Functional characterisation of *cis*-regulatory elements governing dynamic *Eomes* expression in the early mouse embryo

Claire S. Simon[1], Damien J. Downes[2], Matthew E. Gosden[2], Jelena Telenius[2], Douglas R. Higgs[2], Jim R. Hughes[2], Ita Costello[1], Elizabeth K. Bikoff[1], and Elizabeth J. Robertson[1]

Author Affiliations:

1. The Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom
2. MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, United Kingdom

Corresponding Author: Elizabeth Robertson
Email: elizabeth.robertson@path.ox.ac.uk

Key Words:
Eomesodermin, enhancer, Capture-C, Nodal signaling, definitive endoderm

Summary Statement

Targeted genetic deletion and chromatin-conformation capture based characterisation of cis-regulatory elements governing dynamic Eomes expression identify an important endoderm enhancer required during mouse development.

Development • Advance article

Abstract

The T-box transcription factor (TF) Eomes is a key regulator of cell fate decisions during early mouse development. The *cis*-acting regulatory elements that direct expression in the anterior visceral endoderm (AVE), primitive streak (PS) and definitive endoderm (DE) have yet to be defined. Here, we identified three gene-proximal enhancer-like sequences (PSE_a, PSE_b and VPE) that faithfully activate tissue specific expression in transgenic embryos. However, targeted deletion experiments demonstrate that PSE_a and PSE_b are dispensable and only the VPE is required for optimal *Eomes* expression *in vivo*. Embryos lacking this enhancer display variably penetrant defects in anterior-posterior axis orientation and DE formation. Chromosome conformation capture experiments reveal VPE-promoter interactions embryonic stem cells (ESC), prior to gene activation. The locus resides in a large (500kb) pre-formed compartment in ESC and activation during DE differentiation occurs in the absence of 3D structural changes. ATAC-seq analysis reveals that VPE, PSE_a, and four additional putative enhancers display increased chromatin accessibility in DE associated with Smad2/3 binding coincident with transcriptional activation. In contrast, activation of the *Eomes* target genes *Foxa2* and *Lhx1* is associated with higher order chromatin reorganisation. Thus diverse regulatory mechanisms govern activation of lineage specifying TFs during early development.

Introduction

Reciprocal signaling cues between the pluripotent epiblast and adjacent tissues, namely the extra-embryonic ectoderm (ExE) and visceral endoderm (VE), precisely co-ordinate cell fate decisions during gastrulation. Nodal/Smad signals from the epiblast are required for specification of the AVE, a discrete signaling center that establishes anterior-posterior (A-P) polarity (Brennan et al., 2001; Robertson, 2014; Stower and Srinivas, 2014). The A-P axis initially becomes visible at gastrulation, when proximal posterior cells undergo an epithelial-to-mesenchymal transition (EMT) at the PS to form nascent mesoderm. Slightly later, following distal extension of the streak, endoderm progenitors delaminate and emerge onto the surface of the embryo (Kwon et al., 2008).

The T-box transcription factor (TF) Eomesodermin (Eomes), acting downstream of Nodal/Smad signals, is required to promote AVE formation and orientation of the A-P axis (Arnold et al., 2008a; Ciruna and Rossant, 1999; Nowotschin et al., 2013), as well as EMT of nascent mesoderm cells (Arnold et al., 2008a; Costello et al., 2011; Russ et al., 2000; van den Ameele et al., 2012). At post-implantation stages *Eomes* is expressed in the ExE and embryonic-VE, robustly induced at the onset of gastrulation in the PS, maintained in the anterior PS as it extends, before being abruptly lost coincident with node formation (Kwon and Hadjantonakis, 2007). Fate mapping experiments demonstrate that transient *Eomes* expression marks progenitors of the cardiovascular lineage, definitive endoderm (DE), node and midline (Costello et al., 2011).

Transgenic and targeted deletion approaches have provided insight into cell type specific developmental enhancers governing expression of key genes responsible for partitioning the pluripotent epiblast into discrete cell lineages. Proximal *cis*-regulatory regions within 20kb of the transcriptional start sites (TSS) directing spatiotemporally restricted expression of *Nodal*, *Mesp1/2* and *Lhx1* have been identified. Both the ASE, an intronic autoregulatory enhancer (Adachi et al., 1999; Norris and Robertson, 1999), and the Wnt signaling responsive 5' PEE (Ben-Haim et al., 2006) cooperatively regulate *Nodal* expression. Mutant embryos lacking these genomic sequences display dose-dependent defects in specification of mesoderm and DE/midline progenitors (Norris et al., 2002; Vincent et al., 2003). Similarly, the *Mesp1/2* genes, essential for formation of nascent mesoderm, are jointly regulated by the EME, an Eomes dependent enhancer (Costello et al., 2011; Haraguchi et al., 2001). Our recent work

demonstrates that *Lhx1*, required for AVE and anterior mesendoderm specification (Barnes et al., 1994; Shawlot and Behringer, 1995), is directly controlled by Eomes binding to a proximal promoter element (Nowotschin et al., 2013).

Eomes, rapidly induced in the proximal-posterior epiblast coincident with the acquisition of A-P polarity (Ciruna and Rossant, 1999), is widely viewed as a master regulator of mesendodermal lineages (Costello et al., 2011; Izumi et al., 2007; Teo et al., 2011; van den Ameele et al., 2012). Thus, *Eomes* represents the earliest lineage-specifying gene in the embryo-proper. However, relatively little is known about the *cis*-acting regulatory elements controlling its dynamic pattern of expression. Recent studies of mouse and human ESC have identified a conserved switch enhancer -7kb upstream of the TSS (Beyer et al., 2013; Kartikasari et al., 2013; Rada-Iglesias et al., 2011) that is repressed under self-renewing conditions (Teo et al., 2011), and becomes activated during mesoderm and endoderm differentiation. However, possible functional contributions made by this genomic region have yet to be assessed *in vivo*.

