Enhancers and long noncoding RNAs (lncRNAs) are key determinants of lineage specification during development. Here, we evaluate remodeling of the enhancer landscape and modulation of the lncRNA transcriptome during mesendoderm specification. We sort mesendodermal progenitors from differentiating embryonic stem cells (ESCs) according to Eomes expression, and find that enhancer usage is coordinated with mesendoderm-specific expression of key lineage-determining transcription factors. Many of these enhancers are associated with the expression of IncRNAs. Examination of ESC-specific enhancers interacting in three-dimensional space with mesendoderm-specifying transcription factor loci identifies MesEndoderm Transcriptional Enhancer Organizing Region (Meteor). Genetic and epigenetic manipulation of the Meteor enhancer reveal its indispensable role during mesendoderm specification and subsequent cardiogenic differentiation via transcription-independent and -dependent mechanisms. Interestingly, Meteor-deleted ESCs are epigenetically redirected towards neuroectodermal lineages. Loci, topologically associating a transcribed enhancer and its cognate protein coding gene, appear to represent therefore a class of genomic elements controlling developmental competence in pluripotency.
A complex molecular program controls cardiac differentiation and morphogenesis during development. Understanding the gene regulatory networks (GRNs) governing cardiac organogenesis could lead to innovations for the treatment of both inherited and acquired heart disease. In addition, induction of cardiac regeneration via either cell-based therapies or reactivation of dormant endogenous mechanisms has engendered considerable interest over the past decade. Both approaches require a deep knowledge of the molecular and cellular events regulating cardiac specification and differentiation. However, transcriptional profiling of individual transient cellular intermediates in the developing heart is extremely challenging and requires the use of in vitro cellular systems. During gastrulation, epiblast cells undergo epithelial-to-mesenchymal transition (EMT) and ingress through the primitive streak (PS). This transient precursor cell population, referred to as the mesendoderm (ME), is the source of the definitive endoderm and the mesoderm. Nascent mesoderm cells then rapidly migrate from the posterior side of the embryo to the anterior side, become specified to the cardiac lineage, and ultimately generate the embryonic heart. Mesendodermal precursors are characterized by the expression of genes such as Eomesodermin (Eomes), Goosecoid (Gsc), and LIM-homeobox1 (Lhx1). In particular, Eomes, a T-box transcription factor (TF), is critical for ME specification between embryonic day 6.5–7.5 in the mouse. Eomes expression marks the earliest cardiac mesoderm and dictates the formation of cardiac precursors through regulating the master TF Mesoderm posterior 1 (Mesp1). The transient formation of the ME is under the control of a specialized GRN consisting of cell-fate determining TFs that interact at target sequences known as enhancers. Moreover, enhancers are an important class of distal regulatory elements that are key information processing units within the genome, controlling the precise spatiotemporal expression of their target protein-coding genes (PCGs). Recently, regions of the mammalian genome comprising multiple enhancers have been identified and termed super-enhancers (SEs). SEs are typically an order of magnitude larger than typical enhancers (TEs) and are highly enriched with regulatory TFs and chromatin marks (i.e., H3K27ac). They are master regulators of developmental and cell identity genes, which are critical for cell fate determination and differentiation. Through integrating upstream context-specific developmental signals, TF-bound TE-SE associations mediate appropriate gene programs required to mark and specify mesendodermal fate.

Progress in high-throughput sequencing has advanced our understanding of genome organization and regulation. Only 2% of the genome appears to code for proteins. The remaining 98% represents the noncoding fraction of the genome. Most of the noncoding genome is transcribed into RNAs. In particular, long noncoding RNAs (lncRNAs) represent an important class of regulatory molecules. lncRNAs are typically transcribed by RNA polymerase II and are usually multixonic and polyadenylated. Interestingly, lncRNAs have been shown to be expressed in unique cell types, for instance across various stages of differentiation, suggesting their involvement in regulating cell fate. An important subset of RNAs is associated with enhancers, and named enhancer RNAs (eRNAs). They exist as two different transcripts: bidirectional non-polyadenylated transcripts; and unidirectional, multixonic, spliced, polyadenylated transcripts. Recent studies have demonstrated that targeted degradation of eRNAs is sufficient to reduce expression of adjacent PCGs. In particular, eRNAs appear to be involved in the formation and stabilization of the loop between the enhancer and the promoter in a typical Cis-regulatory manner. In addition, chromatin conformation capture approaches have shown that the genome is hierarchically organized into larger domains known as topologically associating domains (TADs). TADs are linear DNA segments that form independent units in nuclear three-dimensional space. Importantly, TAD boundaries are well conserved across species and cell-types. Disrupting boundaries results in spurious interactions between promoters and enhancers normally residing in different TADs, leading to transcriptional dysregulation. Interestingly, transcribed DNA elements, including enhancers, are emerging as potential regulators of TAD formation. However, an important debate revolves around the nature of lncRNA functions. Whether the simple act of transcription from the lncRNA locus or the mature lncRNA transcript results in observed phenotypes is still unclear.

Within this context, mouse embryonic stem cells (mESCs) harboring reporter genes transcribed from developmental TF promoters represent a unique system to model the formation of cellular intermediates such as the ME precursors. More precisely, the enhancer and the lncRNA landscapes governing ME specification and thereby cardiac differentiation has not been investigated. Here, we use an Eomes reporter mESC line to assess remodeling of the enhancer landscape and to profile the lncRNA transcriptome during ME specification. We identify a large number of previously uncharacterized enhancer-associated lncRNAs. Examination of ESC-specific enhancer-associated lncRNA loci within mesendodermal TADs identified an Eomes-interacting locus, which was named MesEndoTranscriptional Enhancer Organizing Region (Meteor). Interestingly, the lncRNA associated to the Meteor enhancer corresponds to a previously described pluripotency-associated lncRNA. Meteor deletion and epigenetic manipulation reveals its indispensable role during ME determination and subsequent cardiogenic differentiation, supporting a predetermined role for this class of genomic elements in programming developmental competence and ESC specification during development.

**Results**

**Early cell fate specification in mesendodermal progenitors.** We utilized an Eomes reporter mESC line engineered to carry an EGFP cassette inserted into the transcriptional start site of the endogenous Eomes gene (EomesEGFP ESCs) (Supplementary Fig. 1a). These cells were induced to differentiate using the hanging drop model. This method allows the stepwise differentiation of ESCs towards the cardiogenic lineage, generating mesendoderm precursor cells (MEPC), cardiac precursor cells (CPCs) and ultimately differentiated cardiomyocytes (CMs) (Supplementary Fig. 1b). Initially, embryoid bodies (EBs) were harvested every 12 h during a 10-day period. Differentiation was accompanied by the downregulation of pluripotency-associated genes, the transient induction of ME and cardiac mesoderm specifying TFs Eomes, T and Mesp1, and finally the expression of markers of cardiac differentiation including Myh6 and Myh7 (Supplementary Fig. 1c). Terminal differentiation resulted in a significant number of beating EBs at both day 8 and 10 (Supplementary Fig. 1d). Importantly, the EomesEGFP reporter was able to mark Eomes-expressing cells at day 3 of differentiation, a time point at which Eomes is maximally expressed and specifies the nascent mesoderm (Supplementary Fig. 1e). Flow cytometry analysis indicated that half of the differentiating cells at day 3 commit to ME (Supplementary Fig. 1f, g).

We next isolated Eomes-positive (Eo+) and Eomes-negative (Eo−) cells at day 3 using fluorescence-activated cell sorting (FACS) (Supplementary Fig. 2a) and measured the expression of pluripotency genes (Nanog, Sox2, Oct4), markers of ME (Eomes, Lhx1, Mixl1), cardiac mesoderm (Mesp1, Gata4, Nx2-5) and neuroectoderm (Pax6, Nkx6-3, Neurog3) (Supplementary Fig. 2b). As compared to undifferentiated ESCs, significant expression of
**Fig. 1** Global assessment of the transcriptome during mesendoderm specification. a Pie chart showing composition of the Poly (A)+ transcriptome, Protein Coding Genes (PCG, blue), Ensembl lncRNAs (yellow) and non-annotated lncRNAs (red). b Kernel density plot of coding potential (Gene ID score) of PCGs, Ensembl lncRNAs and non-annotated lncRNAs. c Box plot whiskers of transcript abundance (FPKM) of PCGs (blue), Ensembl lncRNAs (yellow) and non-annotated lncRNAs (red). p values were calculated using a two-tailed t test. (**P < 0.0001). d Expression heatmap of representative PCGs in ESC, Eo− and Eo+. Markers of pluripotency, Primitive Streak/Mesendoderm (PS/ME), cardiac mesoderm and early ectoderm/neuroectoderm are shown. e Hierarchical clustering of Ensembl lncRNA expression across ESC, Eo− and Eo+. Enriched GO terms and example PCGs are shown to the right. f Hierarchical clustering of non-annotated lncRNA expression across ESC, Eo−, and Eo+. Enriched GO terms and example PCGs are shown to the right. g Kernel density plot of the specificity of PCG and lncRNA assessed by quantifying the normalized difference of expression in the three conditions (ESC, Eo−, Eo+). h Co-activation of selected PCGs and divergent lncRNAs involved in ME specification. i RNAseq reads in ESC, Eo− and Eo+ at the Lhx1 locus. A divergent lncRNA (Lhx1os) is shown.
mesendodermal and cardiac mesodermal genes was measured in Eo⁺ cells. In addition, neuroectoderm gene expression was higher in Eo⁻ cells than in Eo⁺ cells. To validate these subpopulations for subsequent genome-wide chromatin immunoprecipitation followed by sequencing (ChiP-Seq) analysis, we performed ChiP-qPCR using antibodies against H3K4me3 (associated with active promoters) and H3K27Ac (associated with active enhancers). Primers were designed within known promoter and enhancer regions associated with pluripotency (Nanog) and ME (Eomes) (Supplementary Fig. 2c). As expected, the Nanog promoter and the associated distal enhancer were enriched with H3K4me3 and H3K27Ac respectively in pluripotent ESCs. On the other hand, the Eomes promoter and enhancer were enriched with the H3K4me3 and H3K27Ac marks in the Eo⁺ sorted cells (Supplementary Fig. 2d). Our data thus indicates that Eo⁺ cells express a unique transcriptional and enhancer signature reflecting their potential to become ME-derived lineages, including cardiac mesoderm.

