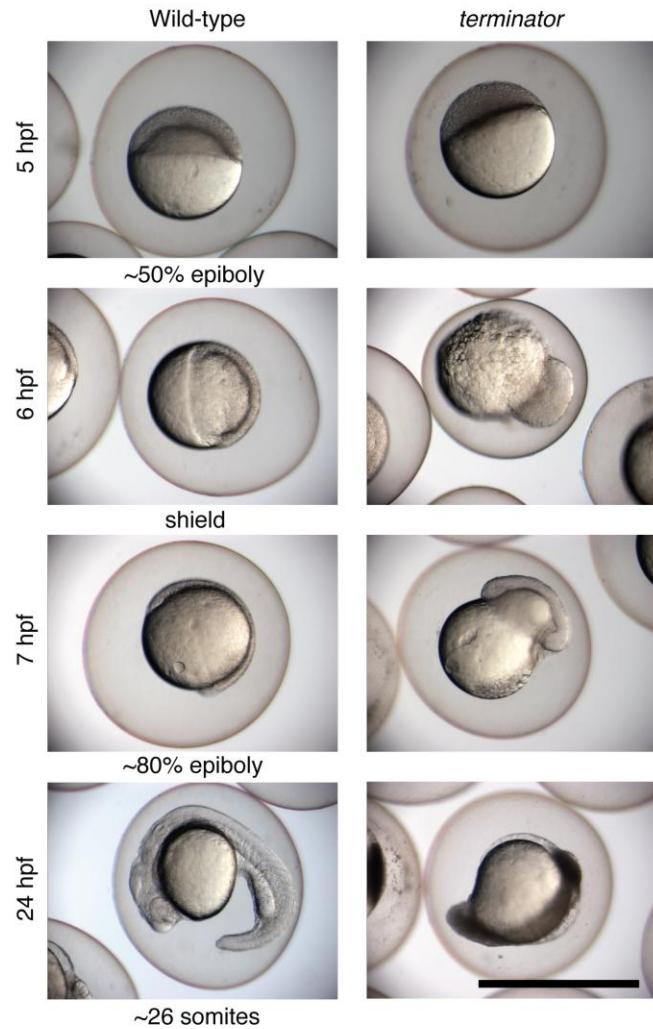


**Supplemental Figure 4. Dynamic expression of identified lncRNAs in respective cell types and identification of key molecular partner networks.**

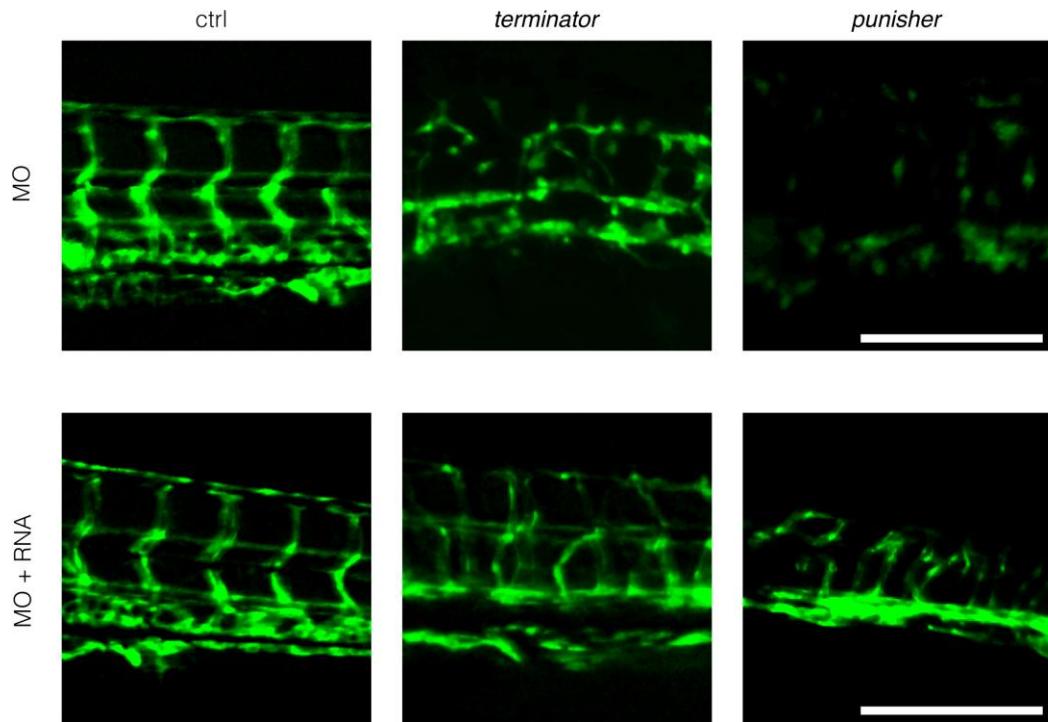
**A**, Expression of the different lncRNAs at indicated differentiation stages in relevant cell types as determined by qPCR ( $n \geq 5$ ). **B,C**, Dynamics of transcripts showing significant correlated (B) and anti-correlated (C) expression pattern to the different lncRNAs during vascular differentiation. **D**, Enriched Gene Ontology terms (gene function) of direct interacting partners for each lncRNAs, as identified by RNA immunoprecipitation coupled with mass spectrometry (RIP-MASS SPEC). Data are represented as mean  $\pm$  s.d., \*  $P < 0.05$ .



**Supplemental Figure 5. Dynamics of gene expression of key developmental markers during mouse and zebrafish vascular development.** **A**, Expression dynamics of pluripotency (*Pou5F1*, *Nanog*), mesoderm (*T*, *Meox1*, *Mixl1*) and endothelial markers (*Fli1*, *Cdh5*) during mouse embryonic development ( $n \geq 5$ ). **B**, Expression dynamics of pluripotency (*pou5f1*, *nanog*), mesoderm (*t*, *meox1*, *mixl1*) and endothelial markers (*fli1*, *cdh5*) during zebrafish embryonic development ( $n \geq 5$ ). Data are represented as mean  $\pm$  s.d.



**Supplemental Figure 6. *Terminator* loss of function results in lethality during early stages of vertebrate embryonic development.** Morpholino-mediated *Terminator* knockdown resulted in arrested development at 24 hpf in zebrafish embryos. At 6 hpf treated embryos showed detachment of the cell mass from the yolk. hpf: hours post-fertilization. Scale bar: 1mm



**Supplemental Figure 7. Phenotypic rescue for *terminator* and *punisher* in *fli1::GFP* zebrafish.** Phenotypic defects observed after antisense morpholinos (MO) for *terminator* or *punisher* were rescued by the co-injection of the human transcript of the lncRNA with its corresponding MO. Scale bars: 10 $\mu$ m.

# Gene Ontology Processes Dysregulated Upon Blockade of the Identified lncRNAs

## Gene Ontology processes upregulated upon Punisher knockdown

TermID	Term	Enrichment	logP	Genes in Term	Target Genes in Term	Fraction of Targets in Term	Total Target Genes	Total Genes	Gene Symbols
GO:0009615	response to virus	1.18E-06	-13.649417	247	24	0.047619048	504	15891	DDX58,MX1,PVR,GBP3,IFI44,IFIT3,OAS1,ELMOD2,ICAM1,IFITM1,ISG15,IFI44L,POLR3G,IFIT2,CXCL10,OAS2,PMAIP1,GBP1,IFIT5,SLC20A2,ACTA2,ITGAV,IFIT1,CDK6
GO:0051607	defense response to virus	4.66E-06	-12.277565	148	17	0.033730159	504	15891	DDX58,MX1,GBP3,IFIT3,OAS1,ELMOD2,IFITM1,ISG15,IFI44L,POLR3G,CXCL10,OAS2,IFIT2,PMAIP1,GBP1,IFIT5,IFIT1
GO:0001822	kidney development	4.89E-06	-12.227924	232	22	0.043650794	504	15891	BDNF,CREM1,ROBO2,LGR4,ARID5B,SULF1,TGFBR1,ADAMTS1,HOXA11,GCNT4,BMPER,C1GALT1,ZNF354A,CEP290,NOG,DLG1,TFAP2A,FSTL3,KIF3A,ACTA2,PKD2,CD34
GO:0071357	cellular response to type I interferon	5.57E-06	-12.098746	65	11	0.021825397	504	15891	MX1,IFIT2,OAS2,IRF6,IFI27,IFIT1,IFIT3,OA S1,XAF1,IFITM1,ISG15
GO:0060337	type I interferon signaling pathway	5.57E-06	-12.098746	65	11	0.021825397	504	15891	MX1,IFIT2,OAS2,IRF6,IFI27,IFIT1,IFIT3,OA S1,XAF1,IFITM1,ISG15
GO:0034340	response to type I interferon	6.49E-06	-11.945025	66	11	0.021825397	504	15891	MX1,IFIT2,OAS2,IRF6,IFI27,IFIT1,IFIT3,OA S1,XAF1,IFITM1,ISG15
GO:0072001	renal system development	1.10E-05	-11.418845	244	22	0.043650794	504	15891	BDNF,CREM1,ROBO2,LGR4,ARID5B,SULF1,TGFBR1,ADAMTS1,HOXA11,GCNT4,BMPER,C1GALT1,ZNF354A,CEP290,NOG,DLG1,FSTL3,TFAP2A,KIF3A,ACTA2,PKD2,CD34
GO:0070482	response to oxygen levels	1.94E-05	-10.850496	253	22	0.043650794	504	15891	BACH1,BDNF,SCFD1,CRYAB,CCL2,SDC2,TCEB1,TGFB2,BIRC2,ANGPTL4,FAM162A,ICAM1,EDNRA,LPAR1,ZNF354A,ANG,PMAIP1,PLOD2,STC2,UBE2B,MGARP,CPEB2
GO:0015992	proton transport	2.01E-05	-10.816109	117	14	0.027777778	504	15891	ATP6V1D,ATP6V1F,UQCR10,COX7A1,SLC35A3,ATP6V0A4,ATP5I,NDUFA4,ATP5E,COX7B,COX6C,ATP6V1E1,ATP6AP1L,ATP6V1G1
GO:0001666	response to hypoxia	2.12E-05	-10.762893	236	21	0.041666667	504	15891	BACH1,BDNF,SCFD1,CRYAB,CCL2,SDC2,TCEB1,TGFB2,BIRC2,ANGPTL4,FAM162A,ICAM1,EDNRA,ZNF354A,ANG,PMAIP1,PLOD2,STC2,UBE2B,MGARP,CPEB2
GO:0006818	hydrogen transport	2.44E-05	-10.622485	119	14	0.027777778	504	15891	ATP6V1D,ATP6V1F,UQCR10,COX7A1,SLC35A3,ATP6V0A4,ATP5I,NDUFA4,ATP5E,COX7B,COX6C,ATP6V1E1,ATP6AP1L,ATP6V1G1
GO:0036293	response to decreased oxygen levels	2.72E-05	-10.511486	240	21	0.041666667	504	15891	BACH1,BDNF,SCFD1,CRYAB,CCL2,SDC2,TCEB1,TGFB2,BIRC2,ANGPTL4,FAM162A,ICAM1,EDNRA,ZNF354A,ANG,PMAIP1,PLOD2,STC2,UBE2B,MGARP,CPEB2
GO:0030198	extracellular matrix organization	3.02E-05	-10.407763	358	27	0.053571429	504	15891	CREM1,FBLN5,MMP10,DMD,RECK,ITGA4,LEPREL1,A2M,COL8A1,SDC2,SULF1,TGFBR1,TGFB1,TGFB2,ICAM1,HAPLN1,DST,MMP7,POSTN,ABI3BP,PLOD2,VCAN,TFA P2A,ITGA11,ITGAV,LAMC2,LOX