Here, we investigate the structural features of the locus that govern *Eomes* expression during early mouse development. Gain of function transgenic reporter assays identified three gene-proximal *Eomes* enhancer-like sequences (PSE_a, PSE_b and VPE). However, when we engineered germline deletions to evaluate their functional contributions *in vivo*, surprisingly, only the VPE was found to influence expression in the early embryo. We also exploited Next Generation (NG) Capture-C technology (Davies et al., 2016) to describe the 3D structural features of the locus. The *Eomes* promoter occupies a discrete 500kb regulatory compartment in ESC, and this chromatin conformation is not appreciably altered during DE differentiation. However, our ATAC-seq analysis revealed that the VPE, PSE_a and four additional distal regulatory elements located within this pre-formed compartment display increased chromatin accessibility and acquire Smad2/3 occupancy during DE differentiation. This mode of 3D genome organisation probably serves to facilitate rapid Nodal/Smad-dependent activation of the locus. In contrast, developmentally regulated *Foxa2* and *Lhx1* promoter-promoter and promoter-enhancer interactions seem to require substantial structural changes during the shift from transcriptionally inactive to active conformation.

Results

Identification of proximal *Eomes* enhancers active during gastrulation

Putative enhancer elements containing DNase hypersensitive sites and marked by H3K4me1, are considered to be active if also enriched for H3K27ac, or alternatively viewed as poised if enriched for H3K27me3 (Rada-Iglesias et al., 2011; Zentner et al., 2011). To identify candidate enhancers at the *Eomes* locus we examined ChIP-seq datasets from undifferentiated ESC, epiblast like cells (EpiLC) and mesodermal precursors (MES) (Alexander et al., 2015; Buecker et al., 2014; Consortium, 2012), corresponding to the E4.5 epiblast (ESC), the E5.5 epiblast (EpiLC) or E6.5 primitive streak (MES) cell populations.

We identified three DNase hypersensitive sites close to the *Eomes* promoter marked by H3K4me1 that show increased H3K27ac upon differentiation, including two sites (PSE_a and PSE_b) located close together, spanning a 5kb region between -11kb to -6kb upstream of the transcriptional start site (TSS), and a third candidate region (VPE) lying +8kb downstream of the TSS (Fig. 1A, Fig. S1A). Notably, the upstream cluster contains the previously described switch enhancer (PSE_b) activated during ESC differentiation to DE and mesendoderm (Beyer et al., 2013; Kartikasari et al., 2013) Additionally, two downstream DNaseI hypersensitive sites bound by CCCTC-binding factor (CTCF) were identified in ESC (Fig. S1A). The three proximal regions are highly conserved amongst mammals (Fig. S1A) and associated with H3K4me1/H3K27me3 in ESC, and thus probably represent poised enhancers, primed for activation. Consistent with a shift to the active state during the transition from pluripotency to lineage commitment, these regions contain increased H3K27ac and decreased H3K27me3 in EpiLC and MES. The homologous regions are also associated with active enhancer marks in human DE cultures (Fig. S1B).

To test activities of these candidate enhancers we generated transgenic strains carrying LacZ reporter constructs and subsequently examined embryonic expression at early post-implantation stages (Kothary et al., 1989). The 5kb upstream region was designated the PSE (**P**rimitive **S**treak **E**nhancer) because PSE-LacZ activity is restricted to the PS at early (ES), mid (MS) and late-streak (LS) stages (Fig. 1B). There was no detectable LacZ expression in the ExE or VE. On the other hand, the 0.7kb downstream enhancer designated the VPE (**V**isceral endoderm and **P**rimitive streak **E**nhancer), showed activity in the proximal-

posterior epiblast, and also in the AVE at pre-streak (PrS) stages (Fig. 1C). Slightly later, LacZ staining was detectable in the PS, nascent mesendoderm and the AVE, subsequently became restricted to the anterior PS, and was lost by LS stages. Collectively these three enhancers faithfully recapitulate the endogenous *Eomes* expression patterns within both the VE and embryo proper.

The PSE is dispensable for normal embryonic development

The 5kb PSE contains both an upstream element, PSE_a, as well as the previously described PSE_b switch enhancer reported to interact with the *Eomes* promoter during DE differentiation (Fig. S1A) (Beyer et al., 2013; Kartikasari et al., 2013). To investigate their functional activities in the context of the developing embryo we generated discrete germline targeted deletions (Fig. 2A, Fig. S2). Surprisingly, homozygous mice lacking the 2kb PSE_b genomic fragment ~8kb – ~6kb upstream of the TSS (ΔPSE_b) were recovered at Mendelian ratios and are indistinguishable from wild type littermates (Fig. 2B). These results demonstrate that the PSE_b is dispensable *in vivo*. It is well known that heterozygous mice carrying null alleles (*Eomes*$^{GFP/+,}$ *Eomes*$^{LacZ/+}$ or *Eomes*$^{\Delta exon2\text{-}5/+}$) are fully viable (Arnold et al., 2008a; Arnold et al., 2009; Russ et al., 2000). To investigate whether the PSE_b deletion may compromise transcriptional output, we crossed *Eomes*$^{\Delta PSE\_b\ /\ \Delta PSE\_b}$ mice to those carrying the *Eomes*$^{GFP/+}$ allele (hereafter referred to as *Eomes* null; *Eomes*$^{+/-}$). The resulting *Eomes*$^{\Delta PSE\_b/-}$ compound mutants develop normally (Fig 2C).

Next, we engineered a deletion that eliminates the entire 5kb PSE cluster (referred to as ΔPSE, Fig. S3). However, as for the PSE_b, removal of the entire PSE region in *Eomes*$^{\Delta PSE/\ \Delta PSE}$ mice has no noticeable effect on viability (Fig. 2B). Finally, crossing these deletion mutants with mice carrying the *Eomes* null allele also failed to perturb embryonic development (Fig. 2C). Thus, it appears that the PSE can activate expression in gain of function transgenic embryos. Nonetheless, this genomic region is clearly dispensable for *Eomes* expression *in vivo*.