Transcriptome assessment during mesendoderm specification.
To characterize the transcriptome, and in particular the long noncoding transcriptome, in pluripotent ESCs and in sorted Eo⁺ and Eo⁻ cells at day 3 of differentiation, we performed very deep sequencing (>500 million reads per sample) coupled to ab initio reconstruction (Supplementary Fig. 3a). We integrated our reconstructed transcripts with the Ensembl gene annotation. Using this pipeline, we reconstructed 22,187 transcripts of which 16,440 corresponded to annotated PCGs. In addition, 5747 lncRNAs were identified. This included 1913 previously annotated lncRNAs and 3834 multixonic non-annotated lncRNAs (Fig. 1a; Supplementary Data 1). The non-annotated lncRNAs encode minimal and comparable protein coding potential to Ensembl-annotated lncRNAs (Fig. 1b). At the end, we disregarded any transcripts with a coding potential score greater than 4. Ensembl and non-annotated lncRNAs were globally expressed at significantly lower levels than PCGs (Fig. 1c). Unsupervised hierarchical clustering of all PCGs, Ensembl annotated lncRNAs and non-annotated lncRNAs identified three distinct clusters in ESCs, Eo⁻ and Eo⁺ cells (Supplementary Fig. 3b), demonstrating that the transcriptome was representative of the developmental events associated with ME specification.

LncRNA exons were determined to be less conserved than PCG exons. In contrast, promoters of both annotated and non-annotated lncRNAs were significantly more conserved than the neutrally evolving genomic background (Supplementary Fig. 3c). We proceeded to determine whether this reflected evolutionary constraint at lncRNA promoters impacting on specific developmental cellular intermediates during cardiac mesoderm differentiation. We used in this analysis the current RNA-Seq datasets and two previously generated datasets for cardiac precursor cells and the adult heart. We first identified the proximal promoter DNA sequences for PCGs, Ensembl and non-annotated lncRNAs in each of these five populations, i.e., ESCs, Eo⁻, Eo⁺, CPC, and adult heart, defining four inferred branch points. We next calculated the mean level of evolutionary constraint for each set of promoters (Supplementary Fig. 3d). We found that promoters of ME-associated transcripts were significantly more constrained than those specific to populations at other stages of cardiac differentiation. Furthermore, non-annotated lncRNA promoters specific to Eo⁺ cells were significantly more constrained than promoters of non-annotated lncRNAs expressed in Eo⁻ cells with this difference not detectable for Ensembl lncRNAs. Interestingly, promoter conservation at non-annotated lncRNAs recapitulates the hourglass model of development, and supports therefore an evolutionary conserved role for non-annotated lncRNA loci with respect to ME specification.

We next analyzed the expression of a series of PCGs that are typically associated with pluripotency, primitive streak/ME, cardiac mesoderm, early ectoderm and neuroectoderm. ME and cardiac mesoderm TFs, including for instance Eomes, T, and Mesp1, were highly enriched in Eo⁺ cells while pluripotency TFs were enriched in ESCs (Fig. 1d). In contrast, the Eo⁻ cells were not associated with a well-defined transcription signature. Importantly, PCG expression was characterized by stage-specific induction or repression of key lineage-determining TFs (Supplementary Fig. 4a). Unsupervised clustering yielded specific groups of PCGs expressed specifically within each cell population (Supplementary Fig. 4b). These cell population-specific clusters were associated to expected gene ontology (GO) terms (Supplementary Data 2). For example, PCGs enriched in Eomes-expressing cells were strongly associated with processes linked to gastrulation and mesoderm formation. Extending this analysis to the long noncoding transcriptome, we found that both Ensembl and non-annotated lncRNAs, encompassing all known biotypes (Supplementary Fig. 4c), exhibited striking stage-specific expression (Fig. 1e, f; Supplementary Data 3). We determined GO terms for the nearest PCGs relative to Ensembl and non-annotated lncRNAs enriched in Eomes-expressing cells (Fig. 1e, f). Interestingly, non-annotated lncRNAs enriched in Eo⁺ cells were proximal to PCGs specifically linked to heart processes, implicating non-annotated lncRNAs as potentially important and specific regulators of cardiac mesoderm specification and differentiation. Furthermore, non-annotated lncRNAs exhibited greater cell type-specific expression as compared to PCGs and annotated lncRNAs (Fig. 1g). Recently, it has been demonstrated that the expression of the lncRNA biotype known as divergent lncRNAs strongly correlates with that of their cognate developmental PCGs. We therefore selected and assessed the expression of six key ME and ectoderm TFs, in addition to their divergently expressed lncRNAs (Fig. 1h). A prototypic example is Hhex and its divergent noncoding transcript Hhexos (Fig. 1i). In agreement with previous findings, divergent lncRNAs were exquisitely correlated in their expression with their paired PCG (Supplementary Fig. 4d).

**Tissue-enrichment characteristics of mesendodermal lncRNAs.**
Very deep sequencing of RNA samples obtained from defined cell subpopulations allows the identification of noncoding transcripts that usually escape previous annotation. These transcripts are more likely to demonstrate high cell-type and tissue specificity. Many of the non-annotated lncRNAs exhibited particular enrichment in ME-committed cells (Fig. 1f). We therefore suspected that these transcripts could also be enriched in ME-derived tissues, including the heart. To evaluate this possibility, we selected all transcripts that were significantly expressed in either the Eo⁻ or Eo⁺ cells. We then computationally mapped twelve mouse ENCODE RNA-Seq datasets obtained from the heart and eleven non-cardiac tissues using the Ensembl and our non-annotated lncRNA annotations. Expression data in Fig. 2a identify heart-enriched transcripts in Eo⁻ and Eo⁺ cells, highlighted in red. Importantly, when we directly compared heart enrichment in Eo⁻ and Eo⁺ cells populations, we found that non-annotated lncRNAs exhibited greater heart enrichment than that of PCGs and Ensembl lncRNAs (Fig. 2b, c). These findings suggest that non-annotated lncRNAs enriched in Eo⁺ cells may mediate important regulatory functions for cardiogenic differentiation. To further explore tissue specificity, we repeated this enrichment analysis individually for the eleven non-cardiac tissues derived from mesoderm, endoderm, and ectoderm, and...
directly compared each individual score to that found for the heart (Fig. 2d). Strikingly, the heart is preferentially associated with Eo+ enriched previously non-annotated lncRNAs, discovered within the frame of this study, whereas this is not the case for PCGs and Ensembl lncRNAs. These findings emphasize again the exquisite cell and tissue-specificity of non-annotated lncRNA expression, and support the notion that these noncoding transcripts represent important cardiogenic factors. An example of an Eo+ cell-enriched transcript that is exclusively expressed in the adult heart as compared to other adult tissues is given in Fig. 2e. Mesendodermal transcripts presenting such a pattern of expression might therefore be involved in the maintenance of cardiac identity in the adult heart.

Enhancer landscape remodeling during mesendoderm specification. Enhancer elements are marked by H3K27Ac, H3K4me1 and by Mediator occupancy. Since these marks largely overlap in
previous ChIP-Seq studies\(^\text{10}\), we used H3K27Ac for the identification of enhancers in pluripotent ESC and in \(E^{+}\) and \(E^{-}\) cells. To identify SEs, we used the ROSE algorithm\(^\text{10}\), which stitches together individual enhancers within 12.5Kb of each other to identify a single continuous genomic locus. Stitched enhancers with a H3K27Ac value above a cut-off defined as the point where the slope of the distribution plot of H3K27Ac ChIP-Seq intensity is 1 are designated SEs\(^\text{10}\) (Fig. 3a). Using this distribution plot all remaining genomic loci with ChIP-Seq signal intensity below 1 are classified as TEs. Examples of TE and SE regions are depicted in Supplementary Fig. 5a, b, respectively. As expected, SEs were globally more enriched with H3K27Ac as compared to TEs (Fig. 3b) and spanned much larger genomic regions (Fig. 3c).

To dissect the enhancer state transition that governs mesendodermal gene expression programs, we clustered TE and SE according to their activity across the three cell populations (Fig. 3d, e; Supplementary Fig. 5c, d). A larger set of both TEs and SEs was active in pluripotent ESCs than in the two committed populations. Consistent with previous observations\(^\text{32}\), a gradual restriction in enhancer usage during differentiation was observed,
with Eomes-fated cells exhibiting the smallest enhancer repertoire (Supplementary Fig. 5e, f). Moreover, a small percentage of TEs and SEs were uniquely active in either Eo+ or Eo− cells (Fig. 3f).