GO:0043062	extracellular structure organization	3.17E-05	-10.359031	359	27	0.053571429	504	15891	GREM1,FBLN5,MMP10,DMD,RECK,ITGA4,LEPREL1,A2M,COL8A1,SDC2,SULF1,TGFBR1,TGFB1,TGFB2,ICAM1,HAPLN1,DST,MMP7,POSTN,ABI3BP,PLOD2,VCAN,TFA,P2A,ITGA11,ITGAV,LAMC2,LOX
GO:0015682	ferric iron transport	5.23E-05	-9.8584007	32	7	0.013888889	504	15891	ATP6V1D,ATP6V1F,ATP6V1E1,MCOLN1,TFR,ATP6V1G1,ATP6V0A4
GO:0033572	transferrin transport	5.23E-05	-9.8584007	32	7	0.013888889	504	15891	ATP6V1D,ATP6V1F,ATP6V1E1,MCOLN1,TFR,ATP6V1G1,ATP6V0A4
GO:0072512	trivalent inorganic cation transport	5.23E-05	-9.8584007	32	7	0.013888889	504	15891	ATP6V1D,ATP6V1F,ATP6V1E1,MCOLN1,TFR,ATP6V1G1,ATP6V0A4
GO:2000810	regulation of tight junction assembly	6.32E-05	-9.6686287	8	4	0.007936508	504	15891	PKP2,NEDD4L,GJA1,SNAI2
GO:0022904	respiratory electron transport chain	7.57E-05	-9.4889794	100	12	0.023809524	504	15891	BDNF,NDUFC1,NDUFB6,UQCR10,NDUFA1,ATP5I,NDUFA4,ATP5E,COX7B,COX6C,NDUFA13,NDUFS4

## Gene Ontology processes downregulated upon Punisher knockdown

TermID	Term	Enrichment	logP	Genes in Term	Target Genes in Term	Fraction of Targets in Term	Total Target Genes	Total Genes	Gene Symbols
GO:0000278	mitotic cell cycle	1.21E-53	-121.84441	730	133	0.214516129	620	15891	POLA2,SKA3,NDC80,ENSA,KIF2C,TXNL4B,CCNB1,MCM4,KIFC1, ,WEE1,ANAPC1,PSMB8,XRCC2,TUBG1,KIF22,CDC25B,PLK1,CE NPA,KIF15,RRM2,CDC6,HDAC8,AURKB,CCNF,ANLN,RFC2,CDK 1,GINS4,PRIM1,KIF18B,TYMS,VRK1,E2F8,ANAPC5,MCM6,CLS PN,KHDRBS1,CENPF,RRM1,KIF23,HAUS4,RACGAP1,SPAG5,L NA,NUP88,MCM5,TTK,DNA2,NEK2,TICRR,TIMELESS,KNTC1,T UBGCP3,CDC25A,ORC2,KIF20A,POLA1,MYBL2,HIST2H4B,KIF2 0B,PLCB1,FAM64A,CDC45,FAM83D,CDK2,CCNB2,DLGAP5,SPC 25,TUBB4B,FOXM1,PTTG1,TPX2,ESCO2,NUF2,GSG2,ORC1,NC APD2,CCNA2,NCAPL,CDH13,PLK4,RFC4,KNSTRN,RFC3,CDC20, CDC2A,BUB1B,NUSAP1,NUP155,CDC25C,CDC45,CEP55,MAD2 L1,CCNA1,SGOL1,MCM3,MCM10,DDX11,CIT,CENPI,CDCA3,CD CA8,NCAPG,MCM7,PPP1CC,CENPL,TOP2A,USP3,PRC1,CKS1B, BUB1,CENPM,IQGAP3,SMC4,GINS1,ZWINT,RHOA,ASPM,SKP2, KIF11,SKA1,NCAPD3,POLE,RFC5,POLE2,TFDP1,SPC24,CASC5,U BE2C,NCAPG2,ESPL1,MELK,CDKN3
GO:0007049	cell cycle	5.95E-47	-106.43753	1226	163	0.262903226	620	15891	POLA2,SKA3,NDC80,FANCDC2,MDM4,ENSA,KIF2C,TXNL4B,CC B1,BARD1,FANCI,MCM4,KIFC1,WEE1,ANAPC1,PA2G4,PSMB8, SPIN1,XRCC2,TUBG1,KIF22,CDC25B,TACC3,TRIP13,PLK1,CENP A,KIF15,MKI67,RRM2,CDC6,HDAC8,AURKB,CCNF,ANLN,RFC2, CDK1,GINS4,PRIM1,KIF18B,TYMS,VRK1,E2F8,ANAPC5,MCM6, HJURP,CLSPN,KHDRBS1,CENPF,RRM1,KIF23,HAUS4,RACGAP1, PHGDH,SPAG5,LMNA,DTL,NUP88,MCM5,APEX2,TTK,DNA2,NE K2,TICRR,TIMELESS,KNTC1,DDX12P,TUBGCP3,CDC25A,ORC2, KIF20A,POLA1,MYBL2,HIST2H4B,KIF20B,FANCA,PLCB1,FAM64 A,RAD51,CDC45,FAM83D,AUNIP,CDK2,CCNB2,KRT18,DLGAP5 ,SPC25,TUBB4B,FOXM1,BRIP1,PTTG1,TPX2,ESCO2,NUF2,GSG 2,ORC1,NCAPD2,CCNA2,DCLRE1B,RALB,NCAPL,CDH13,PLK4, MND1,RFC4,KNSTRN,RFC3,CDC20,NUP155,CDC2A,BUB1B,NU SAP1,CDC25C,CDC45,EXO1,CCNA1,CEP55,MAD2L1,H2AFX,SG OL1,MCM3,MCM10,DDX11,CIT,CENPI,CDCA3,CDCA8,RAD54L, BLM,NCAPG,MCM7,PPP1CC,CENPL,EZH2,TOP2A,USP3,GAS2L 3,PRC1,CKS1B,BUB1,CENPM,IQGAP3,SMC4,GINS1,ZWINT,RH OA,ASPM,SKP2,RASSF2,KIF11,SKA1,NCAPD3,RFC5,POLE,POLE 2,MAP2K6,TFDP1,SPC24,CASC5,NCAPG2,UBE2C,ESPL1,MELK, CDKN3