Targeted deletion of the VPE leads to defective gastrulation

To investigate functional contributions made by the VPE we generated a targeted deletion lacking this 0.7kb region (Fig. S4). Homozygous ΔVPE mutants are viable and fertile (Fig. 2B). However, when we crossed *Eomes*$^{\Delta VPE/\Delta VPE}$ mice with *Eomes*$^{+/-}$ heterozygous animals carrying the null allele, we observed a significant under-representation of viable *Eomes*$^{\Delta VPE/-}$ compound heterozygotes (Fig. 2C), with approximately 40% (n=18) of the expected numbers recovered at weaning (equivalent to *Eomes*$^{\Delta VPE/+}$, n=44). These results strongly suggest that *Eomes*$^{\Delta VPE}$ acts as a hypomorphic allele.

Next, to determine the onset of lethality we examined embryos from E6.5 onwards. Approximately one third of *Eomes*$^{\Delta VPE/-}$ embryos are morphologically normal. However, two distinct classes of abnormal embryos were recovered at roughly equivalent numbers. The most severely affected (Class I) mutants arrest at early gastrulation stages while a second group (Class II) progress to mid gestation (Fig. 2D).

In Class I embryos the AVE marker *Hex* is induced at E6.5 but remains localised to the distal tip. Thus, the AVE is specified but fails to migrate towards the prospective anterior side of the embryo. These embryos fail to correctly orient the A-P axis, and lack a discrete PS. At E7.5 mesoderm (*Brachyury*) and DE (*Foxa2*) markers are restricted proximally. Class I mutant embryos, distinguished by the accumulation of disorganised mesenchymal cells in the epiblast cavity and a constriction at the embryonic and extra-embryonic boundary, phenocopy those selectively lacking *Eomes* activity in the VE (Nowotschin et al., 2013). Taken together with results above that demonstrate VPE-LacZ expression in the VE, the simplest explanation is that these abnormalities are caused by loss of *Eomes* function in the VE.

The Class II embryos, representing approximately a third of the *Eomes*$^{\Delta VPE/-}$ embryos, successfully establish normal A-P polarity. However, as gastrulation proceeds they display focal defects in the anterior PS (APS) and its derivatives the DE, midline, node and notochord. *Brachyury* expression in the PS fails to extend to the distal tip of the streak at E7.5. *Foxa2* positive DE progenitors are specified but fail to migrate anteriorly. As judged by *Afp* expression, the VE is retained over the epiblast and fails to become distally restricted. These tissue disturbances probably reflect *Eomes* functional loss within the APS (Arnold et

al., 2008a; Teo et al., 2011). APS derivatives are known to provide essential trophic signals required for patterning the anterior neurectoderm (Arkell and Tam, 2012). Consistent with this, at E9.5 class II mutant embryos display ventral closure and neural tube defects, fused or malformed somites, and loss of forebrain tissue.

The VPE is required for optimal *Eomes* expression levels

To directly test whether targeted loss of the VPE compromises *Eomes* transcriptional output, we eliminated the VPE in the context of our *Eomes*^{GFP} reporter allele containing an EGFP-pA cassette inserted in-frame at the translational start site in exon 1 (Fig. 3A, Fig. S5) (Arnold et al., 2009) and performed flow cytometry analysis to quantify expression levels. The *Eomes*^{GFP} reporter is robustly activated during ESC differentiation to embryoid bodies (EBs) (Costello et al., 2011) (Fig. 3B). As shown in Fig. 3C, GFP expression is dramatically reduced in *Eomes*$^{GFP\Delta VPE/+}$ EBs as compared to *Eomes*$^{GFP/+}$ EBs. The VPE deletion results in markedly reduced expression to 42% of the control *Eomes*$^{GFP/+}$ EBs (student's t-test p=0.05) (Fig. 3D).

These heterogenous EB cultures contain mixtures of cardiac mesoderm, DE and VE Eomes^{+} cell populations. To investigate the impact of the VPE deletion *in vivo*, we generated *Eomes*$^{GFP\Delta VPE/+}$ mice and examined expression during gastrulation. GFP expression in *Eomes*$^{GFP\Delta VPE/+}$ embryos recapitulates domains of the *Eomes*$^{GFP/+}$ control embryos at E6.5, in the ExE, PS, nascent mesoderm, and VE (Fig. 3E,F). The VPE deletion reduced expression levels but tissue specific expression patterns were unperturbed. Similar conclusions were reached by whole-mount *in situ* hybridisation (WISH) experiments examining *Eomes* mRNA expression in *Eomes*$^{\Delta VPE/\Delta VPE}$ embryos (Fig. S4E). Thus, reduced *Eomes* transcription (~50%) as in *Eomes*$^{+/-}$ or *Eomes*$^{\Delta VPE/\ \Delta VPE}$ embryos is sufficient to promote A-P axis specification and gastrulation. However, as shown above, further reduced expression (~25%) in Eomes$^{\Delta VPE/-}$ embryos results in gastrulation defects.

FoxH1-independent Nodal/Smad2/3 signals regulate VPE activity

Eomes activation in the VE and PS depends on Nodal/Smad signals (Brennan et al., 2001; Nowotschin et al., 2013). To investigate Nodal/Smad requirements in cultured EBs, we used the small molecule SB-431542 (SB), a potent inhibitor of type 1 Activin receptor like kinases

4, 5 and 7. As expected, in control cultures maximal *Eomes* expression was detectable between d3.5 and d4 (Fig. 4A). *Eomes* expression was dramatically reduced in cultures treated with the SB inhibitor from d3, and by d4 is severely compromised to just 2% of that seen in controls (Fig. 4A). These results confirm that Nodal signaling is required to induce *Eomes* expression during the transition from pluripotency to lineage commitment. Additionally when we compared Smad2/3 ChIP-seq datasets in ESC and DE cultures (Yoon et al., 2015), we found evidence for Smad2/3 occupancy at the VPE specifically in DE cultures (Fig. 4B). These observations strengthen the idea that Nodal/Smad signals controlling *Eomes* expression activate transcription via the VPE.