Next, we performed a GO analysis of PCGs adjacent to TEs and SEs in each of the three subpopulations (Supplementary Data 4). TEs specific to Eo+ cells were associated with PCGs linked to relevant terms such as mesoderm formation and heart development (Fig. 3d; Supplementary Fig. 5g). Interestingly, active enhancers in this cell population were adjacent to master regulators of ME and cardiac mesoderm specification, such as Eomes and Mesp1. SEs identified in ESCs were linked to key regulatory TFs associated with pluripotency, in particular Sox2 and Klf4. Moreover, six SEs were found uniquely active in Eo+ cells. Although this small number did not allow a GO analysis to be performed, PCGs adjacent to these SE were master lineage determining TFs such as Id3, Pitx2, and Otx2 (Fig. 3e; Supplementary Fig. 5h). We then quantified the expression of PCGs proximal to either TEs or SEs across the three cell populations. In ESCs, PCGs proximal to SEs were typically more expressed than TE-associated PCGs (Fig. 3g). In addition, we investigated the dynamic changes in PCG expression during cell fate determination by measuring expression of PCGs adjacent to TEs or SEs modulated during ESC specification in to Eo+ cells (Fig. 3h).

Interestingly, the transition from ESC to mesendodermal specification was accompanied by a greater downregulation in expression of PCGs linked to inactivated SEs than to inactivated TEs (Fig. 3h; top panel). Finally, a significant activation in expression was observed for PCGs associated to TEs and SEs inactive in ESCs and activated in Eo+ cells (Fig. 3h; bottom panel). Altogether, these results demonstrate that dynamic remodeling at enhancer loci during ME specification is accompanied by a corresponding modulation of cognate PCGs.

Enhancers are typically associated with the production of lncRNAs, which are believed to contribute to enhancer function. We therefore categorized lncRNAs according to their association with regions marked by either H3K4me3 or H3K27Ac. lncRNAs were classified as being associated with either a canonical promoter signature (H3K4me3, demethylated pLncRNAs) or an active enhancer (H3K27Ac, TE or SE lncRNAs) (Fig. 4a). Non-annotated lncRNAs were found to be more associated with both TE (60%) and SE (10%) signatures when compared to annotated Ensembl lncRNAs (18% and 4% associated with TE and SE, respectively) (Fig. 4b). Enhancer-associated lncRNAs were expressed at lower levels as compared to promoter-associated lncRNAs (Fig. 4c). We then visualized pLncRNA, TE lncRNA and SE lncRNA expression across the three cell populations (Fig. 4d; Supplementary Data 5). LncRNA expression segregates in defined cell-specific patterns. Moreover, the number of cell-specific enhancer-associated lncRNAs expressed in Eo+ cells, both TE lncRNAs and SE lncRNAs, is much lower than in the two other populations. Interestingly, both TE and SE associated lncRNAs exhibit greater cell-type specificity as compared to pLncRNAs (Fig. 4e). Examples of the three classes of lncRNAs are depicted in Fig. 4f–h. These highly specialized expression profiles support specific transcriptional functions for enhancer-associated lncRNA loci during cell-fate determination and differentiation.

Mesendoderm-specifying loci in pluripotent ESCs. In addition to the enhancer landscape, chromatin structure and topology is emerging as a critical regulatory feature in pluripotent stem cells that subsequently dictates cell-fate determination and lineage specification17. Recently, it emerged that lncRNAs associated with cis-regulatory sequences could play important roles in the nuclear organizing processes that dictate cell fate18. We hypothesized that enhancers associated with lncRNAs expressed specifically in pluripotent ESCs, and distal to mesendodermal TFs, may represent critical functional elements dictating developmental competence and ultimately ME specification. We manually selected three mesendodermal TFs, namely Eomes, Sox17 and Gsc, based on their location adjacent to an active enhancer expressing an associated lncRNA exclusively in pluripotent ESCs (Table 1). Interestingly, both the enhancer activity, marked by H3K27Ac, and the expression of the associated lncRNA were specific to the ESC stage, and neither the enhancer nor expression of the lncRNA was activated in Eo+ cells (Fig. 5a–d). Conversely, the adjacent TEs were lowly expressed in ESCs and significantly upregulated in Eo+ cells. At the pluripotent stage, cell-type invariant topologically associating domains (TAD) are established, which are critical for configuring the chromatin structure ensuring correct temporal and spatial communication between distal enhancers and their cognate cell-fate determining TFs. We thus utilized publicly available high-throughput conformation capture (Hi-C) datasets from murine ESCs to interrogate the topological nature of the different loci containing Eomes, Sox17, and Gsc. All three regions were encompassed within highly interacting chromatin domains (Fig. 5e–g). These data indicated the presence in pluripotent ESCs of chromatin loops that place the enhancers and the mesendodermal TF promoters in close proximity within the three-dimensional nuclear architecture. Altogether, this suggested these loci could play fundamental roles in subsequent cis-regulatory events occurring during ME lineage specification.

The Meteor locus dictates mesendoderm specification. To evaluate the importance of the identified loci for mesendodermal cell fate determination, we selected the enhancer locus upstream of Eomes for a detailed analysis. We named the enhancer Meteor...
Fig. 4 Identification and characterization of promoter, typical and super enhancer associated long noncoding RNAs during mesendoderm specification. 

**a** Pie chart showing distribution of Ensembl (yellow) and non-annotated (red) lncRNAs associated with a canonical promoter signature (H3K4me3, green), typical enhancer signature (blue) or super enhancer signature (dark red). 

**b** Percentage of TE lncRNAs (blue) and SE lncRNAs (dark red) in Ensembl and non-annotated lncRNAs. 

**c** Box plot whiskers of transcript abundance (FPKM) of plncRNAs (green), SE lncRNAs (dark red), TE lncRNAs (blue) and unmarked lncRNAs (black). Box plot whiskers show median value and 10–90 percentiles. p values were calculated using a two-tailed t test. (**P < 0.01**). 

**d** Hierarchical clustering of plncRNA, TE lncRNA and SE lncRNA expression across ESC, Eo⁻ and Eo⁺. 

**e** Kernel density plot of the specificity of plncRNAs, TE lncRNAs and SE lncRNAs assessed by quantifying the Normalized difference of expression in the three conditions (ESC, Eo⁻, Eo⁺). RNAseq and H3K4me3 reads in ESCs, Eo⁻ and Eo⁺ cells for example plncRNA (ALIEN). RNAseq and H3K27Ac reads in ESCs, Eo⁻ and Eo⁺ cells for example TE lncRNA (XLOC_053874). RNAseq and H3K27Ac reads in ESCs, Eo⁻ and Eo⁺ cells for example SE lncRNA (XLOC_050466).
for MesEndoderm Transcriptional Enhancer Organizing Region. The enhancer was highly active in ESCs, coupled with high expression of its associated IncRNA, while Eomes was not significantly expressed at this pluripotent stage. Conversely, Eomes was enriched and the Meteor IncRNA was depleted in Eos+ cells (Supplementary Fig. 6a, b). The expression kinetics during cardiogenic differentiation confirmed the downregulation of the Meteor IncRNA during differentiation concomitant with the transient induction of Eomes (Supplementary Fig. 6c). Meteor was bound in pluripotent ESCs by Oct4 and Sox2 but not Nanog (Supplementary Fig. 6d). To probe the molecular and cellular function of this locus, we utilized CRISPR-Cas9 to delete the Meteor enhancer in Eomes-EGFP ESCs (Fig. 6a). The deleted fragment included the transcriptional start site of the Meteor IncRNA (Supplementary Fig. 6e), and resulted in a complete loss of Meteor IncRNA expression (Fig. 6b). Importantly, Meteor knockout (KO) cells exhibited normal expression of pluripotency markers such as NANOG/SSEA-1, and high alkaline phosphatase activity, demonstrating that these cells maintain stemness properties comparable to wild-type (WT) ESCs (Supplementary Fig. 6f–h). Meteor KO cells were next analyzed following induction of differentiation. On day 3, approximately fifty percent of wild-type ESC-derived cells expressed EGFP as assessed by flow cytometry (Fig. 6c). In sharp contrast, no EGFP expression was detected in differentiating Meteor KO cells (Fig. 6c, d). Furthermore, cell surface analysis of PDGFRα expression, a marker of early mesodermal cells, demonstrated the lack of mesodermal specification during differentiation of Meteor KO ESCs (Supplementary Fig. 6i). Consistent with the observed phenotype, differentiating Meteor KO ESCs exhibited a major transcriptional defect in the induction of mesendodermal PCGs (Eomes, T, Mix1, Gsc, Foxa2, Lhx1), and consequently of cardiac mesoderm (Mesp1), cardiac precursor (Gata4, Nkx2-5, Mef2c) and cardiomyocyte (Myh7) PCGs (Fig. 6e; Supplementary Fig. 6i).