GO:0022402	cell cycle process	1.13E-43	-98.886579	950	138	0.222580645	620	15891	POLA2,SKA3,NDC80,FANCD2,MDM4,ENSA,KIF2C,TXNL4B,CCNB1,BARD1,MCM4,KIFC1,WEE1,ANAPC1,PA2G4,PSMB8,TUBG1,KIF22,CDC25B,TACC3,TRIP13,PLK1,CENPA,KIF15,RRM2,CDC6,HDAC8,AURKB,CCNF,ANLN,RFC2,CDK1,PRIM1,KIF18B,TYMS,VRK1,ANAPC5,MCM6,CLSPN,KHDRBS1,CENPF,KIF23,HAUS4,RA,CGAP1,PHGDH,SPAG5,LMNA,DTL,NUP88,MCM5,TTK,DNA2,N,EK2,TICRR,TIMELESS,KNTC1,TUBGCP3,CDC25A,ORC2,KIF20A,POLA1,MYBL2,KIF20B,PLCB1,FAM64A,RAD51,CDCA5,FAM83D,AUNIP,CDK2,CCNB2,DLGAP5,SPC25,TUBB4B,FOXM1,BRIP1,PTTG1,TPX2,NUF2,GSG2,ORC1,NCAPD2,CCNA2,DCLRE1B,RALB,NCAPH,PLK4,RFC4,KNSTRN,RFC3,CDC20,CDCA2,BUB1B,NUSA P1,NUP155,CDC25C,CDC45,CEP55,MAD2L1,CCNA1,H2AFX,SGOL1,MCM3,MCM10,DDX11,CIT,CDCA3,CDCA8,BLM,NCAPG,M CM7,EZH2,TOP2A,GAS2L3,PRC1,CKS1B,BUB1,IQGAP3,SMC4,Z WINT,RHOA,ASPM,SKP2,KIF11,SKA1,NCAPD3,POLE,RFC5,POL E2,MAP2K6,TFDP1,SPC24,CASC5,NCAPG2,UBE2C,ESPL1,MELK,CDKN3
GO:1903047	mitotic cell cycle process	2.18E-42	-95.928484	661	113	0.182258065	620	15891	POLA2,SKA3,SPC25,DLGAP5,FOXM1,TUBB4B,PTTG1,NDC80,TPX2,ENSA,GSG2,NUF2,KIF2C,ORC1,TXNL4B,CCNB1,MCM4,NCAPD2,KIFC1,WEE1,ANAPC1,CCNA2,PSMB8,NCAPH,PLK4,TUBG1,KIF22,CDC25B,PLK1,KNSTRN,CENPA,CDC20,KIF15,RRM2,NU P155,CDC6,NUSAP1,BUB1B,CDCA2,AURKB,CCNF,ANLN,CDC25 C,CDK1,PRIM1,CDC45,KIF18B,CCNA1,MAD2L1,CEP55,TYMS,S GOL1,VRK1,MCM3,ANAPC5,MCM6,MCM10,CLSPN,DDX11,CIT ,KHDRBS1,CENPF,CDCA8,CDCA3,KIF23,RACGAP1,HAUS4,MC M7,NCAPG,SPAG5,LMNA,TOP2A,NUP88,MCM5,TTK,PRC1,NE K2,TICRR,CKS1B,TIMELESS,BUB1,KNTC1,IQGAP3,TUBGCP3,SM C4,CDC25A,ORC2,RHOA,ZWINT,POLA1,ASPM,MYBL2,SKP2,KIF20B,SKA1,PLCB1,KIF11,NCAPD3,POLE2,FAM64A,POLE,CDCA5,TFDP1,FAM83D,SPC24,CASC5,NCAPG2,CDK2,UBE2C,CCNB2,E SPL1,MELK,CDKN3
GO:0007067	mitotic nuclear division	2.99E-38	-86.404529	305	74	0.119354839	620	15891	SKA3,SPC25,DLGAP5,PTTG1,NDC80,TPX2,ENSA,GSG2,NUF2,KIF2C,TXNL4B,CCNB1,NCAPD2,KIFC1,WEE1,ANAPC1,CCNA2,NCAPH,KIF22,CDC25B,PLK1,KNSTRN,CDC20,KIF15,CDC6,CDCA2, BUB1B,NUSAP1,AURKB,CCNF,ANLN,CDC25C,CDK1,KIF18B,CC NA1,MAD2L1,CEP55,SGOL1,VRK1,ANAPC5,DDX11,CIT,CENPF,CDCA8,CDCA3,KIF23,RACGAP1,HAUS4,NCAPG,SPAG5,NEK2,TIMELESS,BUB1,KNTC1,SMC4,CDC25A,ZWINT,RHOA,ASPM,MY BL2,KIF20B,KIF11,SKA1,NCAPD3,FAM64A,CDCA5,FAM83D,SP C24,CASC5,NCAPG2,CDK2,CCNB2,UBE2C,ESPL1
GO:0000280	nuclear division	1.26E-37	-84.965968	428	86	0.138709677	620	15891	SKA3,SPC25,DLGAP5,PTTG1,NDC80,FANCD2,TPX2,ENSA,GSG2 ,NUF2,KIF2C,TXNL4B,CCNB1,NCAPD2,KIFC1,WEE1,ANAPC1,CC NA2,SPIN1,NCAPH,RCSC2,KIF22,CDC25B,TRIP13,PLK1,KNSTR N,MND1,CDC20,KIF15,MKI67,CDC6,CDCA2,BUB1B,NUSAP1,A URKB,CCNF,ANLN,CDC25C,CDK1,EXO1,KIF18B,CCNA1,MAD2L 1,CEP55,H2AFX,SGOL1,VRK1,ANAPC5,DDX11,CIT,CENPF,CDCA 8,CDCA3,KIF23,RACGAP1,HAUS4,RAD54L,NCAPG,SPAG5,TOP 2A,NEK2,TIMELESS,BUB1,KNTC1,SMC4,CDC25A,ZWINT,RHOA,ASPM,MYBL2,KIF20B,FANCA,KIF11,SKA1,NCAPD3,FAM64A,RAD51,CDCA5,FAM83D,SPC24,CASC5,NCAPG2,CDK2,CCNB2,UBE2C,ESPL1

GO:0048285	organelle fission	1.81E-36	-82.300442	453	87	0.140322581	620	15891	SKA3,SPC25,DLGAP5,PTTG1,FANCD2,NDC80,TPX2,ENSA,NUF2,GSG2,KIF2C,CCNB1,TXNL4B,NCAPD2,KIFC1,WEE1,ANAPC1,CCNA2,SPIN1,NCAPH,XRCC2,KIF22,CDC25B,TRIP13,PLK1,MND1,KNSTRN,CDC20,KIF15,MKI67,CDC6,CDCA2,BUB1B,NUSAP1,AURKB,CCNF,ANLN,CDC25C,CDK1,EXO1,KIF18B,MAD2L1,CEP55,CCNA1,H2AFX,SGOL1,VRK1,ANAPC5,DDX11,CIT,CENPF,CDCA3,CDCA8,KIF23,RACGAP1,HAUS4,RAD54L,NCAPG,SPAG5,TOP2A,NEK2,TIMELESS,BUB1,KNTC1,SMC4,CDC25A,RHOA,ZWINT,ASPM,MTFR2,MYBL2,KIF20B,FANCA,SKA1,KIF11,NCAPD3,FA M64A,RAD51,CDCA5,FAM83D,SPC24,CASC5,NCAPG2,UBE2C,CCNB2,CDK2,ESPL1
GO:0006259	DNA metabolic process	1.91E-30	-68.431707	810	108	0.174193548	620	15891	POLA2,BRIP1,FOXM1,PTTG1,FANCD2,ORC1,BARD1,RPAIN,FA NCI,MTRR,NCAPD2,MCM4,RBM14,DCLRE1B,NCAPH,PARPBP,XRCC2,DNMT3B,KIF22,HIST1H2AB,DNMT1,TRIP13,MND1,RFC4,NONO,RFC3,CENPA,PARP1,PPT1,MKI67,HIST1H2BK,RRM2,C DC6,NUSAP1,RFC2,CDC25C,FANCE,CDK1,GINS4,ASF1B,PRIM1,EXO1,CDC45,INO80D,KIAA0101,H2AFX,HIST1H1B,TYMS,IGFB P4,SMG1,MCM3,POLO,SSRP1,MCM6,MCM10,CLSPN,DDX11, HJURP,CENPI,CENPF,RRM1,HIST1H1E,RAD54L,BLM,MCM7,NC APG,TOP2A,DTL,ALYREF,USP3,MCM5,APEX2,DNA2,TICRR,HIS T2H3D,HIST1H2BH,CDK2AP1,DDX12P,POLR2B,SMC4,HIST1H2 BB,GINS1,CDC25A,TK1,ORC2,APEX1,POLA1,ERCC6L2,HIST2H4 B,FANCA,NFIB,XRCC1,HIST1H2AK,NCAPD3,RFC5,POLE2,POLE, HIST1H2BM,RAD51,LOC100133315,NEIL3,CDCA5,UNG,ARRB1,CASC5,NCAPG2,CDK2,NHEJ1
GO:0051301	cell division	3.57E-30	-67.804606	594	91	0.146774194	620	15891	SKA3,SPC25,DLGAP5,PTTG1,NDC80,FANCD2,TPX2,ENSA,GSG2 ,NUF2,KIF2C,TXNL4B,CCNB1,NCAPD2,KIFC1,WEE1,ANAPC1,CCNA2,SPIN1,RALB,NCAPH,XRCC2,KIF22,CDC25B,TRIP13,PLK1,K NSTRN,MND1,CDC20,KIF15,MKI67,CDC6,CDCA2,BUB1B,NUSA P1,AURKB,CCNF,ANLN,CDC25C,CDK1,EXO1,KIF18B,CCNA1,M AD2L1,CEP55,H2AFX,SGOL1,VRK1,ANAPC5,DDX11,CIT,CENPF, CDCA8,CDCA3,KIF23,RACGAP1,HAUS4,RAD54L,NCAPG,PPP1C C,SPAG5,TOP2A,PRC1,NEK2,CKS1B,TIMELESS,BUB1,KNTC1,SMC4,CDC25A,ZWINT,RHOA,KIF20A,ASPM,MYBL2,KIF20B,FAN CA,KIF11,SKA1,NCAPD3,FA M64A,RAD51,CDCA5,FAM83D,SPC24,CASC5,NCAPG2,CDK2,CCNB2,UBE2C,ESPL1