It is well known that the forkhead transcription factor FoxH1 functions as a Smad2/3 co-factor governing Nodal/Smad target gene expression (Attisano et al., 2001; Izzi et al., 2007). FoxH1 has been proposed to act as a pioneer factor and recruit Smad2/3 complexes to switch enhancers, activated as ESC transition to DE fates (Beyer et al., 2013; Cirillo et al., 2002; Cirillo and Zaret, 1999; Kim et al., 2011). Interestingly, the VPE Smad2/3 peak also contains a conserved FoxH1 binding motif. Moreover, the VPE region is co-bound by FOXH1, SMAD2/3, and SMAD4 in human DE cultures (Fig. S6) (Beyer et al., 2013; Brown et al., 2011; Kim et al., 2011; Teo et al., 2011). Consistent with the idea that FoxH1 cooperatively activates *Eomes* expression via the VPE, homozygous null *FoxH1*$^{-/-}$ embryos phenocopy the *Eomes*$^{\Delta VPE/-}$ embryos, displaying either defective AVE formation prior to gastrulation, or disturbances in APS specification at later stages (Hoodless et al., 2001; Yamamoto et al., 2001).

To directly evaluate *FoxH1* functional contributions, we analysed *Eomes* expression at E6.5 and E7.5 in the context of *FoxH1*$^{-/-}$ mutant embryos (Fig. 4C). In mutants with AVE/DVE defects at E6.5 *Eomes* is expressed in the thickened VE at the distal tip of the embryo, and at E7.5 in the chorion and proximal epiblast. *FoxH1* mutants with APS defects express *Eomes* in the ExE and PS. *Eomes* is clearly expressed in both classes of *FoxH1* mutant embryos. Slightly reduced levels in the PS can be explained due to the loss of FoxH1-dependent activation of the auto-regulatory ASE *Nodal* enhancer (Norris et al., 2002). In striking contrast to *Eomes/Nodal* double heterozygotes (Arnold et al., 2008a), we found no evidence here for *Eomes* and *FoxH1* genetic interactions. Indeed, *Eomes* and *FoxH1* compound mutant mice are fully viable (Fig. 4D). Finally, to confirm that VPE activity is FoxH1 independent, we examined expression of the VPE-LacZ transgene in *FoxH1* mutant

embryos. LacZ staining is detectable throughout the epiblast at E6.5 (Fig. 4E), and also in the thickened VE at the distal tip. FoxH1 function is nonessential for VPE-LacZ reporter activity. Thus, we conclude that Nodal/Smad signals activate *Eomes* expression in a FoxH1-independent manner raising the possibility that other forkhead family members may recruit Smad2/3 complexes during Eomes induction in vivo.

Characterisation of the *Eomes* 3D regulatory chromatin compartment during endoderm differentiation

The finding that the VPE targeted deletion partially reduces but fails to completely eliminate *Eomes* expression, strongly suggests that additional regulatory elements contribute to transcriptional output of the locus. Enhancer interactions with target promoters have been analysed by chromatin conformation capture techniques (de Wit and de Laat, 2012). We took advantage of the recently developed Next Generation (NG) Capture-C methodology (Davies et al., 2016) to screen for *Eomes* regulatory enhancer elements. During DE differentiation *Eomes* expression increased by ~600 fold (Fig. S7B) resulting in activation of the Eomes target genes, *Lhx1* and *Foxa2* (Fig. S7C) (Nowotschin et al., 2013; Teo et al., 2011).

NG Capture-C using viewpoints from the PSE_a and PSE_b exhibited promoter interactions in ESC (Fig. S8) when analysed with FourCseq (Klein et al., 2015). These interactions were marginally reduced in DE. However the overall change was not statistically significant. By contrast NG Capture-C revealed significant interactions between the VPE and the *Eomes* promoter in both ESC and DE cells (Fig. S8). Thus, the locus appears to be primed for activation prior to expression.

Next, performing Capture-C using a viewpoint from the *Eomes* promoter revealed that the *Eomes* locus, together with an upstream 300kb gene desert, and its neighboring genes *Azi2* and *Cmc1*, occupies a discrete ~500kb chromatin compartment (Fig. 5A). This region contains numerous CTCF binding sites (Handoko et al., 2011). Consistent with CTCF-mediated chromatin loops forming the compartment boundaries, motif analysis suggests that the outermost binding sites face inwards (Fig. 5A). This compartment structure is readily detectable in both ESC and DE cells but is completely absent in control terminally differentiated erythrocytes lacking *Eomes* expression (Fig. 5A, Fig. S9). Comparison of the NG Capture-C data from ESC and DE, in which the Eomes locus is transcriptionally silent or

active, respectively, demonstrates that the compartment is highly stable. Moreover there were no detectable changes in long-range promoter interactions within the compartment (Fig. S10).

To map changes in regions of open chromatin associated with *Eomes* activation and identify potential novel DE enhancers within the compartment we performed ATAC-seq. We identified 85,581 total peaks in ESC and DE, and of these 19% were gained and 32.5% lost during differentiation (Fig. S9). Within the *Eomes* compartment we identified 6 regions that show increased accessibility in DE, including the VPE and the PSE_a as well as four additional sites at -93kb, -45kb, -38kb and +9kb relative to the *Eomes* TSS (Fig. 5B).

Next, we examined Smad2/3 binding across the compartment (Yoon et al., 2015). Smad2/3 occupancy was detectable in DE but not in ESC at all six of the differentially accessible sites (Fig. 5B). These findings demonstrate the *Eomes* locus is organised into a large 3D regulatory chromatin compartment in pluripotent ESC that is maintained upon DE differentiation. Global structural changes are not required for *Eomes* induction during DE differentiation. Rather, transcriptional activation seems to reflect increased chromatin accessibility and Smad2/3 recruitment at DE enhancers. The -95kb and -45kb regions, and to a lesser extent the -38kb region, are associated with poised and active enhancer marks as cells transition from ES to Epi to MES states respectively (Fig. S11). Additionally recently published TF ChIP-seq data demonstrate that the -45kb ATAC-seq peak together with the PSE_a and VPE are co-bound by Tcf3 in DE (Wang et al., 2017), suggesting that both Nodal and Wnt signaling converge on these enhancer regions during gastrulation (Ben-Haim et al., 2006). Consistent with its activities as a key Eomes regulatory element during DE specification, the VPE is also bound by Otx2 and Lhx1 in EpiLC and mesendoderm cultures respectively (Buecker et al., 2014; Costello et al., 2015).