To gain deeper molecular insights into the mechanisms governing this lineage-specifying defect, we performed RNA-Seq on WT and Meteor KO ESCs at day 0 (undifferentiated state) and day 3 of differentiation (Supplementary Data 6). Compared to controls, 1267 PCGs were differentially expressed in Meteor KO cells of which 871 were downregulated and 396 upregulated (Supplementary Fig. 7a). Furthermore, at day 3, 1667 PCGs were modulated of which 1101 were downregulated and 566 upregulated (Supplementary Fig. 7b). Core pluripotency factors were not impacted in Meteor KO ESCs (Fig. 6f). However, key regulators of primitive streak and ME were strikingly downregulated in differentiating Meteor KO cells. As a result, markers linked to cardiac mesoderm were also downregulated (Fig. 6f; Supplementary Fig. 7c). Consistent with this observation, GO analysis of PCGs downregulated in Meteor-deleted cells at day 3 of differentiation revealed terms linked to gastrulation and mesoderm formation, supporting that these biological processes are primarily affected in Meteor-deficient cells (Supplementary Fig. 7b; Supplementary Data 7). Importantly, Lhx1 and Nkx6-3 were respectively the most downregulated and upregulated PCGs in Meteor-KO ESCs at day 0 (Fig. 6g). These data indicate Meteor deletion during pluripotency is associated with a global transcriptional reprogramming favoring an ectodermal fate while blocking subsequent mesodermal specification. Accordingly, Mesp1 was the most downregulated PCG at day 3 while Neurog3 was the most upregulated (Supplementary Fig. 7d). As a consequence, Meteor KO cells were unable to initiate their cardiogenic program and did not generate any beating EBs (Supplementary Fig. 7e). Considering the topological nature of the Meteor locus, PCGs encompassed both within and proximal to the TAD harboring Meteor may be impacted by Meteor deletion in a cis-dependent manner. We examined the expression of PCGs within a 4 Mb region centered on the Meteor locus in wild-type and Meteor-KO undifferentiated ESCs. Only two of the twenty-six PCGs within this region were significantly downregulated in deleted cells, specifically Snc5a and Mobp (Supplementary Fig. 7f). Globally, Meteor deletion was not associated with a large scale transcriptional dysregulation of local gene expression. In particular, Eomes was not modulated in Meteor-KO ESCs.

Increased neurogenic differentiation in Meteor-deleted ESCs. Interestingly, regulatory factors linked to early ectoderm and neuroectoderm were upregulated both at day 0 and day 3 in Meteor-KO cells. In particular, Nkx6-3 was the most upregulated gene in undifferentiated KO cells, and Neurog3 was the most upregulated PCG at day 3 of differentiation (Fig. 6g; Supplementary Fig. 7d). Early ectoderm markers, i.e. Otx2, Pax6, Gbx2, were also significantly induced upon differentiation in Meteor-deleted cells, suggesting that, while losing their ability to be specified towards ME, these cells were redirected towards the ectoderm lineage (Supplementary Fig. 7g). To investigate whether Meteor-KO cells could harbor greater propensity for differentiating into ectoderm-derived tissues, WT and Meteor-KO cells were differentiated using a neurogenic differentiation protocol. Consistent with the induced ectodermal gene program, Meteor-KO cells were able to produce increased numbers of TUBB3 positive neurons (Fig. 6h). Enhanced expression of Nestin and Nefin, two other markers of neuronal differentiation, demonstrated the increased production of mature neurons. The transcriptional reprogramming leading to a blockade of ME specification and activation of a neuroectodermal gene network in Meteor-deleted ESCs suggests that Meteor may enact a more global Trans-regulatory role, either directly via the IncRNA produced from this locus or indirectly by regulating the expression of nearby genes. We therefore investigated whether reprogramming was associated with modifications on chromatin at specific PCG promoters and distal enhancers (Fig. 6i). ChIP-qPCR was performed against H3K4me3 and H3K27Ac using primers targeting promoters and enhancers of PCGs exhibiting the greatest fold changes between WT and Meteor KO cells. For example, for both Nkx6-3 and Neurog3, the promoters and the closest distal enhancers were significantly enriched in H3K4me3 and H3K27Ac respectively in Meteor-deficient cells. Conversely, both the promoters and enhancers of the ME TFs, Eomes and Lhx1, were significantly depleted of these respective marks in Meteor KO cells. We also assessed the chromatin state at promoters and enhancers of other relevant ectoderm and ME PCGs.

| Table 1 Pluripotency enhancer-associated IncRNA loci distal to ME TFs |
| Mesendoderm PCGs | Enhancer expressing an associated IncRNA exclusively in ESCs |
|-------------------|------------------------------------------------------------|
| Tbx3              | NO                                                         |
| Gsc               | YES                                                       |
| Eomes             | YES                                                       |
| T                 | NO                                                        |
| Mix1              | NO                                                        |
| Foxa2             | NO                                                        |
| Sex17             | YES                                                       |
| Bmp4              | NO                                                        |
| Kdr               | NO                                                        |
| Cax4              | NO                                                        |
| Pdgfr-a           | NO                                                        |
| Evx1              | NO                                                        |
| Lhx1              | NO                                                        |
and confirmed that significant epigenetic remodeling occurred in the absence of the *Meteor* locus (Supplementary Fig. 7h). Finally, we repeated and validated these findings using an alternative, independently generated, *Meteor*-deleted clone (Supplementary Fig. 8a–e). These results support therefore an important role for *Meteor* during pluripotency in controlling the

---

**Fig. 5** Pluripotency specific enhancer associated IncRNA loci interact with mesendodermal transcription factors. a RNAseq and H3K27Ac reads in ESCs and Eo⁺ for *Eomes* and the IncRNA *XLOC_073715*. b RNAseq and H3K27Ac reads in ESCs and Eo⁺ for *Sox17* and the IncRNA *XLOC_000012*. c RNAseq and H3K27Ac reads in ESCs and Eo⁺ for *Gsc* and the IncRNA *XLOC_015227*. d Heatmap showing expression of three IncRNAs candidates (*XLOC_073515*, *XLOC_000012*, and *XLOC_015227*) and their cognate PCGs (*Eomes*, *Sox17*, and *Gsc*) in ESCs and Eo⁺. e Hi–C interaction density heatmap of a genomic region where *Eomes* and IncRNA *XLOC_073715* are centered. f Hi–C interaction density heatmap of a genomic region where *Sox17* and IncRNA *XLOC_000012* are centered. g Hi–C interaction density heatmap of a genomic region where *Gsc* and IncRNA *XLOC_015227* are centered.
developmental competence to commit into ME and neuroectoderm.

**RNA-independent Meteor function in pluripotency.** Genetic deletion of the Meteor enhancer and its associated lncRNA led to a global transcriptional reprogramming that abrogated the developmental competence of pluripotent ESCs for subsequent ME specification. The functionality of the enhancer can therefore depend on mechanisms that involve the DNA element (e.g., recruiting TFs), the RNA transcript itself and/or processes associated with its production, including the process of transcription.
and splicing of the transcript. As a first step, we utilized a CRISPR-based gain-of-function approach (CRISPR-On) to boost transcription at the locus and increase Meteor IncRNA expression. Embryonic stem cell-like P19CL6 cells were transfected with components of the synergistic activator mediator (SAM)33, in combination with a guide-RNA targeting sequences upstream the transcription start site (TSS) of the Meteor IncRNA. Induced Meteor IncRNA expression was associated with increased expression of key putative downstream ME target genes including Eomes, T, Gsc, in addition to cardiogenic regulators including Gata4 and Isil (Fig. 7a). This result formally demonstrates that the activation of a pluripotent-specific enhancer such as Meteor is sufficient to stimulate downstream pathways controlling cardiogenic differentiation. Of note, early ectoderm and neuroectodermal markers were not affected following Meteor activation (Supplementary Fig. 8f). To determine whether Meteor enhancer activation was mediated by the Meteor IncRNA or was purely a consequence of increased transcription at this locus, we utilized both siRNA and modified antisense oligonucleotides (ASOs) to deplete Meteor IncRNA in ESCs (Fig. 7b, c). Both approaches were able to decrease Meteor IncRNA levels by ~50%. However, transcript depletion was not associated with perturbed expression of target ME or neuroectodermal genes as observed in Meteor KO ESCs (Fig. 7b, c; Supplementary Fig. 8g, h). These data support an RNA-independent mechanism in transcriptional programming of developmental competence occurring in pluripotency.

To confirm these observations, we utilized genetically modified ESC lines engineered to incorporate an early polyadenylation signal (pAS) 1.1 kb downstream of the Meteor IncRNA TSS (previously described as linc1405)21. Genetic modifications were performed in 129/Castaneus (Cas) F1 hybrid mouse ESCs that contain a polymorphic site every 140 bp, enabling to distinguish allele-specific expression. Importantly, ~90% of Meteor IncRNA was transcribed from the 129 allele (Fig. 7d). Insertion of a pAS in the 129 allele (pAS/Cas), therefore, completely abrogated the expression of the Meteor IncRNA in engineered ESCs while keeping the Meteor enhancer intact (Fig. 7e). Depletion of Meteor IncRNA in pluripotent ESCs had no impact on expression of target genes. Altogether, these data demonstrate that the effects of the Meteor locus on ME specification in pluripotent cells is RNA-independent and thus likely secondary to enhancer activity.