GO:1902589	single-organism organelle organization	3.64E-23	-51.668335	1441	136	0.219354839	620	15891	ARHGAP26,POLA2,SKA3,NF1,VAMP8,NDC80,AQP1,FANCD2,E NSA,KIF2C,TXNL4B,CCNB1,KIFC1,WEE1,ANAPC1,RBM14,XRCC 2,DNMT3B,TUBG1,KIF22,CDC25B,TACC3,TRIP13,PLK1,CENPA, SYNE3,KIF15,CDC6,HDAC8,AURKB,CCNF,ANLN,RFC2,CDK1,PE X1,PRIM1,COPZ1,KIAA0101,KIF18B,VRK1,ANAPC5,CENPF,KIF2 3,HAUS4,RACGAP1,SPAG5,LMNA,CFLAR,TTK,DNA2,NEK2,TO MM22,TIMELESS,KNTC1,TUBGCP3,CDC25A,KIF20A,POLA1,MY BL2,HIST2H4B,KIF20B,BAP1,PHACTR4,FAM64A,RAD51,CDCA5 ,FAM83D,DYSF,AUNIP,CDK2,KRT18,CCNB2,DLGAP5,CEP68,SP C25,PTTG1,TPX2,TESK2,NUF2,GSG2,NCAPD2,CCNA2,DCLRE1B ,NCAPH,PLK4,TADA1,RFC4,KNSTRN,RFC3,PARP1,SYNE2,CDC2 0,PREX1,CDCA2,BUB1B,NUSAP1,GIMAP5,CDC25C,CEP55,MA D2L1,CCNA1,SGOL1,DDX11,CIT,CDCA3,CDCA8,BLM,NCAPG,EZ H2,MSRB1,TOP2A,GATC,USP3,GAS2L3,PRC1,BUB1,SMC4,ZWI NT,RHOA,MSL2,ASPM,KIF11,SKA1,NCAPD3,POLE,RFC5,POLE2, ELMO1,ARRB1,FGD5,SPC24,CASC5,NCAPG2,UBE2C,ESPL1,CO RO2B
GO:0006260	DNA replication	1.14E-20	-45.919457	216	45	0.072580645	620	15891	POLA2,MCM6,MCM10,CLSPN,CENPF,RRM1,ORC1,BLM,MCM 7,RPAIN,TOP2A,MCM4,DTL,RBM14,ALYREF,MCM5,DNA2,TIC RR,CDK2AP1,GINS1,RFC4,CDC25A,RFC3,TK1,ORC2,POLA1,RR M2,CDC6,RFC2,CDC25C,NFIB,CDK1,GINS4,PRIM1,CDC45,KIAA 0101,POLE2,POLE,RFC5,RAD51,TYMS,CDK2,MCM3,SSRP1,POL Q
GO:0006996	organelle organization	4.87E-20	-44.469222	2320	177	0.285483871	620	15891	ARHGAP26,POLA2,SKA3,NF1,VAMP8,FANCD2,NDC80,AQP1,E NSA,KIF2C,CCNB1,TXNL4B,KIFC1,WEE1,ANAPC1,RBM14,SPIN 1,TIMM23,XRCC2,DNMT3B,TUBG1,KIF22,CDC25B,HIST1H2AB, DNMT1,TRIP13,TACC3,PLK1,CENPA,SYNE3,KIF15,MKI67,HIST1 H2BK,CDC6,HDAC8,AURKB,CCNF,ANLN,RFC2,CDK1,ASF1B,PRI M1,PEX1,COPZ1,KIAA0101,KIF18B,MPV17,HIST1H1B,CLU,VRK 1,ANAPC5,HJURP,CENPF,KIF23,HAUS4,RACGAP1,HIST1H1E,L MNA,SPAG5,CFLAR,NUP88,VPS72,TTK,DNA2,NEK2,TIMELESS, TOMM22,KNTC1,TUBGCP3,TOMM34,NOS3,USE1,CDC25A,KIF 20A,POLA1,MTRF2,HIST2H4B,MYBL2,HIP1,KIF20B,BAP1,FANC A,PHACTR4,KIF4A,FAM64A,HIST1H2BM,RAD51,CDCA5,FAM8 3D,DYSF,AUNIP,CDK2,CCNB2,KRT18,SPC25,DLGAP5,CEP68,PT TG1,TPX2,TESK2,GSG2,NUF2,SEMA6A,NCAPD2,RRN3,CCNA2, DCLRE1B,SRPX,NCAPH,PLK4,TADA1,RFC4,KNSTRN,MND1,RFC 3,PARP1,PPT1,CDC20,SYNE2,PREX1,NUSAP1,BUB1B,NUP155, CDCA2,RAB38,GIMAP5,ERAL1,CDC25C,EXO1,MAD2L1,CEP55, CCNA1,H2AFX,SGOL1,DDX11,CENPF,PHF19,CIT,CDCA3,CDCA8, RAD54L,BLM,NCAPG,MSRB1,EZH2,TOP2A,GATC,USP3,GAS2L3 ,PRC1,HIST2H3D,BUB1,HIST1H2BH,SMC4,HIST1H2BB,RHOA,Z WINT,MSL2,ASPM,NLGN1,PDE2A,SKA1,KIF11,HIST1H2AK,NCA PD3,POLE2,POLE,RFC5,ELMO1,FGD5,ARRB1,SPC24,CASC5,NC APG2,UBE2C,ESPL1,CORO2B

TermID	Term	Enrichment	logP	Genes in Term	Target Genes in Term	Fraction of Targets in Term	Total Target Genes	Total Genes	Gene Symbols
GO:0000278	mitotic cell cycle	1.21E-53	-121.84441	730	133	0.214516129	620	15891	POLA2,SKA3,NDC80,ENSA,KIF2C,TXNL4B,CCNB1,MCM4,KIFC1, ,WEE1,ANAPC1,PSMB8,XRCC2,TUBG1,KIF22,CDC25B,PLK1,CE NPA,KIF15,RRM2,CDC6,HDAC8,AURKB,CCNF,ANLN,RFC2,CDK 1,GINS4,PRIM1,KIF18B,TYMS,VRK1,E2F8,ANAPC5,MCM6,CLS PN,KHDRBS1,CENPF,RRM1,KIF23,HAUS4,RACGAP1,SPAG5,L NA,NUP88,MCM5,TTK,DNA2,NEK2,TICRR,TIMELESS,KNTC1,T UBGCP3,CDC25A,ORC2,KIF20A,POLA1,MYBL2,HIST2H4B,KIF2 0B,PLCB1,FAM64A,CDCA5,FAM83D,CDK2,CCNB2,DLGAP5,SPC 25,TUBB4B,FOXM1,PTTG1,TPX2,ESCO2,NUF2,GSG2,ORC1,NC APD2,CCNA2,NCAPH,CDH13,PLK4,RFC4,KNSTRN,RFC3,CDC20, CDC2,BUB1B,NUSAP1,NUP155,CDC25C,CDC45,CEP55,MAD2 L1,CCNA1,SGOL1,MCM3,MCM10,DDX11,CIT,CENPI,CDCA3,CD CA8,NCAPG,MCM7,PPP1CC,CENPL,TOP2A,USP3,PRC1,CKS1B, BUB1,CENPM,IQGAP3,SMC4,GINS1,ZWINT,RHOA,ASPM,SKP2, KIF11,SKA1,NCAPD3,POLE,RFC5,POLE2,TFDP1,SPC24,CASC5,U BE2C,NCAPG2,ESPL1,MELK,CDKN3
GO:0007049	cell cycle	5.95E-47	-106.43753	1226	163	0.262903226	620	15891	POLA2,SKA3,NDC80,FANCD2,MDM4,ENSA,KIF2C,TXNL4B,CCN B1,BARD1,FANCI,MCM4,KIFC1,WEE1,ANAPC1,PA2G4,PSMB8, SPIN1,XRCC2,TUBG1,KIF22,CDC25B,TACC3,TRIP13,PLK1,CEN A,KIF15,MKI67,RRM2,CDC6,HDAC8,AURKB,CCNF,ANLN,RFC2, CDK1,GINS4,PRIM1,KIF18B,TYMS,VRK1,E2F8,ANAPC5,MCM6, HJURP,CLSPN,KHDRBS1,CENPF,RRM1,KIF23,HAUS4,RACGAP1, PHGDH,SPAG5,LMNA,DTL,NUP88,MCM5,APEX2,TTK,DNA2,NE K2,TICRR,TIMELESS,KNTC1,DDX12P,TUBGCP3,CDC25A,ORC2, KIF20A,POLA1,MYBL2,HIST2H4B,KIF20B,FANCA,PLCB1,FAM64 A,RAD51,CDCA5,FAM83D,AUNIP,CDK2,CCNB2,KRT18,DLGAP5 ,SPC25,TUBB4B,FOXM1,BRIP1,PTTG1,TPX2,ESCO2,NUF2,GSG 2,ORC1,NCAPD2,CCNA2,DCLRE1B,RALB,NCAPH,CDH13,PLK4, MND1,RFC4,KNSTRN,RFC3,CDC20,NUP155,CDCA2,BUB1B,NU SAP1,CDC25C,CDC45,EXO1,CCNA1,CEP55,MAD2L1,H2AFX,SG OL1,MCM3,MCM10,DDX11,CIT,CENPI,CDCA3,CDCA8,RAD54L, BLM,NCAPG,MCM7,PPP1CC,CENPL,EZH2,TOP2A,USP3,GAS2L 3,PRC1,CKS1B,BUB1,CENPM,IQGAP3,SMC4,GINS1,ZWINT,RH OA,ASPM,SKP2,RASSF2,KIF11,SKA1,NCAPD3,RFC5,POLE,POLE 2,MAP2K6,TFDP1,SPC24,CASC5,NCAPG2,UBE2C,ESPL1,MELK, CDKN3