Foxa2 and *Lhx1* promoters form long-range interactions in polycomb bodies

The forkhead TF Foxa2 and the LIM domain homeobox TF Lhx1 function together with Eomes as master regulators of APS cell fates (Ang and Rossant, 1994; Costello et al., 2015; Perea-Gomez et al., 1999; Shawlot and Behringer, 1995). One possible model is that this pre-configured genomic structure might be a common feature shared by endoderm specific transcriptional factors (Fig. S7C). As for *Eomes*, Capture-C of the *Foxa2* and *Lhx1* promoters demonstrates localisation within pre-formed compartments (both ~350kb) in ESC, but not

erythrocytes where the genes are inactive (Fig. 6A,B). However, these *Foxa2* and *Lhx1* compartments were found to undergo significant rearrangements during DE differentiation (Fig. 6A,B). Unlike *Eomes*, in ESC *Lhx1* and *Foxa2* promoters both make long-range contacts with neighboring developmental genes lying outside the compartment boundaries (Fig. 6A,B). These long-range interactions range from 370kb to 1.8Mb in size and are almost entirely specific to gene promoters (Table S3), and are lost as cells acquire a DE fate (Fig. 6A,B).

Both *Foxa2* and *Lhx1* are repressed by polycomb in ESC (Leeb et al., 2010). Examination of published ESC ChIP-seq data-sets for Polycomb components Ezh2, Suz12 (PRC2) and Ring1b (PRC1) (Chen et al., 2008; Ku et al., 2008) as well as the polycomb repressive mark H3K27me3 (Yue et al., 2014) showed they are present at all of the promoters of the adjacent genes with which *Lhx1* and *Foxa2* interact (Fig. 6), suggesting that these genes are present in Polycomb bodies (Pirrotta and Li, 2012). Interestingly, these Polycomb repressive components are also present at the *Eomes* promoter in ESC, but we found no evidence for long-range interactions with gene promoters lying outside the compartment (Fig. S10). Collectively, results above demonstrate that three essential TFs required for cell fate specification, *Eomes*, *Foxa2*, and *Lhx1*, were found to exhibit distinct modes of 3D chromatin organisation during differentiation.

Discussion

The spatiotemporal expression of key lineage specifying transcription factors (TF) is tightly controlled during early mouse development to ensure correct cell fate decisions. Interactions of cell-type specific *cis*-acting enhancer elements with gene promoters, within topologically discrete chromatin compartments, directs developmentally regulated patterns of expression (de Laat and Duboule, 2013). Our recent studies demonstrate that the T-box TF Eomes, dynamically expressed in the VE, ExE and PS during gastrulation, acts downstream of the Nodal signaling pathway as an essential master-regulator of the DE and cardiac mesoderm cell lineages. Here, we exploit transgenic reporter assays, targeted deletion, and NG Capture-C strategies to investigate the regulatory landscape at the *Eomes* locus.

We demonstrate in gain of function experiments that conserved proximal *cis*-regulatory elements, namely the so-termed PSE (comprising of PSE_a and PSE_b) and the VPE, have the ability to drive reporter activity in the PS, or VE and PS, respectively. The conserved

Eomes PSE_b region, representing an archetypal poised developmental enhancer in both human and mouse ESC, was recently shown to be activated upon mesendoderm induction in response to Nodal (Smad2/3, FoxH1) and Wnt (β-cat) signaling pathways (Beyer et al., 2013; Brown et al., 2011; Buecker and Wysocka, 2012; Estaras et al., 2015; Funa et al., 2015; Kartikasari et al., 2013; Kim et al., 2011; Rada-Iglesias et al., 2011). However, surprisingly our targeted deletion experiments demonstrate that this switch enhancer, and the adjacent PSE_a, are dispensable for correct developmentally regulated *Eomes* expression in the early embryo. Moreover, mutant mice that entirely lack this genomic region develop normally and are viable and fertile.

Eomes is required for the maintenance and migration of the AVE (Nowotschin et al., 2013). Additionally, robust expression in the PS is essential for formation of APS progenitors (Arnold et al., 2008a). The present results demonstrate that the VPE activates expression in both the AVE and PS, and makes important functional contributions governing Eomes activities during gastrulation. We found that removal of this element halves transcriptional output from the locus as assessed *in vitro*. Moreover, $Eomes^{\Delta VPE/-}$ embryos exhibit pleiotropic tissue defects due to compromised specification of AVE or APS, that closely resemble those caused by defective Nodal signaling or loss of the Smad2/3/4 co-factor, *FoxH1* (Arnold et al., 2008a; Hoodless et al., 2001; Norris et al., 2002; Yamamoto et al., 2001).

Our NG Capture-C experiments revealed that the VPE directly interacts with the *Eomes* promoter in both ESC and DE. Moreover the *Eomes* locus lies within a large pre-formed 3D regulatory chromatin compartment in pluripotent ESC that is maintained upon differentiation to DE. Thus activation of the locus occurs in the absence of remodeling long-range interactions. In contrast, previous studies of mouse and human ESC implicate *de novo* enhancer-promoter interactions during DE and mesendoderm differentiation (Estaras et al., 2015; Kartikasari et al., 2013). These inconsistencies probably reflect technical differences since a target lead (one-versus-some) 3C PCR technique was used previously, as compared to the unbiased (one-versus-all) NG Capture-C sequencing approach exploited here.

NG Capture-C analysis of the direct Eomes targets, *Foxa2* and *Lhx1*, known to regulate APS fates, demonstrates they similarly occupy discrete regulatory compartments in transcriptionally silent ESC. However, in contrast to *Eomes*, *Foxa2* and *Lhx1* promoters display contacts with polycomb associated gene promoters lying far outside their compartments. These associations are specifically lost during DE differentiation (Fig. 7).