Probing Meteor functionality during differentiation. We next investigated the importance of Meteor IncRNA transcription during ESC differentiation into cardiomyocytes. Upon differentiation, the pAS-modified hybrid cells (pAS/Cas ESCs) lacking Meteor IncRNA expression demonstrated a defect in the activation of the ME and cardiogenic gene programs (Fig. 7f). Specifically, the absence of transcript production from the Meteor locus abolished expression of key mesendodermal genes such as Eomes, T and Lhx1, and of cardiac mesoderm and cardiac genes, for instance Mesp1, Myh6, and Myh7, resulting in reduced number of beating EBs (Supplementary Fig. 8i). Interestingly, the dynamic downregulation of pluripotency genes were not affected (Supplementary Fig. 8j). Therefore, despite the RNA being dispensable for the programming of ME competence during pluripotency, these data highlighted a dependence on IncRNA transcription during the process of cardiogenic differentiation.

Considering the importance of the Meteor locus, we aimed at identifying whether the Meteor enhancer was associated with the production of an IncRNA in human (h) ESCs. Publically available RNA and ChIP Seq data suggested transcription at the locus was indeed conserved. As a consequence, the associated transcript (METEOR IncRNA) was determined to be highly expressed in hESCs (Fig. 7g). Using an efficient differentiation protocol for ME specification and cardiac differentiation (Fig. 7h; Supplementary Fig. 8k)34, we next demonstrated that the human METEOR IncRNA was dynamically modulated in an evolutionarily conserved manner during cardiogenic differentiation (Fig. 7i). In particular, the METEOR IncRNA was downregulated at the mesendodermal stage, coinciding with the upregulation of key ME markers including EOMES, T and downstream cardiogenic markers such as ISL1 and TMEM88.

Discussion

The enhancer landscape is pivotal in establishing the developmental competence of pluripotent ESCs, in particular during the response to inductive signals for ME specification12,13. Enhancers are also emerging as key determinants governing both cell fate and identity during the development of the cardiovascular system14. In this context, the small subset of enhancers associated with the production of multieXonic IncRNAs is of particular importance during commitment to specific fates14,15. We therefore disregarded bidirectional eRNAs in the present study. Indeed, enhancers producing multieXonic and polyadenylated IncRNAs exhibit greater chromatin accessibility and are associated with increased binding of key developmental TFs14. They are also more likely to elicit formation of promoter-enhancer loops and show greater enhancing activity on target PCGs than non-transcribed enhancers15. IncRNA loci are unique in their ability to spatially amplify regulatory information encoded by their underlying DNA16. Based on these observations, we characterized the activity of enhancers, profiled the IncRNA transcriptome during ME formation, and finally identified the Meteor locus encompassing an enhancer element lying upstream the Eomes gene. The Meteor enhancer appears crucial to control cell fate determination between the ME and the neuroectoderm, and is essential for cardiac differentiation. An important finding is the...
fact that the mesendodermal competence of ESCs, for which Meteor seems indispensable, is hardwired in pluripotency. These findings are summarized visually in Supplementary Fig. 9.

ChIP-Seq analysis allowed us to map the TE and SE landscapes during mesendodermal commitment. Both TEs and SEs undergo significant restriction during cell fate determination and differentiation. Furthermore, a concomitant activation of a small number of lineage-specific enhancers is observed. This is compatible with a gradual narrowing of the spectrum of cell-fate competence32. Importantly, ME-specific TEs were linked to PCGs associated with relevant biological processes such as gastrulation and mesoderm formation. Only six SEs were found uniquely activated in mesendodermal cells. Although this number reflects in part the smaller number of SEs as compared to TEs, this finding is consistent with SEs representing a limited but crucial class of Cis-regulatory elements that defines cell identity10. Along
this vein, global SE activation was associated with significantly greater changes in target PCG expression during differentiation, supporting an important role in controlling dynamics of gene expression. Transcriptomic profiling revealed that the enhancer landscape was associated with the transcription of thousands of multiexonic lncRNAs. This large number of non-annotated transcripts is in part a result of the extreme depth of sequencing utilized in this study. Nevertheless, non-annotated lncRNAs demonstrated interesting characteristics when compared to previously annotated transcripts. In particular, they exhibited more restricted and specialized expression patterns. This is likely a consequence of their enriched association with stage-specific TFs and SEs. Importantly, both TE and SE-associated lncRNAs exhibited significant lineage restriction of expression during differentiation, while pheRNAs exhibited less restricted profiles. SE-associated lncRNAs displayed the greatest restriction, suggesting again such transcripts may mediate important roles during specification and differentiation. Several lncRNAs associated with SEs, including Carmen, CCAT-2, Meteor, and Wisper, have recently emerged as important modulators of cell fate determination and maintenance of cell identity, notably in the heart. Globally, non-annotated lncRNAs expressed in mesendodermal precursors were shown to be enriched in the adult heart. The model used to induce ME specification, which favors cardiac differentiation and not definitive endoderm, may partially explain this characteristic. Nevertheless, a significant fraction of the identified lncRNAs may mediate specialized functions in cardiac homeostasis. We also observed a significant difference in the evolutionary characteristics of non-annotated lncRNAs expressed in ME-speciﬁed cells. Conservation at promoter regions of those lncRNA loci was greater than that of lncRNAs expressed in other cellular intermediates during cardiogenic differentiation, and greater than that measured in cells not committed to ME. ME formation is a key evolutionary branch point that is represented by Eos cells in our experiments. This point has been termed the phytotopic stage and has given rise to the hourglass model of development. Interestingly, our evolutionary analyses are in accord with this model and indicate that the hourglass phenomenon may be associated with selection of discrete sets of enhancer-associated lncRNA promoters.

Developmental competence is typically mediated by chromatin states at lineage specifying enhancer and promoter loci. In this context, transcribed enhancers are emerging as key elements for modulating chromatin architecture, in particular as regulators of TAD formation. TADs are typically established in ESCs and critical for developmental competence and subsequent germ layer specification. We identiﬁed three ESC-speciﬁc enhancer-associated lncRNA loci highly interacting in three-dimensional space with key ME-specifying TFs, and examined the functional requirements of one speciﬁc locus, Meteor, containing a transcribed enhancer highly interacting with the Eomes promoter. Deleting this single genomic locus in ESCs led to the complete abolition of Eomes expression, ME speciﬁcation and subsequent differentiation into the cardiomyocyte fate. Eomes has been previously deleted in ESCs and resulted in perturbed ME speciﬁcation. However, Eomes-deleted cells do not exhibit absolute lack of ME formation. Some ME-specifying TFs such as Gsc and T are not impacted by Eomes deﬁciency while Meteor deletion completely disrupts the mesendodermal gene program. This suggests Meteor occupies an upstream position within the GRN regulating ME speciﬁcation. Importantly, Meteor-deleted cells appear to maintain core stemness features suggesting this locus is not required for self-renewal. Meteor lncRNA is transcribed from its associated enhancer in a convergent orientation to another enhancer that is activated during ME speciﬁcation. The importance of this second distal enhancer for Eomes expression should be examined in future studies. Interestingly, different proximal enhancers dictating the induction of Eomes in the anterior visceral endoderm, primitive streak and definitive endoderm have been recently investigated. A poised preformed chromatin architecture at the Eomes locus appears permissive for rapid transcriptional induction in response to nodal signaling during gastrulation via SMAD2/3 binding at these enhancers. The role of Meteor and its associated lncRNA in modulating the activity of these proximal enhancers remains to be demonstrated. Nevertheless, based on these results, it is tempting to speculate that this preformed chromatin compartment may be established by Meteor during pluripotency. Finally, Meteor-deleted ESCs appear to lose their competence for ME speciﬁcation while maintaining their capacity to produce neuroectodermal lineages. This was already evident at the ESC stage suggesting that Meteor is able to epigenetically prime pluripotent cells at key lineage-determining loci prior to commitment into the three germ layers. As a consequence, Meteor KO cells give rise to increased number of neurons upon induction of neurogenic differentiation. This is in accordance with previously published studies identifying molecular mechanisms controlling a binary cell fate decision between the mesoderm and the neuroectoderm.

Mechanistically, the Meteor locus encodes both an enhancer and a lncRNA. Transcription at the locus and/or the RNA itself may be involved in remodeling the local chromatin topology, thereby priming the epigenome for developmental signals. However care needs to be taken when interpreting the roles of the enhancer and of its associated lncRNA. A number of transcribed enhancers have been shown to mediate multifunctional roles. For instance, the Haunt, Hand2 and Nanog enhancers have been shown to encode both RNA-dependent and RNA-dependent functions.
RNA extraction, RT-PCR and real-time PCR analysis. Total RNA from cultured cells was extracted using miRNeasy kit (Qiagen) according to the manufacturer’s instructions and quantified with Nanodrop (Thermo scientific). The quality control was performed with bioanalyzer Agilent 2100 (Agilent Technologies). Two steps cDNA synthesis was performed with SuperScript II (Invitrogen), and quantification was carried out using Quantstudio™ 6F (ThermoFisher). Gene expression was normalized to Gapdh and quantified using the ΔΔCt method. Primers or probes used in the manuscript are described in Supplementary Table 1.