## Gene Ontology processes upregulated upon Alien knockdown

TermID	Term	Enrichment	logP	Genes in Term	Target Genes in Term	Fraction of Targets in Term	Total Target Genes	Total Genes	Gene Symbols
GO:0007155	cell adhesion	1.38E-08	-18.098974	822	66	0.107142857	616	15891	MPZL3,MGP,DDR2,PKP2,IGFBP7,CDH1,NPNT,CDH10,CDH3,TLN2,ITGB6,SLIT2,CYP1B1,DSC3,AATF,MPZL2,ARHGAP5,CADM1,CDON,CLDN6,DSP,SEMA5A,COL12A1,CXADR,NRXN3,FERMT1,LMLN,MFAP4,CLCA2,DPP4,KAL1,ADAM9,SPON1,LAMC2,EMR2,C,LDN1,PCDH18,OLR1,BVES,PTPRD,NEDD9,ITGA6,S,POCK1,TGFB2,LAMA1,COL6A3,THBS1,SPP1,ITGB8,DSG2,FREM2,TENM3,FBLN5,CNTN5,TNFAIP6,RELN,EPHA1,GPNMB,POSTN,VCAN,EFNA1,ITGB4,SVEP1,NLG4Y,EGFL6,PCDHB4
GO:0022610	biological adhesion	1.59E-08	-17.957071	825	66	0.107142857	616	15891	MPZL3,MGP,DDR2,PKP2,IGFBP7,CDH1,NPNT,CDH10,CDH3,TLN2,ITGB6,SLIT2,CYP1B1,DSC3,AATF,MPZL2,ARHGAP5,CADM1,CDON,CLDN6,DSP,SEMA5A,COL12A1,CXADR,NRXN3,FERMT1,LMLN,MFAP4,CLCA2,DPP4,KAL1,ADAM9,SPON1,LAMC2,EMR2,C,LDN1,PCDH18,OLR1,BVES,PTPRD,NEDD9,ITGA6,S,POCK1,TGFB2,LAMA1,COL6A3,THBS1,SPP1,ITGB8,DSG2,FREM2,TENM3,FBLN5,CNTN5,TNFAIP6,RELN,EPHA1,GPNMB,POSTN,VCAN,EFNA1,ITGB4,SVEP1,NLG4Y,EGFL6,PCDHB4
GO:0030198	extracellular matrix organization	1.85E-08	-17.808164	358	38	0.061688312	616	15891	DCN,MPZL3,MMP9,ITGA6,LEPREL1,SDC4,MFAP5,DDR2,COL4A5,FSHR,TGFB2,LAMA1,HAS2,COL6A3,THBS1,FMOD,SPP1,CDH1,ITGB8,NPNT,APBB2,MMP2,TFAP2A,ITGB6,CYP1B1,GREM1,FBLN5,COL12A1,POSTN,VCAN,MFAP4,ADAM9,SERpine1,BMP4,C,OL1A2,ITGB4,LAMC2,EGFL6
GO:0043062	extracellular structure organization	1.99E-08	-17.732774	359	38	0.061688312	616	15891	DCN,MPZL3,MMP9,ITGA6,LEPREL1,SDC4,MFAP5,DDR2,COL4A5,FSHR,TGFB2,LAMA1,HAS2,COL6A3,THBS1,FMOD,SPP1,CDH1,ITGB8,NPNT,APBB2,MMP2,TFAP2A,ITGB6,CYP1B1,GREM1,FBLN5,COL12A1,POSTN,VCAN,MFAP4,ADAM9,SERpine1,BMP4,C,OL1A2,ITGB4,LAMC2,EGFL6
GO:0030155	regulation of cell adhesion	2.04E-07	-15.405821	314	33	0.053571429	616	15891	ITGA6,SDC4,SPOCK1,ASS1,TGFB2,LAMA1,HAS2,WNT5A,THBS1,EPCAM,SPP1,CDH1,NPNT,ADAMTS18,PRKG1,SFRP1,TACSTD2,TGM2,ERBB3,CDK6,LRRC16A,CYP1B1,TDGF1,GREM1,PIK3R1,EPHA1,SEMA5A,PPP2CA,ADAM9,DPP4,SERpine1,EFNA1,EGFL6
GO:0072001	renal system development	9.62E-07	-13.854386	244	27	0.043831169	616	15891	DCN,RARB,BDNF,GREM1,ENPEP,EYA1,SDC4,PLCE1,SMAD5,HAS2,SIX4,RPGRIPL1,WNT5A,SGPL1,EPCAM,TMED10,RRM2B,FGFR2,NPNT,GPC3,TP63,TFAP2A,ACTA2,SFRP1,TACSTD2,BMP4,SLT2

GO:0051270	regulation of cellular component movement	1.04E-06	-13.777932	573	47	0.076298701	616	15891	MMP9,ANXA3,DDR2,PKP2,NEDD4L,TGFB2,LAMA1,HAS2,PARD6B,WNT5A,THBS1,P2RY6,IGF1R,DSG2,TFAP2A,SFRP1,TACSTD2,PDE4D,MET,AMOT,ENPP2,SLIT2,LRRK16A,CDK6,NOX4,CYP1B1,TDGF1,GRM1,APEX1,CGA,PIK3R1,ARHGAP5,CMKL1,NUP155,RELN,EPHA1,DSP,PLCB1,NEXN,SEMA5A,BMP5,IL6,ADAM9,SERpine1,EFNA1,BMP4,HACE1
GO:0030334	regulation of cell migration	2.09E-06	-13.079813	482	41	0.066558442	616	15891	MMP9,ANXA3,DDR2,PKP2,TGFB2,LAMA1,HAS2,PARD6B,WNT5A,THBS1,P2RY6,IGF1R,TFAP2A,SFRP1,TACSTD2,MET,AMOT,ENPP2,SLIT2,LRRK16A,NOX4,CYP1B1,TDGF1,GRM1,APEX1,CGA,PIK3R1,ARHGAP5,CMKL1,RELN,EPHA1,PLCB1,NEXN,SEMA5A,BMP5,IL6,ADAM9,SERpine1,EFNA1,BMP4,HACE1
GO:0022612	gland morphogenesis	2.39E-06	-12.945948	104	16	0.025974026	616	15891	PTHLH,CAV1,TGFB2,LAMA1,WNT5A,NCOA3,CDH1,IGF1R,FGFR2,TP63,IL6,SEMA3C,SFRP1,TGM2,BMP4,SLIT2
GO:0040012	regulation of locomotion	2.66E-06	-12.837683	557	45	0.073051948	616	15891	MMP9,ANXA3,TRIM5,BVES,TMPRSS2,DDR2,PKP2,TGFB2,LAMA1,HAS2,PARD6B,WNT5A,THBS1,P2RY6,IGF1R,TFAP2A,SFRP1,TACSTD2,MET,AMOT,ENPP2,SLIT2,LRRK16A,CDK6,NOX4,CYP1B1,TDGF1,GRM1,APEX1,CGA,PIK3R1,ARHGAP5,CMKL1,RELN,EPHA1,PLCB1,NEXN,SEMA5A,BMP5,IL6,ADAM9,SERpine1,EFNA1,BMP4,HACE1
GO:2000145	regulation of cell motility	3.23E-06	-12.64193	508	42	0.068181818	616	15891	MMP9,ANXA3,DDR2,PKP2,TGFB2,LAMA1,HAS2,PARD6B,WNT5A,THBS1,P2RY6,IGF1R,TFAP2A,SFRP1,TACSTD2,MET,AMOT,ENPP2,SLIT2,LRRK16A,CDK6,NOX4,CYP1B1,TDGF1,GRM1,APEX1,CGA,PIK3R1,ARHGAP5,CMKL1,RELN,EPHA1,PLCB1,NEXN,SEMA5A,BMP5,IL6,ADAM9,SERpine1,EFNA1,BMP4,HACE1
GO:0001822	kidney development	3.91E-06	-12.451874	232	25	0.040584416	616	15891	DCN,RARB,BDNF,GRM1,ENPEP,EYA1,SDC4,PLCE1,SMAD5,HAS2,SIX4,RPGRIPL1,SGPL1,EPCAM,TMED10,NPNT,RRM2B,FGR2,GPC3,TFAP2A,ACTA2,SFRP1,TACSTD2,BMP4,SLIT2
GO:0001655	urogenital system development	4.21E-06	-12.377463	279	28	0.045454545	616	15891	DCN,RARB,BDNF,GRM1,ENPEP,EYA1,SDC4,PLCE1,SMAD5,HAS2,SIX4,RPGRIPL1,WNT5A,SGPL1,EPCAM,IGF1R,TMED10,RRM2B,FGR2,NPNT,GPC3,TP63,TFAP2A,ACTA2,SFRP1,TACSTD2,BMP4,SLIT2
GO:0040017	positive regulation of locomotion	4.29E-06	-12.359946	295	29	0.047077922	616	15891	MMP9,ANXA3,CGA,BVES,TMPRSS2,PIK3R1,ARHGAP5,DDR2,CMKL1,RELN,EPHA1,TGFB2,HAS2,SEMA5A,WNT5A,THBS1,P2RY6,IGF1R,IL6,TFAP2A,ADAM9,SERpine1,BMP4,MET,AMOT,SLIT2,LRRK16A,NOX4,TDGF1