Promoter-promoter interactions within ESC are often occupied by polycomb repressive complexes (PRC) that organise the 3D chromatin structure into polycomb bodies to silence gene expression (Denholtz et al., 2013; Schoenfelder et al., 2015; Sexton et al., 2012; Williamson et al., 2014). These epigenetic barriers are thought to block lineage-specifying gene activation and thus prevent precocious differentiation. We demonstrate here that in contrast to *Foxa2* and *Lhx1* the *Eomes* locus exhibits a distinct mode of regulation. Rather, in the absence of polycomb mediated repressive contacts, the *Eomes* promoter can rapidly respond to dynamic signaling cues during gastrulation (Fig. 7).

Considerable evidence suggests that stable enhancer-promoter interactions within pre-formed chromatin compartments initiate transcription through the release of paused polymerase (de Laat and Duboule, 2013; Ghavi-Helm et al., 2014; Jin et al., 2013; Williamson et al., 2016). We found that promoter-enhancer interactions are relatively stable. However our ATAC-seq experiments reveal significant changes in open chromatin regions during DE differentiation. We identified several candidate enhancers within the *Eomes* compartment, that display increased chromatin accessibility and are greatly enriched for Smad2/3 occupancy upon DE differentiation (Yoon et al., 2015). Moreover, we confirm that Smad2/3 is required for *Eomes* activation, as inhibition of receptor mediated Nodal/Smad2/3 signaling blocks transcription. Smad2/3 associations with the histone demethylase Jmjd3 are known to be required for the activation of Nodal target genes (Dahle et al., 2010; Kartikasari et al., 2013). Jmjd3 activates poised developmental genes by removing promoter-proximal H3K27me3 and releasing paused polymerase (Chen et al., 2012). We propose that the poised chromatin architecture at the *Eomes* locus is permissive for rapid transcriptional induction in response to localised Nodal signaling during gastrulation, primarily via enhancer binding of Smad2/3/Jmjd3 complexes to release promoter-paused polymerase.

The 3C technologies developed over the past two decades have provided important new insights into the regulatory chromatin landscapes that orchestrate tissue-specific transcription. Here, we characterise for the first time *cis*-regulatory elements that activate *Eomes* expression during gastrulation, and describe the higher order chromatin architecture of the locus. We speculate that the pre-formed chromatin compartment and the absence of additional epigenetic safeguards prior to expression facilitates the rapid induction of *Eomes* expression in response to dynamic signaling cues at the onset of gastrulation. However, the stage of embryonic development that these compartments are established, and later

dismantled, remains elusive. Future studies will investigate whether these enhancers and permissive chromatin configuration are tissue-invariant and may also control cell type specific *Eomes* expression governing cell fate decisions at other sites such as the developing cortex, and adult NK and CD8+ T cell lineages (Arnold et al., 2008b; Gordon et al., 2012; Pearce et al., 2003).

Materials and Methods

Animals and PCR genotyping

Eomes$^{GFP/+}$ (Arnold et al., 2009) and *FoxH1*$^{+/-}$ (Hoodless et al., 2001) strains were genotyped as described. *Eomes*$^{\Delta PSE/+}$, *Eomes*$^{\Delta PSE\_b/+}$, *Eomes*$^{\Delta VPE/+}$ and *Eomes*$^{GFP\Delta VPE/+}$ strains were generated from targeted ES cell clones using standard methods (Arnold et al., 2009) (Fig. S2-S5, Supplementary Methods), and maintained on a mixed 129Sv/Ev/C57BL/6 background. To generate PSE.LacZ and VPE.LacZ transgenic constructs the 4.6 kb HincII – KpnI PSE fragment and a 696bp PCR amplified VPE sequence (Table S1), were cloned upstream of a hsp68 promoter, LacZ cassette and SV40 polyA signal (Sasaki and Hogan, 1996). Zygotes were injected with NotI linearised plasmid and transferred into pseudo-pregnant foster females. Embryos were either collected at E6.5-E7.5 or used to establish stable transgenic mouse lines. PCR genotyping primers are listed in Table S1. All animal experiments were performed in accordance with Home Office (UK) regulations and approved by the University of Oxford Local Ethical Committee.

ESC differentiation

ES cell lines were maintained in DMEM (Invitrogen) supplemented with 15% fetal calf serum (Gibco), 1% Penicillin/Streptomycin (Invitrogen), 0.1mM 2-mercaptoethanol (Sigma), 1% Glutamine (Invitrogen), 1% MEM Non-essential amino acids (Gibco), 1mM Sodium Pyruvate (Sigma), 1000U/ml LIF (ESGRO) on gelatin coated plates.

For analysis of GFP reporter expression wild type (CCE), *Eomes*$^{GFP\Delta VPE/+}$ and *Eomes*$^{GFP/+}$ ES cells were seeded as 10 μl hanging drops (1x10^4 cells/ml) in the absence of LIF to induce EB formation. After 2 days EBs were transferred to suspension culture. For SB inhibition experiments, ES cells were seeded in suspension at low density (1x10^4 cells/ml) in the absence of LIF to form EBs. On day 3 EBs were cultured in the presence or absence of 10 μM SB431542 inhibitor (Tocris). For DE differentiation, ES cells were induced to form EBs in suspension as described above, but were transferred on day 2 into N2B27 medium (Cellartis) supplemented with 20ng/ml ActivinA (R&D systems) and 20ng/ml EGF (Peprotech) to induce DE differentiation (Morrison et al., 2008). For Capture-C, ChIP-seq, and ATAC-seq experiments EBs were dissociated by incubation with 0.25% trypsin (Gibco) for 3 min at 37˚C with constant agitation followed by gentle pipetting to obtain a single cell suspension.

RNA analysis

RNA was isolated from using Qiashredder homogenizer (Qiagen), RNeasy mini kit (Qiagen) and RNase-Free DNase Set (Qiagen). RNA was reverse transcribed to cDNA using Superscript III First Strand Synthesis System (Life Technologies) and qRT-PCR was carried out in triplicate using SYBR-green kit (Qiagen) on a Rotagene cycler (Qiagen) with primers listed in Table S1. Relative gene expression was normalised to *Gapdh* and calculated as $2^{\Delta\Delta Ct}$.