RNA sequencing and analysis in ESC, Eo− and Eo+. Total RNA was isolated using the RNeasy kit (Qiagen). Sequencing libraries were prepared according to Illumina RNA Seq library kit instructions with Poly(A) selection. Libraries were sequenced on the Illumina HiSeq2000 (2 × 100 bp) using 2 lanes/sample with a multiplex level of 1 (~5 × 10⁹ reads per sample) for a total of twelve samples from four different differentiation sets: four ESC, four Eo− and four Eo+. 100 nt paired-end reads were mapped to mm10 reference genome using STAR software version 2.4.0_g, using Ensembl GRCh38.77 reference genes GTF. An ab initio gene prediction reconstruction was performed using Cufflinks, version 2.2.1. As the RNASeq data is stranded, parameter library-type was set to ‘fr-firststrand’. The other parameters were default. The resulting GTFs were merged using Cuffmerge, version 2.2.1.44, using option ‘–g’ with Ensembl GRCh38.77 GTF as reference, allowing distinguishing known and non-annotated transcripts. Read counts were calculated per gene from the alignment bam files using HTSeq (0.6.1) with options ‘-m union’–stranded reverse. Genes were then filtered for minimal expression (at least one condition with average > 0.1 FPKM).

Classification of lncRNA. Using the GTF output of Cuffmerge, the transcripts were classified into 3 categories: known mRNAs, known lncRNAs (using Ensembl as reference) and non-annotated lncRNAs. Non-annotated transcripts were filtered for minimal length of 200 bp and at least 2 exons. lncRNA genes were classified into several categories by comparing the lncRNA exon and gene coordinates with coordinates of known protein coding genes.

LncRNA analysis. Coding potential: The protein-coding potential of transcripts was evaluated using the program Gene3D, version v1.4.4, applied to transcript sequences in FASTA format, with 8 parameters adapted for vertebrates as provided by the authors in file Gene3D.107.0123_preferences.sgd-G. Expression heatmaps and genome ontology analysis: Unsupervised clustering of PCGs, Ensembl lncRNAs and non-annotated lncRNAs was generated using the Euclidian distance between the FPKM values of the genes. The PCG, Ensembl lncRNA and non-annotated lncRNA expression heatmaps were generated by clustering the genes by Pearson correlation of the FPKMs, and clustering using the hclust function (method = “complete”). Heatmap.2 was used to generate the heatmaps. Values were scaled by row. The clusters were manually rearranged.

Differential expression analysis of lncRNAs: Differentially expressed genes were detected using the limma package in R by transforming the raw count data to log2-tpm (counts per million reads) using the voom function. Empirical Bayes moderated t statistics and corresponding p-values were then computed for the 3 comparisons: Eo−/ESC; Eo+ / ESC and Eo+/Eo−. P values were adjusted for multiple comparisons using the Benjamini Hochberg procedure.25 Genes with an adjusted p value of < 0.05 were considered to be differentially expressed.

Transcript cell specificity: The specificity of PCGs and lncRNAs was assessed by quantifying the Normalized Difference (ND) of expression in the three conditions (ESC, Eo−, Eo+). The Normalized Difference of a PCG or lncRNA x was quantified as the maximum difference between its expression gx (FPKM normalized) in the three conditions, divided by its average expression (as normalizing factor). Formally:

$$ND(x) = \frac{\max(g_{x}) - \min(g_{x})}{\text{mean}(g_{x})}$$

The distributions of the Normalized Difference in the three classes (PCGs, Ensembl lncRNAs and non-annotated lncRNAs) were formally compared to each other using the one-tail Wilcoxon rank sum test. Furthermore, the Normalized Difference was used to measure the specificity of the three classes of lncRNAs (plncRNA, TE lncRNA and SE lncRNA). The density plots of the distribution of

Methods

Culture and differentiation of mouse ES cells. Mouse embryonic EomesEGFP reporter stem cell line was a kind gift of Elizabeth Robertson (University of Oxford, UK). Cells were cultured on mouse embryonic fibroblast feeders in standard ES cell medium, which consisted of DMEM high-glucose (Life Technologies #11965-092), 20% Fetal Bovine Serum (FBS) (Invitrogen #16140-079), 1% Non-Essential Amino Acids (Life Technologies 11140-035), 0.1 mM β-mercaptoethanol (Life Technologies #13350-010), Gentamicin (1:500, Gibco #15730-037) and 1000U ml⁻¹ of Leukemia Inhibitor Factor (ESGRO #ESG1107). Cardiac differentiation of ES cells was induced by aggregating aLk quots containing ~ 1000 cells in hanging drop of 25 µl medium supplemented with 200mM L-glutamine (Life Technologies #21980-032) medium. Cardiac differentiation was carried out using Quantstudio™ 6F (ThermoFisher). Gene expression was normalized to Gapdh and quantified using the ΔΔCt method. Primers or probes used in the manuscript are described in Supplementary Table 1.

Immunofluorescence analysis. Cells and/or EBs were fixed for 10 min in 4% paraformaldehyde in PBS and permeabilized with 0.1% Triton X-100 in PBS (Sigma). After blocking in blocking buffer (PBS containing 0.1% Triton X-100 and 1% BSA) they were incubated overnight at 4 °C with the following primary antibodies: chicken anti-GFP antibody (1:1000, Abcam #AB19790), rabbit anti-

Alkaline phosphatase staining. Wild type ESCs and Meteor KO at day 0 were stained for the expression of Alkaline Phosphatase with the Alkaline Phosphatase detection kit (Milipore, #CR3004). An Axioplan microscope (Carl Zeiss) was used for this analysis.

Flow cytometry. Mouse ES cells and EBs were dissociated using FACS medium and filtered through a 40-μm cell strainer. Live cells were gated on the basis of side scatter, forward scatter and propidium iodide exclusion. Undifferentiated ES cells were gated for the GFP channel to exclude any possible background of GFP signal. ESC and day 3 cells were analyzed for GFP expression (Beckman Coulter Life Sciences). Cells obtained from the dissociation of the day3 embryoid bodies were sorted for GFP. A total of 12 × 10⁶ of GFP negative (Eo−) and GFP positive (Eo+) cells were sorted (BD FACSAria IIu, BD Biosciences) for performing RNA isolation and Chromatin Immunoprecipitation (ChIP) assay.

Downloaded day 0 Meteor KO were analyzed for SSEA-1 (1:100, Abcam #560142) using the Gallios analyser. Day0 WT, day0 Meteor KO, day3 WT and day3 Meteor KO were analyzed for PDGFB expression (1:10, Milenyi Biotec #130-102-473) using the Gallios analyser. Flow cytometry analysis was performed on day15 human iPSC-derived cardiomyocytes. Cells were stained using cardiac troponin T antibody (0.2 mg ml⁻¹, Thermo Fisher, MA5-12960) or the corresponding isotype control (0.5 mg ml⁻¹, Thermo Fisher, #14-3714-82). Cells were analyzed using a BD FACSCANTO II (Beckton Dickinson), with FACSDiva software (BD Biosciences). Data analysis for all flow cytometry analyzes was performed using Flowjo (Tree Star).

RNA sequencing and analysis in ESC, Eo− and Eo+. Total RNA was isolated using the RNeasy kit (Qiagen). Sequencing libraries were prepared according to Illumina RNA Seq library kit instructions with Poly(A) selection. Libraries were sequenced on the Illumina HiSeq2000 (2 × 100 bp) using 2 lanes/sample with a multiplex level of 1 (~5 × 10⁹ reads per sample) for a total of twelve samples from four different differentiation sets: four ESC, four Eo− and four Eo+. 100 nt paired-end reads were mapped to mm10 reference genome using STAR software version 2.4.0_g, using Ensembl GRCh38.77 reference genes GTF. An ab initio gene prediction reconstruction was performed using Cufflinks, version 2.2.1. As the RNASeq data is stranded, parameter library-type was set to ‘fr-firststrand’. The other parameters were default. The resulting GTFs were merged using Cuffmerge, version 2.2.1.44, using option ‘–g’ with Ensembl GRCh38.77 GTF as reference, allowing distinguishing known and non-annotated transcripts. Read counts were calculated for each gene from the alignment bam files using HTSeq (0.6.1) with options ‘-m union’–stranded reverse. Genes were then filtered for minimal expression (at least one condition with average > 0.1 FPKM).

Classification of lncRNA. Using the GTF output of Cuffmerge, the transcripts were classified into 3 categories: known mRNAs, known lncRNAs (using Ensembl as reference) and non-annotated lncRNAs. Non-annotated transcripts were filtered for minimal length of 200 bp and at least 2 exons. lncRNA genes were classified into several categories by comparing the lncRNA exon and gene coordinates with coordinates of known protein coding genes.

LncRNA analysis. Coding potential: The protein-coding potential of transcripts was evaluated using the program Gene3D, version v1.4.4, applied to transcript sequences in FASTA format, with 8 parameters adapted for vertebrates as provided by the authors in file Gene3D.107.0123_preferences.sgd-G. Expression heatmaps and genome ontology analysis: Unsupervised clustering of PCGs, Ensembl lncRNAs and non-annotated lncRNAs was generated using the Euclidian distance between the FPKM values of the genes. The PCG, Ensembl lncRNA and non-annotated lncRNA expression heatmaps were generated by clustering the genes by Pearson correlation of the FPKMs, and clustering using the hclust function (method = “complete”). Heatmap.2 was used to generate the heatmaps. Values were scaled by row. The clusters were manually rearranged.

Differential expression analysis of lncRNAs: Differentially expressed genes were detected using the limma package in R by transforming the raw count data to log2-tpm (counts per million reads) using the voom function. Empirical Bayes moderated t statistics and corresponding p-values were then computed for the 3 comparisons: Eo−/ESC; Eo+ / ESC and Eo+/Eo−. P values were adjusted for multiple comparisons using the Benjamini Hochberg procedure. Genes with an adjusted p value of < 0.05 were considered to be differentially expressed.