GO:0009888	tissue development	4.56E-06	-12.297555	1430	89	0.144480519	616	15891	DCN,EXPH5,RSCO2,PTHLH,GRHL2,UGCG,MGP,DDR2,PKP2,SMAD5,HAS2,RPGRIPI1,CYP26A1,WNT5A,EYS,EPCAM,CDH1,NPNT,MARVELD2,TXNIP,MAP7,CDH3,ACVR2A,TACSTD2,PDE4D,TGM2,MET,CDK6,SLIT2,TAGLN,TYRP1,CYP1B1,TDGF1,THRIB,BNC2,CAV1,IRF6,CDON,DSP,NEXN,SEMA5A,LCP1,CXADR,PPP2CA,MSCE,NPPI,ADAM9,SEMA3C,FST,LAMC2,EDA2R,EYA1,FRMD6,SDC4,NOTCH2,TGFB2,FSHR,LAMA1,SIX4,THBS1,SPP1,IGF1R,NEBL,TP63,GPC3,FREM2,TFAP2A,ACTA2,SFRP1,ERBB3,MYOCD,INHBA,RARB,BDNF,GREM1,COPS2,KCNMA1,PRICKLE1,GPNMB,IFT80,BMP5,NCOA3,POSTN,FGFR2,IL6,TUFT1,EFNA1,BMP4,SFN
GO:0001657	ureteric bud development	5.14E-06	-12.178046	86	14	0.022727273	616	15891	EPCAM,RARB,BDNF,GREM1,FGFR2,NPNT,EYA1,GP C3,SDC4,SFRP1,SMAD5,TACSTD2,BMP4,SLIT2
GO:0072163	mesonephric epithelium development	5.91E-06	-12.038275	87	14	0.022727273	616	15891	EPCAM,RARB,BDNF,GREM1,FGFR2,NPNT,EYA1,GP C3,SDC4,SFRP1,SMAD5,TACSTD2,BMP4,SLIT2
GO:0072164	mesonephric tubule development	5.91E-06	-12.038275	87	14	0.022727273	616	15891	EPCAM,RARB,BDNF,GREM1,FGFR2,NPNT,EYA1,GP C3,SDC4,SFRP1,SMAD5,TACSTD2,BMP4,SLIT2
GO:0016477	cell migration	9.56E-06	-11.558432	659	49	0.079545455	616	15891	MMP9,EMR2,ENPEP,OLR1,ITGA6,CD177,SPOCK1,TGFB2,PLAT,SIX4,SH3KBP1,SGPL1,THBS1,WNT5A,SPP1,APBB2,FUT8,PRKG1,SDCBP,PDE4D,ZRANB1,MET,AMOT,LRRC16A,SLIT2,PARP9,CYP1B1,TDGF1,FAP,GREM1,PIK3R1,WWC1,RELN,CAV1,GPC6,SEMA5A,LCP1,CXADR,FERMT1,VCAN,IL6,PEX2,MATN2,DPP4,CXCL14,SEMA3C,EFNA1,COL1A2,SLC7A8

## Gene Ontology processes downregulated upon Alien knockdown

TermID	Term	Enrichment	logP	Genes in Term	Target Genes in Term	Fraction of Targets in Term	Total Target Genes	Total Genes	Gene Symbols
GO:0001525	angiogenesis	3.76E-13	-28.60836295	277	26	0.102362205	254	15891	VAV3,DLL4,HSPG2,MMRN2,TMEM100,FGF6,CALCRL,RAMP2,RASIP1,JAM3,FOXC2,ROBO4,NRARP,ECM1,CXCR4,EPHB2,SOX18,ADAM15,EGFL7,TIE1,CXCL12,ESM1,SOX17,PIK3CG,ANGPT2,NOS3
GO:0001568	blood vessel development	8.19E-12	-25.5280277	424	30	0.118110236	254	15891	GJA4,VAV3,DLL4,HSPG2,MMRN2,TMEM100,FGF6,CALCRL,RAMP2,RASIP1,PECAM1,JAM3,FOXC2,ROBO4,NRARP,HHEX,ECM1,CXCR4,EPHB2,CDH5,SOX18,ADAM15,EGFL7,TIE1,CXCL12,ESM1,SOX17,PIK3CG,ANGPT2,NOS3
GO:0048514	blood vessel morphogenesis	1.87E-11	-24.6999345	355	27	0.106299213	254	15891	VAV3,DLL4,HSPG2,MMRN2,TMEM100,FGF6,CALCRL,RAMP2,RASIP1,JAM3,FOXC2,ROBO4,NRARP,HHEX,ECM1,CXCR4,EPHB2,SOX18,ADAM15,EGFL7,TIE1,CXCL12,ESM1,SOX17,PIK3CG,ANGPT2,NOS3
GO:0001944	vasculature development	4.03E-11	-23.93527381	452	30	0.118110236	254	15891	GJA4,VAV3,DLL4,HSPG2,MMRN2,TMEM100,FGF6,CALCRL,RAMP2,RASIP1,PECAM1,JAM3,ROBO4,FOXC2,NRARP,HHEX,CXCR4,ECM1,EPHB2,CDH5,SOX18,ADAM15,EGFL7,TIE1,CXCL12,ESM1,SOX17,PIK3CG,ANGPT2,NOS3
GO:0072359	circulatory system development	6.19E-08	-16.59853729	727	33	0.12992126	254	15891	GJA4,DLL4,HSPG2,MMRN2,TMEM100,RASIP1,PECAM1,ECM1,EPHB2,ADAM15,CXCL12,SOX17,ANGPT2,NOS3,VAV3,PPARG,FGF6,CALCRL,RAMP2,JAM3,FOXC2,ROBO4,NRARP,HHEX,CD3,CXCR4,SALL1,CDH5,SOX18,TIE1,EGFL7,ESM1,PIK3CG
GO:0072358	cardiovascular system development	6.19E-08	-16.59853729	727	33	0.12992126	254	15891	GJA4,DLL4,HSPG2,MMRN2,TMEM100,RASIP1,PECAM1,ECM1,EPHB2,ADAM15,CXCL12,SOX17,ANGPT2,NOS3,VAV3,PPARG,FGF6,CALCRL,RAMP2,JAM3,FOXC2,ROBO4,NRARP,HHEX,CD3,CXCR4,SALL1,CDH5,SOX18,TIE1,EGFL7,ESM1,PIK3CG
GO:0048646	anatomical structure formation involved in morphogenesis	3.34E-07	-14.91306767	782	33	0.12992126	254	15891	DLL4,HSPG2,MMRN2,TMEM100,TLX2,RASIP1,HOXA11,ECM1,EPHB2,ADAM15,CXCL12,SOX17,ANGPT2,NOS3,VAV3,FGF6,CALCRL,RAMP2,FOXD1,JAM3,FOXC2,ROBO4,NRARP,HHEX,NRG3,CXCR4,SOX7,SALL1,SOX18,TIE1,EGFL7,ESM1,PIK3CG
GO:1901342	regulation of vasculature development	2.44E-06	-12.92520427	191	14	0.05511811	254	15891	HHEX,DLL4,ECM1,MMRN2,GATA2,TMEM100,NPR1,RAMP2,TIE1,ROCK1,MAPK7,ANGPT2,NOS3,FOXC2
GO:0009888	tissue development	3.34E-06	-12.60990093	1430	46	0.181102362	254	15891	GJA4,DLL4,HSPG2,MMRN2,UPK1A,OTOP1,TLX2,HOXA11,RAPIP1,PECAM1,CRABP2,EMP1,ECM1,VAMP5,GAP43,ASCL2,ADM15,CXCL12,SOX17,SMAD6,PTCH2,TGFB1,SMAD9,ERF,PPARG,CYP7B1,FGF6,HTN1,S100A4,FABP5,FOXD1,HOXB5,FOXC2,NRARP,CD3,HHEX,TCF7L1,CXCR4,SOX7,SALL1,STAT6,DBI,SOX18,KITLG,TIE1,ROCK1
GO:0045765	regulation of angiogenesis	4.26E-06	-12.36609144	173	13	0.051181102	254	15891	HHEX,DLL4,ECM1,MMRN2,GATA2,NPR1,RAMP2,TIE1,ROCK1,MAPK7,ANGPT2,NOS3,FOXC2
GO:0022904	respiratory electron transport chain	4.47E-06	-12.31804857	100	10	0.039370079	254	15891	NDUFC1,NDUFB6,NDUFB4,UQCRCQ,NDUFA1,NDUFA4,ATP5I,COX7B,ATP5E,ATP5J2
GO:0022900	electron transport chain	5.35E-06	-12.13854783	102	10	0.039370079	254	15891	NDUFC1,NDUFB6,NDUFB4,UQCRCQ,NDUFA1,NDUFA4,ATP5I,COX7B,ATP5E,ATP5J2
GO:0009653	anatomical structure morphogenesis	5.98E-06	-12.02711611	1933	56	0.220472441	254	15891	DLL4,HSPG2,MMRN2,TMEM100,DYNLL1,TLX2,OTOP1,RASIP1,HOXA11,HOXB8,CRABP2,SHANK3,PHLDA2,ECM1,EPHB2,GATA2,GAP43,ADAM15,FLI1,CXCL12,SOX17,ANGPT2,NOS3,TGFB1,VA3,SMAD9,ERF,CYP7B1,S100A4,FGF6,CALCRL,MAFB,RAMP2,FOXD1,DPYSL4,JAM3,HOXB5,FOXC2,ROBO4,NRARP,CD3,TCF7L1,HHEX,NRG3,CXCR4,SOX7,SALL1,STAT6,SOX18,EGFL7,TIE1,ROCK1,EFNA2,ESM1,PIK3CG,IGF2