In situ hybridisation, X-gal staining and Immunofluorescence

Whole-mount *in situ* hybridisation (WISH) was performed according to published protocols (Behringer et al., 2013). LacZ activity was visualised by whole-mount X-gal staining as described (Behringer et al., 2013). Whole-mount WISH and X-gal stained embryos were photographed after clearing in 80% glycerol.

For immunofluorescence, embryos were fixed overnight in 1% PFA. EBs were fixed in 4% PFA for 30 min at room temperature. Samples were washed in 0.1% Triton-X in PBS, permeabilised in 0.5% Triton-X in PBS for 15 min, washed in 0.1% Triton-X in PBS, then blocked in 0.1% Triton-X, 0.2% BSA, 5% donkey serum in PBS for 2 hours at room temperature. Samples were incubated with primary antibodies (Table S2) overnight at 4˚C, washed, incubated with secondary antibodies or Phalloidin AlexaFluor 633 stain (A22284; Invitrogen) in block solution for 2 hours at room temperature, counterstained with DAPI and mounted in Vectashield (Vector Laboratories) on chamber slides (LabTek). Images were acquired using an Olympus FV1000 inverted confocal microscope.

Flow cytometry

Day 4 EBs were incubated in 0.25% trypsin for 5 min at 37˚C and dissociated into single cells using a 20-guage needle. FACS analysis was performed using a BD FACSCalibur 4 (BD Biosciences) and data analysed using FlowJo.

ATAC-seq

Tagmentation and indexing of single cell suspensions of ESC, DE and erythrocytes from phenylhydrazine treated mice (Davies et al., 2016) was performed as previously described (Buenrostro et al., 2013; Hay et al., 2016). Samples were sequenced using a 75-cycle paired-end kit on the Illumina NextSeq platform.

ChIP-seq

Single cell suspensions ($5x10^6$) were cross-linked in 1% formaldehyde for 15 minutes at room temperature and processed using standard methods. Briefly, cells were lysed on ice for 20 minutes (5mM PIPES, 85mM KCl, 0.5% Igepal-CA 630), and pelleted nuclei lyzed (50mM Tris-HCl, 10mM EDTA, 1% SDS). Sonicated chromatin was incubated overnight with anti-H3K4me3 (2µl; 07-473; Millipore) and Protein A/G Dynabeads (Invitrogen). Beads were washed with RIPA buffer variants (10mM Tris-HCl, 1mM EDTA, 0.5mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Sodium Deoxycholate): RIPA, High Salt RIPA (500mM NaCl), RIPA with 250mM LiCl and T.E. Buffer before RNase A (Roche) and Proteinase K (Thermo Fisher) treatment. Phenol-chloroform extracted DNA was indexed using NebNext Ultra II (New England BioLabs), multiplexed and sequenced using a 75-cycle paired-end kit on the Illumina NextSeq platform.

ATAC-seq and ChIP-seq analysis

ATAC-seq and ChIP-seq data were analysed as described (Hay et al., 2016) using a custom pipeline (http://userweb.molbiol.ox.ac.uk/public/telenius/PipeSite.html). Sequenced reads were aligned using Bowtie to the mm9 build of the mouse genome. Genomic browser tracks were generated from pooled data from multiple replicates and normalized per million mapped reads using a custom Perl script. Peak detection was performed with the MACS2 (Feng et al., 2012). For differential analysis, a union set of peaks for each cell type generated from ≥2 peak calls per site. Peaks were filtered for high ploidy regions using MIG Viewer (McGowan et al., 2013). CTCF-motifs were identified using the FIMO function of MEME Suite (Bailey et al., 2009; Grant et al., 2011).

NG Capture-C and analysis

NG Capture-C was performed as described (Davies et al., 2016) on single cell suspensions of ESC, DE or erythrocytes. Samples were indexed for multiplexing and co-capture of enhancers or promoters using biotinylated 120mers (Sigma, IDT) designed with the CapSequm webtool (http://apps.molbiol.ox.ac.uk/CaptureC/cgi-bin/CapSequm.cgi) (Hughes et al., 2014) and pooled to a final concentration of 2.9nM (Table S4). Captured material was pooled and sequenced using the Illumina NextSeq platform with 150-bp paired-end reads (300 cycle kit, Illumina). Reads were mapped using Capture-C scripts (https://github.com/telenius/captureC/releases), and analyzed as previously described (Hay et

al., 2016), and additionally with FourCSeq (Klein et al., 2015) and DESeq2 (Love et al., 2014).

Acknowledgements

We thank Julie Baker for valuable discussions during the initial stage of the project, Ben Davies and Chris Preece (Wellcome Trust Centre for Human Genetics) and Jonathan Godwin (Department of Biochemistry, University of Oxford) for generating transgenic mice, the Dunn School Bioimaging Facility, Nigel Rust for flow cytometry, Xin Sun for help with gene targeting, and Thomas Clague for genotyping assistance.

Competing Interests

The Authors declare no competing or financial interests

Author Contributions

C.S.S, I.C., E.K.B. and E.J.R. designed the project; C.S.S., D.J.D., M.E.G. and E.J.R. performed the experiments; C.S.S., D.J.D, J.T., D.R.H., J.R.H, E.K.B. and E.J.R. analysed and interpreted the data; C.S.S., D.J.D., E.K.B and E.J.R. wrote the paper.

Funding

This work was supported by a Wellcome Trust Grant [WT 102811 to E.J.R.] an Edward Penley Abraham studentship [to C.S.S.], and by a Wellcome Trust Strategic Award [WT 106130 to J.R.H. and D.R.H] and a Medical Research Council Award [4050189188 to J.R.H. and D.R.H.] E.J.R. is a Wellcome Trust Principal Research Fellow.