Transcript cell specificity: The specificity of PCGs and lncRNAs was assessed by quantifying the Normalized Difference (ND) of expression in the three conditions (ESC, Eo−, Eo+). The Normalized Difference of a PCG or lncRNA x was quantified as the maximum difference between its expression gx (FPKM normalized) in the three conditions, divided by its average expression (as normalizing factor). Formally:
normalized difference were generated using the Gaussian kernel density estimator implemented in the R package.

**Analysis of gene conservation.** PhastCons: The scores calculated on a multiple alignments of 60 vertebrate genomes to the mm10 mouse genome by chromosome were downloaded from the UCSC website. For each gene, scores per base for exons, introns and promoters (defined as 1000 bp upstream from TSS) were summed. The divided fragment length was used as the numerator score per fragment. In total, 50,000 random intergenic regions were generated (size = 3400 bp ± 20%). The same score was calculated. Log10 of the scores was plotted by category using R package lattice. The scores of the intergenic regions were calculated as a comparison.

PhyloCSF: The scores calculated on a multiple alignments of 60 vertebrate genomes to the mm10 mouse genome by chromosome were downloaded from the UCSC website. The details of the 60 vertebrate genomes can be visualized at the following link: http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phastCons60way/. For each transcript, the maximum per base phyloCSF score was taken over 600 bp (500 bp upstream from the TSS, 100 downstream). The maximum value by transcript was used as value for the gene. In Supplementary Fig. 3d enriched transcripts for ESCs, Eo⁺ and Eo⁻ are defined from the hierarchical clustering of expression in Supplementary Fig. 4b and Fig. 1e, f. Cardiac Precursor Cells (CPC) enriched transcripts are transcripts significantly overexpressed (adj.p < 0.05; fold change > 2) day0 differentiating ESCs vs. day0 undifferentiated ESCs in our previous study. Adult heart enriched transcripts are transcripts significantly expressed (FPKM > 0.5) in the mouse adult heart in our previous study.

**Gene ontology analysis.** Gene Ontology (GO) analysis was performed using GREAT (Genomic Regions Enrichment of Annotations Tool) to analyze biological processes ontology terms. The whole genome was used as background.

**Gene expression across tissues.** Expression of the genes (PCGs, Ensembl IncRNA and non-annotated IncRNAs) in 12 mouse tissues (Thymus, Liver, Stomach, Colon, Ovary, Spleen, Heart, Kidney, Mammary gland, Frontal lobe, Cortex, Cerebellum) was measured on ENCODE public data (CHSL Long RNA-seq, PI Gingeras, Lab CSHL-m) (The ENCODE Project Consortium, 2011). Counts on plus and minus strands were summed and mean counts were taken for the two replicates per tissue. Between sample normalization was performed using DESeq estimateSizeFactors function. Only genes with minimal expression were kept (at least 0.1 FPKM in one of the conditions).

Heart Enrichment score (per gene) was defined as:

\[ HS = \frac{\mu_{\text{cardiac}}}{\sigma_{\text{non-cardiac}}} \]

Where \( \mu \) cardiac is the average expression per gene in ENCODE adult heart tissues, \( \mu \) non-cardiac is the average expression per gene in the 11 other ENCODE samples, and \( \sigma \) is the standard deviation per gene in non-cardiac ENCODE samples. A gene was considered tissue enriched with enrichment score > 1. The clustering was performed using hclust, version 1.3.1.1, using Spearman correlation and euclidean distance, average linkage clustering. A scaling by row was applied. The specificity side bars (in red) were generated using the Heart Enrichment score and defined above. Heatmaps were generated using heatmap.2 from the package gplots in R, version 2.17.0. The analysis was performed in parallel on subsets of genes, de novo and pre-existing. The whole mouse genome was used as background.

**Epigenomic annotation of IncRNAs.** The Ensembl and non-annotated IncRNAs were classified as: (1) Promoter associated IncRNA if the region ± 1 kb around the TSS was marked positive for H3K4me3 (using the Poissonian model and ratio over input as described above). (2) Super-enhancer associated IncRNA if the region ± 1 kb around the TSS was not marked as H3K4me3 positive and if any part of the transcript was marked as Super Enhancer. (3) Typical-enhancer associated IncRNA: same as above, but with any part of the transcript was marked as Typical Enhancer. (4) Unmarked IncRNA if no chromatin mark was present on the gene.

**Hi-C data.** Hi-C plots were generated using the 3D Genome Browser (http://promoter.bx.psu.edu/hi-c/index.html) using publicly available Hi-C data on mouse ES/SCs.

**Chromatin immunoprecipitation (ChIP) assay.** Cells were cross-linked with 1% formaldehyde for 10 min at RT. Cross-linking reaction was stopped by addition of 0.125 M glycine for 5 min. Chromatin extracted from 1 x 10⁷ cross-linked cells was sonicated to an average of 200–700 bp with the Covaris sonicator (S220 Focused-ultrasonicator). Chromatin Immunoprecipitation (ChIP) was carried out as follows: magnetic dynabeads (Dynabeads Protein G, Novex ref #410004D) were pre-coated with the specific ChIP antibody (Ab) for 4h in a rotating platform at 4°C. The following Abs were used: H3K27Ac (2 μg of Ab with 60 μL of magnetic beads in 1 mL of total volume, Abcam #4729) and H3K4me3 (4 μg of Ab with 60 μL of magnetic beads in 1 mL of total volume, Abcam #8580). The fragmented chromatin and the coated beads were incubated ON at 4°C under 10 RPM rotation. After the IP the samples were washed and treated with Proteinase K (500 μg/ml) final and RNaseA (20 μg/ml) final and finally purified with the MinElute PCR Purification Kit (QUIAGEN, cat #28006). Immunoprecipitated DNA was quantified by Quibit (Life Technologies) and subject to qPCR (primers in the manuscript are described in Supplementary Table 2) or high-throughput sequencing analysis. For ChIP-seq experiments, libraries for sequencing were prepared with MicroPlex v2 kit followed by sequencing using 10 ng of Chromatin with input chromatin and immunoprecipitated DNA. Libraries were subjected to 50-bp single-end read analysis on an Illumina HiSeq 2500.

**ChIP-sequencing analysis.** Calling of Constituent Enhancer (CE) regions: 50 bp paired-end short reads were aligned to the mm10 mouse genome using Bowtie (option – non-deterministic was used, the rest was default). Duplicated reads were removed using picard-tools (V. 1.80), MarkDuplicates function. Sequences were extended to 200 bp and allocated in 25-bp bins. Counts per bin were generated using a custom script. Biological replicate whole cell extracts were sequenced for each time point and combined by time point. A Poissonian model was used to determine statistically enriched bins with a p-value threshold set at 1 x 10⁻⁵. In addition, we required that genomic bins were at least 0.5-fold over input to be considered enriched peaks. To obtain a score by condition, the bins were marked as positive if both replicates of the condition were positive (intersection). The final list of constituent enhancers was defined as the union of contiguous marked bins from the three conditions (so CE existing in at least 1 condition), provided the size was at least 200 bp long. Calling of Super Enhancer (SE) and Typical Enhancer (TE) regions: The ROSE software (Version 0.1, April 2013, http://bitbucket.org/young_computation/rose) was then run on each replicate of each condition, using the CE defined by the ROSE software. The ROSE flagged CE as super enhancer if each replicate of the two replicates was taken in each condition, defining a list of SE per condition. The plots shown in Fig. 3a have been generated by the ROSE software. They show 1 replicate for each of the conditions. Then the union of the SE coordinates in the 3 conditions was taken to define a global list of Super Enhancers. The Typical Enhancers were defined as the Constituent Enhancers not overlapping with the union of Super Enhancers in the 3 conditions.

The ChIPseq scores per SE and TE were then calculated using the same custom script. The heatmaps of active and inactive H3K27Ac regions (Fig. 3d, e, f, TE and SE) were generated by taking the overlap (of any coordinate-wise, of the global list of enhancers (resp. TE and SE) with the list of each individual condition. If a particular enhancer of one of the conditions overlapped with the union of the enhancers, it was marked as active in this condition. The clusters of the heatmaps were then sorted to have the active in 3 conditions on the top, and the active only in Eo⁺ on the bottom. The bigWig tracks used in the UCSC genome browser were generated using the Useq software (V. 8.9.3), Sam2Useq function. It generated per base read depth visualization. The values have been scaled per million mapped reads. The useq files have first been converted to wig, then to bigwig, using UseqText and UCSC’s software, wigToBigWig.