GO:0048731	system development	8.61E-06	-11.66236583	3501	86	0.338582677	254	15891	DLL4,HSPG2,MMRN2,TMEM100,DYNLL1,OTOP1,RASIP1,EMP1,ECM1,ASCL2,HBZ,SMAD9,CYP7B1,PPARG,ERF,UQCRCQ,FGF6,S100A4,FABP5,RAMP2,JAM3,NRARP,HHEX,TCF7L1,NRG3,SALL1,STAT6,KITLG,ROCK1,LTK,ESM1,GJA4,TLX2,B2M,HOXB8,HOXA11,PECAM1,CRABP2,SHANK3,PHLDA2,SCARF1,EPHB2,GATA2,VAMP5,TOX2,GAP43,ADAM15,FLI1,CXCL12,COX7B,SOX17,SMAD6,ANGPT2,NOS3,TGFB1,PTCH2,VAV3,APLN,HTN1,DCHS1,CRIP2,VLDLR,LIN28A,CALCRL,SNRK,MAFB,STMN3,BEX1,FOXD1,NR5A2,DPVSL4,HOXB5,FOXC2,ROBO4,ID3,CXCR4,CDH5,DBI,SOX18,VWF,EGFL7,TIE1,EFNA2,PCDH12,PIK3CG,IGF2
GO:0001570	vasculogenesis	9.84E-06	-11.52875726	66	8	0.031496063	254	15891	HHEX,TMEM100,SOX18,TIE1,RAMP2,EGFL7,RASIP1,SOX17,GJA4,DLL4,TMEM100,SOX18,PECAM1,SOX17,HOXB5,FOXC2
GO:0003158	endothelium development	1.53E-05	-11.08686108	70	8	0.031496063	254	15891	TGFB1,NRARP,HHEX,DLL4,CXCR4,SALL1,CXCL12,RASIP1,HOXA11,FOXD1,FOXC2
GO:0048754	branching morphogenesis of an epithelial tube	2.15E-05	-10.7474937	145	11	0.043307087	254	15891	TGFB1,ESAM,PROCR,VWF,PROS1,PIK3CG,IGF2,ANGPT2,NOS3
GO:0009611	response to wounding	3.32E-05	-10.31222491	677	26	0.102362205	254	15891	TGFB1,VAV3,FABP5,HBE1,HBG1,PECAM1,LCP2,JAM3,GNG2,HHEX,ESAM,PROCR,VWF,PROS1,PIK3CG,IGF2,ANGPT2,NOS3
GO:0050878	regulation of body fluid levels	4.17E-05	-10.08509647	606	24	0.094488189	254	15891	TGFB1,VAV3,APLN,NPR1,HBE1,HBG1,PECAM1,LCP2,JAM3,GNG2,ESAM,EPHB2,SERPING1,GATA2,F2RL2,KCNMB3,VWF,PROCR,FLI1,PROS1,PIK3CG,IGF2,ANGPT2,NOS3

## Gene Ontology processes upregulated upon Terminator knockdown

TermID	Term	Enrichment	logP	Genes in Term	Target Genes in Term	Fraction of Targets in Term	Total Target Genes	Total Genes	Gene Symbols
GO:0006334	nucleosome assembly	9.08E-09	-18.5170082	117	14	0.052830189	265	15891	HIST1H2BB,HIST1H3F,HIST1H2AK,HIST1H1D,HIST1H2BK,HIST2H4B,HIST2H3D,HIST1H2AH,HIST1H2BL,HIST2H2AC,HIST1H2AB,HIST2H2AB,HIST1H1C,HIST1H2BO
GO:0031497	chromatin assembly	3.22E-08	-17.2500154	129	14	0.052830189	265	15891	HIST1H2BB,HIST1H3F,HIST1H2AK,HIST1H1D,HIST1H2BK,HIST2H4B,HIST2H3D,HIST1H2AH,HIST1H2BL,HIST2H2AC,HIST1H2AB,HIST2H2AB,HIST1H1C,HIST1H2BO
GO:0009653	anatomical structure morphogenesis	4.07E-08	-17.01637166	1933	64	0.241509434	265	15891	DCN,ENPEP,EPAS1,CRYAB,DYNLL1,SLIT3,GATA3,ADAMTS1,ISL1,GCM1,NPNT,DACT1,NOG,FAM20A,MMP14,CER1,MYL6,TACSTD2,TGM2,SOD1,WLS,DLX3,RSP03,LCP1,ANPEP,RBP4,MSX2,HSPB11,HAND1,AMBN,DLX5,VCAM1,ROCK1,PRKD1,HOXC13,EFNB2,EYA1,SLC1A3,TGFB2,APOB,SRPX2,COL6A3,CA2,FAM20C,FLRT3,TFAP2A,LGI1,RARB,SLC40A1,UNC5C,FMN1,RELN,TGFBI,REMF1,FGF16,ID3,SLC6A4,FRZB,TBX3,IGFBP5,DUOX2,NR2F2,NTRK2,IGF2
GO:0009887	organ morphogenesis	6.71E-08	-16.51722896	770	35	0.132075472	265	15891	DCN,EFNB2,EYA1,SLIT3,GATA3,ADAMTS1,TGFB2,ISL1,CA2,FAM20C,NOG,FAM20A,TFAP2A,CER1,RARB,SLC40A1,SOD1,FMN1,DLX3,REMF1,FGF16,RBP4,ID3,MSX2,HAND1,SLC6A4,FRZB,IGFBP5,TBX3,AMBN,DLX5,DUOXA2,NTRK2,IGF2,HOXC13
GO:0034728	nucleosome organization	9.14E-08	-16.20814693	140	14	0.052830189	265	15891	HIST1H2BB,HIST1H3F,HIST1H2AK,HIST1H1D,HIST1H2BK,HIST2H4B,HIST2H3D,HIST1H2AH,HIST1H2BL,HIST2H2AC,HIST1H2AB,HIST2H2AB,HIST1H1C,HIST1H2BO
GO:0065004	protein-DNA complex assembly	9.14E-08	-16.20814693	140	14	0.052830189	265	15891	HIST1H2BB,HIST1H3F,HIST1H2AK,HIST1H1D,HIST1H2BK,HIST2H4B,HIST2H3D,HIST1H2AH,HIST1H2BL,HIST2H2AC,HIST1H2AB,HIST2H2AB,HIST1H1C,HIST1H2BO
GO:0048513	organ development	1.65E-07	-15.61897966	2518	75	0.283018868	265	15891	DCN,ENPEP,EPAS1,CRYAB,DYNLL1,APOA2,SLIT3,GATA3,ADAMTS1,IGFBP7,ISL1,GCM1,NPNT,FAM20A,ADAMTS18,NOG,MMP14,MYL6,CER1,TACSTD2,TGM2,ZNF750,LOX,SOD1,MEIS2,CGA,LUM,DLX3,PLCE1,AFP,RSPO3,LCP1,RBP4,MSX2,HAND1,AMBN,DLX5,VCAM1,HOXC13,PTN,EFNB2,EYA1,LRP2,TGFB2,COL6A3,CA2,FA2,M20C,TFAP2A,CNFN,IFNE,RARB,PTCH2,UNC5C,SLC40A1,FMN1,VLDLR,RELN,PKDCC,TGFBI,FOS,REMF1,FGF16,ID3,FRZB,SLC6A4,IGFBP5,TBX3,DBI,ASPRV1,DUOX2,NR2F2,NTRK2,APOA1,IFI16,IGF2
GO:0006333	chromatin assembly or disassembly	1.83E-07	-15.51132887	148	14	0.052830189	265	15891	HIST1H2BB,HIST1H3F,HIST1H2AK,HIST1H1D,HIST1H2BK,HIST2H4B,HIST2H3D,HIST1H2AH,HIST1H2BL,HIST2H2AC,HIST1H2AB,HIST2H2AB,HIST1H1C,HIST1H2BO
GO:0048545	response to steroid hormone	4.43E-07	-14.62907318	350	21	0.079245283	265	15891	RARB,CRYAB,APOA2,SLIT3,GATA3,TGFB2,FOS,CRHBP,IGFBP7,ISL1,MSX2,SERPINA1,CA2,SLC6A4,FOSB,MMP14,NR2F2,CPN1,APOA1,IGF2,LOX
GO:0071824	protein-DNA complex subunit organization	6.03E-07	-14.32211738	163	14	0.052830189	265	15891	HIST1H2BB,HIST1H3F,HIST1H2AK,HIST1H1D,HIST1H2BK,HIST2H4B,HIST2H3D,HIST1H2AH,HIST1H2BL,HIST2H2AC,HIST1H2AB,HIST2H2AB,HIST1H1C,HIST1H2BO

GO:0030198	extracellular matrix organization	6.41E-07	-14.25963337	358	21	0.079245283	265	15891	DCN,TTR,TIMP2,LUM,MMP1,FGG,TGFB2,TGFBI,COL6A3,HAPLN1,CCDC80,MMP3,NPNT,VTN,TFAP2A,MMP14,VCAM1,FGB,FGA,ITGB6,LOX
GO:0006323	DNA packaging	6.49E-07	-14.24771039	164	14	0.052830189	265	15891	HIST1H2BB,HIST1H3F,HIST1H2AK,HIST1H1D,HIST1H2BK,HIST2H4B,HIST2H3D,HIST1H2AH,HIST1H2BL,HIST2H2AC,HIST1H2AB,HIST2H2AB,HIST1H1C,HIST1H2BO
GO:0043062	extracellular structure organization	6.71E-07	-14.21424575	359	21	0.079245283	265	15891	DCN,TTR,TIMP2,LUM,MMP1,FGG,TGFB2,TGFBI,COL6A3,HAPLN1,CCDC80,MMP3,NPNT,VTN,TFAP2A,MMP14,VCAM1,FGB,FGA,ITGB6,LOX
GO:0032501	multicellular organismal process	8.84E-07	-13.9389819	5854	136	0.513207547	265	15891	CUBN,PSG5,PSG1,EPAS1,CRYAB,SPIN2A,GATA3,FGG,ADAMTS1,PSG3,CRHBP,ISL1,GCM1,NPNT,IL6ST,ADAMTS18,FAM20A,FOSB,NOG,TACSTD2,TGM2,ZNF750,TFPI2,SMARCA2,SOD1,MEIS2,LUM,WLS,PLCE1,OR4D10,ARHGDIB,RSPO3,LCP1,DPPA3,OR4C12,RBP4,MMP3,DPP4,VCAM1,ROCK1,APOA4,PRKD1,HOXC13,EFNB2,SLC7A7,SLC1A3,COL6A3,MYL4,CA2,FAM20C,FLRT3,CNFN,VTCN1,TMPRSS11E,CNGA1,RARB,LG1,PTCH2,HIST1H3F,UNC5C,VLDDL,RELN,PKDCC,TGFB1,PSG4,GPX2,IGFBP5,FGA,ASPRV1,NTRK2,APOA1,DCN,ENPEP,DYNLL1,APOA2,SLIT3,IGFBP7,BRK1,PSG7,DACT1,MMP14,MYL6,CER1,ITGB6,MEIS1,ARSE,LOX,TIMP2,CGA,CTSE,DLX3,MMP1,AFP,ANPEP,MSX2,HAND1,AKR1D1,AMB,DLX5,DPPA5,PTN,EYA1,LRP2,PSG2,TGFB2,APOB,MR1,SRPX2,FFAR4,PRTG,GPR155,TFAP2A,HIST2H3D,COX7B,ATP2B1,IFNE,SLC40A1,FMN1,MBNL3,FOS,OR13H1,FREM1,FGF16,ID3,TIMP3,SERPINA1,SLC6A4,FRZB,TBX3,DBI,UPK2,FGB,DUOX2,NR2F2,IFI16,IGF2
GO:0007565	female pregnancy	3.75E-06	-12.49282336	164	13	0.049056604	265	15891	PSG5,PSG4,SOD1,PSG1,PSG7,FOSB,PSG2,NR2F2,FOS,PSG3,IGF2,IGFBP7,CRHBP
GO:0033993	response to lipid	3.84E-06	-12.47049497	646	28	0.105660377	265	15891	DCN,RARB,FMO1,CRYAB,APOA2,SLIT3,GATA3,MGST2,TGFB2,FOS,APOB,CRHBP,IGFBP7,ISL1,RBP4,MSX2,CA2,SERPINA1,SLC6A4,FOSB,MMP14,VCAM1,S100A14,NR2F2,CPN1,APOA1,IGF2,LOX
GO:0009888	tissue development	4.73E-06	-12.26079522	1430	47	0.177358491	265	15891	DCN,EYA1,GATA3,TGFB2,ISL1,GCM1,CA2,FAM20C,NPNT,NOG,FAM20A,MMP14,TFAP2A,MYL6,CER1,CNFN,TACSTD2,TGM2,ZNF750,RARB,PTCH2,SLC40A1,SOD1,WLS,LUM,DLX3,PKDCC,TGFB1,FOS,RSPO3,LCP1,RBP4,ID3,MSX2,HAND1,FRZB,DBI,IGFBP5,TBX3,AMBN,UPK2,DLX5,DUOX2,ROCK1,NR2F2,HOXC13,PTN
GO:0010035	response to inorganic substance	7.43E-06	-11.80948641	352	19	0.071698113	265	15891	MTTP,SOD1,MT2A,CRYAB,FGG,FOS,APOB,S100A16,CRHB,SERPINA1,CA2,FOSB,TFAP2A,HP,FGB,VCAM1,DUOX2,FGA,APOA4
GO:0060429	epithelium development	8.48E-06	-11.67789027	909	34	0.128301887	265	15891	EYA1,GATA3,TGFB2,GCM1,CA2,FAM20C,NPNT,NOG,MMP14,TFAP2A,CER1,CNFN,TACSTD2,TGM2,ZNF750,PTCH2,RARB,SLC40A1,SOD1,DLX3,RSPO3,LCP1,MSX2,ID3,HAND1,FRZB,IGFBP5,TBX3,DBI,UPK2,DLX5,ROCK1,NR2F2,HOXC13

## Gene Ontology processes downregulated upon Terminator knockdown

TermID	Term	Enrichment	logP	GenesInTerm	TargetGenesInTerm	FractionOfTargetsInTerm	TotalTargetGenes	TotalGenes	GeneSymbols
GO:0006891	intra-Golgi vesicle-mediated transport	0.000166667	-8.699510081	32	4	0.028776978	139	15891	RAB6A,GOSR1,COPB1,COG5
GO:0007616	long-term memory	0.00182803	-6.304516379	28	3	0.021582734	139	15891	TAC1,PJA2,ADCY8
GO:0098900	regulation of function potential	0.006452518	-5.043284832	14	2	0.014388489	139	15891	TAC1,GJA5
GO:0048193	Golgi vesicle transport	0.007383803	-4.9084664	195	6	0.043165468	139	15891	USE1,RAB6A,GOSR1,COPB1,BLZF1,COG5
GO:0006617	SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition	0.00874709	-4.739034257	1	1	0.007194245	139	15891	SRP54
GO:1900102	negative regulation of endoplasmic reticulum unfolded protein response	0.00874709	-4.739034257	1	1	0.007194245	139	15891	CREBRF
GO:0060816	random inactivation of chromosome	0.00874709	-4.739034257	1	1	0.007194245	139	15891	RLIM
GO:2000691	negative regulation of cardiac muscle cell myoblast differentiation	0.00874709	-4.739034257	1	1	0.007194245	139	15891	PRICKLE1
GO:0086053	AV node cell to bundle of His cell communication by electrical coupling	0.00874709	-4.739034257	1	1	0.007194245	139	15891	GJA5
GO:2000690	regulation of cardiac muscle cell myoblast differentiation	0.00874709	-4.739034257	1	1	0.007194245	139	15891	PRICKLE1
GO:0051892	negative regulation of cardioblast differentiation	0.00874709	-4.739034257	1	1	0.007194245	139	15891	PRICKLE1
GO:1990029	vasomotion	0.00874709	-4.739034257	1	1	0.007194245	139	15891	GJA5
GO:0003294	atrial ventricular junction remodeling	0.00874709	-4.739034257	1	1	0.007194245	139	15891	GJA5
GO:1902897	regulation of postsynaptic density protein 95 clustering	0.00874709	-4.739034257	1	1	0.007194245	139	15891	CRIP1
GO:0006616	SRP-dependent cotranslational protein targeting to membrane, translocation	0.00874709	-4.739034257	1	1	0.007194245	139	15891	SRP54
GO:0010477	response to sulfur dioxide	0.00874709	-4.739034257	1	1	0.007194245	139	15891	ICAM1
GO:0003151	outflow tract morphogenesis	0.009510881	-4.655318804	50	3	0.021582734	139	15891	NPY5R,GJA5,NPY1R
GO:0021591	ventricular system development	0.013018734	-4.341365846	20	2	0.014388489	139	15891	KIF27,RPGRIP1L
GO:0006184	GTP catabolic process	0.014549704	-4.230184623	226	6	0.043165468	139	15891	SRP54,RAB18,RAB6A,RAB38,TUBB8,MX2

## Identification of Novel Long Noncoding RNAs Underlying Vertebrate Cardiovascular Development

Leo Kurian, Aitor Aguirre, Ignacio Sancho-Martinez, Christopher Benner, Tomoaki Hishida, Thai B. Nguyen, Pradeep Reddy, Emmanuel Nivet, Marie N. Krause, David A. Nelles, Concepcion Rodriguez Esteban, Josep M. Campistol, Gene W. Yeo and Juan Carlos Izpisua Belmonte

*Circulation*. 2015;131:1278-1290; originally published online March 4, 2015;  
doi: 10.1161/CIRCULATIONAHA.114.013303

*Circulation* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231  
Copyright © 2015 American Heart Association, Inc. All rights reserved.  
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circ.ahajournals.org/content/131/14/1278>

Data Supplement (unedited) at:

<http://circ.ahajournals.org/content/suppl/2015/03/04/CIRCULATIONAHA.114.013303.DC1.html>

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

**Reprints:** Information about reprints can be found online at:  
<http://www.lww.com/reprints>

**Subscriptions:** Information about subscribing to *Circulation* is online at:  
<http://circ.ahajournals.org//subscriptions/>