**Genetic deletion with CRISPR-Cas9 nickelase.** gRNA design and production of CRISPR-Cas9 nickelase constructs: Guide RNAs (gRNAs) and CRISPR Cas9-D10A nickase were encoded in a modified expression plasmid, pX335 (pX335_G2P), kindly offered by Phillip Groete (Goethe, Universitat Frankfurt). pX335-G2P plasmid was digested using BbsI restriction enzyme (Thermofisher, #ER1011) and gRNA sequences were designed using the web resource http://crispr.mit.edu (gRNAs used in the manuscript are described in Supplementary Table 3). Pairs of DSBs (2 oligonucleotides harboring variable 20 nucleotide sequences (G + 19 bases) were annealed and ligated into the modified pX335_G2P plasmid using the rapid DNA ligation kit (Roche, #11 635 379 001). Electroporation and selection: 300,000 ESCs were plated on gelatinized six-well tissue culture plate the night prior to transfection. The next day, cells were transfected by electroporation (gRNAs using VECTASHIF for TE, for 2000 regression points with gRNA-2000-01-019). After 4h of transfection, mESCs were trypanosed and plated at low density in 10 cm plates and clones grown for 6–7 days in ESC medium supplemented with 2 μg ml⁻¹ of puromycin (Sigma, #P8833-100MG). Individual ESC clones were picked, expanded and analyzed by PCR genotyping. Primers spanning different regions of the deleted region were designed (Supplementary Table 4).

**Sequencing of RNA isolated from WT and Meteor KO cells.** Total RNA was isolated from WT d0 (n = 3), Meteor KO d0 (n = 3), WT d3 (n = 3) and Meteor KO d3 (n = 3) using the miRNAeasy kit (Qiagen). Sequencing libraries were prepared
Neuronal differentiation of ESCs. Mouse embryonic EomesGFP reporter stem cells (WT and Meteor KO) were maintained and differentiated into embryonic stem cell derived neurons (ESNs) in a 37 °C tissue culture incubator at 5% CO2. Briefly, cells were maintained on coated tissue culture flasks (0.1% gelatine, Sigma) in ES medium consisting of DMEM with Glutamax, 15% fetal bovine serum, and 1% non-essential amino acids. Cells were subcultured (1:3 ratio) and filtered for low complexity with seq_crumbs (v. 0.1.8). Cells were adjusted against the Mus musculus GRCm38.82 genome using STAR60 (v. 2.4.2a). The number of read counts per gene locus was summarized with htsqCount61 (v. 0.6.1) using Mus musculus. GRCm38.82 gene annotation. Quality of the RNA-Seq data alignment was assessed using RSeQC62 (v. 2.3.7). Reads were also aligned to the Mus musculus GRCm38.82 transcriptome using STAR62 (v. 2.4.2a), and the estimation of the isoform abundance was computed using RSEM63 (v. 1.2.19). Statistical analysis was performed for genes and isoforms independently in R (version 3.2.1). Genes/isoforms with low coverage were filtered out according to the rule of thumb (at least 3 reads per million (rpm) in at least 1 sample). Library sizes were scaled using TMM normalization64 (EdgeR v 3.3.8.5) and log-transformed with voom function (R version 3.2.2.4). Statistical quality controls were performed through pairwise sample correlations, clustering and sample PCA. Replicates cluster together and are separated between different conditions. Differential expression was computed with limma55 by fitting data into a linear model, adding the factor for the batch effect and comparing GapmR vs. control conditions. The P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method56, controlling for false discovery rate (FDR) or adjusted P-value.

Cardiac differentiation of human induced pluripotent stem cells (iPSCs). hPSCs were plated on Vitronectin XF (Stemcell Technologies #07180) coated plates and maintained in an undifferentiated state in mTeSR1 media (Stemcell Technologies #85580)54. All experiments using human cells were carried out in accordance with human research ethics committee approval at The University of Queensland (Australia). Small molecule cardiac-directed differentiation using a monolayer platform was performed with a modified protocol based on previous reports56,67,68. The differentiation set up was initiated by plating undifferentiated hiPSCs as single cells for 24 h. Cells were induced to differentiate (designated day 0) with RPMI 1640 media (Invitrogen #11879093) containing 3 µM CHIR-99021 (Stemcell Technologies #72054), 213 µg ml−1 L-ascorbic acid 2-phosphate, and 500 µg ml−1 bovine serum albumin (all Sigma Aldrich). On day 5, media was changed to RPMI 1640 media with 1 µM XAV-939 (Stemcell Technologies #72674), L-ascorbic acid 2-phosphate, and bovine serum albumin (all Sigma Aldrich). On day 5, media was changed to RPMI 1640 media with L-ascorbic acid 2-phosphate and bovine serum albumin. From day 7 onward media was replaced with RPMI 1640 containing B27 supplement with insulin (Invitrogen #17504001). Total RNA was isolated using the RNeasy Miniprep kit (Qiagen). First-strand cDNA was synthesized using the Superscript III reverse transcriptase kit (Invitrogen). Quantitative RT-PCR was performed using SYBR Green PCR Master Mix (Invitrogen) on a ViiA7 Real-Time PCR System with 384-Well Block (Applied Biosystems). The gene expression for each transcript is relative to that of HPRT. Primers used for quantitative RT-PCR are listed in Supplementary Table 5.

Statistical analysis. GraphPad Software was used for statistical analysis. Data throughout the paper are expressed as mean ± SEM. Statistical significance between two columns was assessed by two-tailed unpaired Student’s t-test; for more than two columns, one-way ANOVA (Fisher’s LSD test) analysis was used. Two-way ANOVA (Fisher’s LSD test) was used to evaluate statistical significance between two or more groups. Correlation analysis was performed with Pearson (R or R² values; 95% confidence interval) or Spearman (R; 95% confidence interval). P values < 0.05 were considered significant in all events.

The authors declare that all data supporting the findings of this study are available within the article and its Supplementary Information files or from the corresponding author on reasonable request. The RNA-seq and CHIP-seq data reported in this paper have been deposited in NCBI GEO under the accession code: GSE103526 (ESC, Eo− and Eo+) and GSE103583 (WT vs. Meteor KO).

References
1. Kathiriya, I. S., Nora, E. P. & Bruneau, B. G. Investigating the transcriptional control of cardiovascular development. Circ. Res. 116, 700–714 (2015).
2. Bruneau, B. G. Signaling and transcriptional networks in heart development and regeneration. Cold Spring Harbor. Perspect. Biol. 8, a028923 (2015).
3. Kuyumcu-Martinez, M. N. & Bressan, M. C. Rebuilding a broken heart: lessons from developmental and disease. Biochim. Biophys. Acta 1837, 155–168 (2016).
4. Rao, J. et al. Stepwise clearance of repressive roadblocks drives cardiac development from mesoderm to heart tube. Dev. Cell 387, 387–398 (2016).
5. Wu, J., Yang, J., Jiang, J., Wang, J. & Zhang, Z. Cardiogenesis is inhibited by Nodal activation in mouse ESCs. Development 138, 2731–2740 (2011).
6. Wang, L. & Chen, Y. G. Signaling control of differentiation of embryonic stem cells. Developmental Suppl. 2, 215–225 (2014).
7. Wang, L. & Chen, Y. G. Signaling control of differentiation of embryonic stem cells. Developmental Suppl. 2, 215–225 (2014).
8. Costello, I. et al. The T-box transcription factor Eomesodermin acts upstream of Nodal to drive left ventricular outflow tract development. Development 142, 3795–3809 (2015).
9. Wamstad, J. A., Wang, X., Demuren, O. O. & Boyer, L. A. Distal enhancers: from mesoderm to heart tube. Development 142, 3795–3809 (2015).
10. Bagnat, C. et al. The T-box transcription factor Eomesodermin is essential for cardiac looping and thoracoabdominal outflow tract development. Dev. Biol. 387, 194–208 (2014).
11. Ounzain, S. & Pedrazzini, T. Super-enhancer Incs to cardiovascular development and disease. Biochim. Biophys. Acta 1863, 1953–1960 (2016).
12. Wang, L. & Chen, Y. G. Signaling control of differentiation of embryonic stem cells toward mesoderm. J. Mol. Biol. 428, 1409–1422 (2016).
13. Marín-Pérez, I., T. et al. The T-box transcription factor Eomesodermin drives the transcriptional network specific to cardiac mesoderm during mouse gastrulation. Nat. Cell Biol. 13, 1084–1091 (2011).
14. Wamstad, J. A., Wang, X., Demuren, O. O. & Boyer, L. A. Distal enhancers: new insights into heart development and disease. Trends Cell Biol. 24, 294–302 (2014).
15. Li, H., Jiang, J. & Zhao, Y. C. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).
16. Uzunov, S. & Pedrazzini, T. Super-enhancer Incs to cardiovascular development and disease. Biochim. Biophys. Acta 1863, 1953–1960 (2016).
Office for Civil Protection, Bern, Switzerland (S.L.L.) and the Pierre Mercier Foundation, Lausanne, Switzerland (S.O.).

Author contributions
M.A. designed and performed experiments, analyzed data and created the figures; C.-C.T., D.M. and I.P. contributed to ESC culture and differentiation, and performed immunochemistry on wild-type and Meteor KO cells; R.M. and M.N. helped generating the Meteor KO ESCs; D.V. and N.H. carried out library preparation for ChIP-Seq and analysis; S.J., J.K. and S.L. performed the neuronal differentiation and immunochemistry on neuronal-like cells; and analyzed the data; F.B., M.I. and M.M. carried out computational bioinformatics; C.E.F. and N.J.P. performed experiments using human iPSCs; S.O. and T.P. supervised the overall study, designed the experiments and wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01804-w.

Competing interests: S.O. and T.P. filed a patent about therapeutic use of cardiac-enriched lncRNAs. Patent title: “Diagnostic, prognostic and therapeutic uses of long noncoding RNAs for heart disease and regenerative medicine”; International application number: PCT/EP2014/078868; Applicant: University of Lausanne, Switzerland. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017