Data Availability

ChIP-seq, ATAC-seq and NG Capture-C data have been deposited in NCBI GEO (submission in process). Accession numbers of published ChIP-seq data sets used in this study are listed in Supplementary Table S5.

Supplementary Information

Supplementary information in the form of 11 Figures and 5 Tables accompanies this manuscript.

References

Adachi, H., Saijoh, Y., Mochida, K., Ohishi, S., Hashiguchi, H., Hirao, A. and Hamada, H. (1999). Determination of left/right asymmetric expression of nodal by a left side-specific enhancer with sequence similarity to a lefty-2 enhancer. *Genes Dev* **13**, 1589-1600.

Alexander, J. M., Hota, S. K., He, D., Thomas, S., Ho, L., Pennacchio, L. A. and Bruneau, B. G. (2015). Brg1 modulates enhancer activation in mesoderm lineage commitment. *Development* **142**, 1418-1430.

Ang, S. L. and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development. *Cell* **78**, 561-574.

Arkell, R. M. and Tam, P. P. (2012). Initiating head development in mouse embryos: integrating signalling and transcriptional activity. *Open biology* **2**, 120030.

Arnold, S. J., Hofmann, U. K., Bikoff, E. K. and Robertson, E. J. (2008a). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* **135**, 501-511.

Arnold, S. J., Huang, G. J., Cheung, A. F., Era, T., Nishikawa, S., Bikoff, E. K., Molnar, Z., Robertson, E. J. and Groszer, M. (2008b). The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. *Genes Dev* **22**, 2479-2484.

Arnold, S. J., Sugnaseelan, J., Groszer, M., Srinivas, S. and Robertson, E. J. (2009). Generation and analysis of a mouse line harboring GFP in the Eomes/Tbr2 locus. *Genesis* **47**, 775-781.

Attisano, L., Silvestri, C., Izzi, L. and Labbe, E. (2001). The transcriptional role of Smads and FAST (FoxH1) in TGFbeta and activin signalling. *Molecular and cellular endocrinology* **180**, 3-11.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37**, W202-208.

Barnes, J. D., Crosby, J. L., Jones, C. M., Wright, C. V. and Hogan, B. L. (1994). Embryonic expression of Lim-1, the mouse homolog of Xenopus Xlim-1, suggests a role in lateral mesoderm differentiation and neurogenesis. *Dev Biol* **161**, 168-178.

Behringer, R., Gertsenstein, M., Nagy, K. and Nagy, A. (2013). *Manipulating the mouse embryo : a laboratory manual, fourth edition* (Fourth edition. edn): Cold Spring Harbor Laboratory Press.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E. J. and Constam, D. B. (2006). The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. *Dev Cell* **11**, 313-323.

Beyer, T. A., Weiss, A., Khomchuk, Y., Huang, K., Ogunjimi, A. A., Varelas, X. and Wrana, J. L. (2013). Switch enhancers interpret TGF-beta and Hippo signaling to control cell fate in human embryonic stem cells. *Cell reports* **5**, 1611-1624.

Brennan, J., Lu, C. C., Norris, D. P., Rodriguez, T. A., Beddington, R. S. and Robertson, E. J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965-969.

Brown, S., Teo, A., Pauklin, S., Hannan, N., Cho, C. H., Lim, B., Vardy, L., Dunn, N. R., Trotter, M., Pedersen, R., et al. (2011). Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* **29**, 1176-1185.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T. and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838-853.

Buecker, C. and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends in genetics : TIG* **28**, 276-284.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213-1218.

Chen, S., Ma, J., Wu, F., Xiong, L. J., Ma, H., Xu, W., Lv, R., Li, X., Villen, J., Gygi, S. P., et al. (2012). The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. *Genes Dev* **26**, 1364-1375.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell* **9**, 279-289.

Cirillo, L. A. and Zaret, K. S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Molecular cell* **4**, 961-969.

Ciruna, B. G. and Rossant, J. (1999). Expression of the T-box gene Eomesodermin during early mouse development. *Mech Dev* **81**, 199-203.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

Costello, I., Nowotschin, S., Sun, X., Mould, A. W., Hadjantonakis, A. K., Bikoff, E. K. and Robertson, E. J. (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes Dev* **29**, 2108-2122.

Costello, I., Pimeisl, I. M., Dräger, S., Bikoff, E. K., Robertson, E. J. and Arnold, S. J. (2011). The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nat Cell Biol* **13**, 1084-1091.

Dahle, O., Kumar, A. and Kuehn, M. R. (2010). Nodal signaling recruits the histone demethylase Jmjd3 to counteract polycomb-mediated repression at target genes. *Sci Signal* **3**, ra48.

Davies, J. O. J., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R. and Hughes, J. R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nature methods* **13**, 74-+.

de Laat, W. and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499-506.

de Wit, E. and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev* **26**, 11-24.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M. and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602-616.

Estaras, C., Benner, C. and Jones, K. A. (2015). SMADs and YAP Compete to Control Elongation of beta-Catenin:LEF-1-Recruited RNAPII during hESC Differentiation. *Molecular cell* **58**, 780-793.

Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nature protocols* **7**, 1728-1740.

Funa, N. S., Schachter, K. A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honore, C., Hansen, K. and Semb, H. (2015). beta-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell* **16**, 639-652.

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. and Furlong, E. E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96-100.

Gordon, S. M., Chaix, J., Rupp, L. J., Wu, J., Madera, S., Sun, J. C., Lindsten, T. and Reiner, S. L. (2012). The transcription factors T-bet and Eomes control key checkpoints of natural killer cell maturation. *Immunity* **36**, 55-67.

Grant, C. E., Bailey, T. L. and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018.
Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W., Ye, C., Ping, J. L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43**, 630-638.
Haraguchi, S., Kitajima, S., Takagi, A., Takeda, H., Inoue, T. and Saga, Y. (2001). Transcriptional regulation of Mesp1 and Mesp2 genes: differential usage of enhancers during development. *Mech Dev* **108**, 59-69.
Hay, D., Hughes, J. R., Babbs, C., Davies, J. O., Graham, B. J., Hanssen, L. L., Kassouf, M. T., Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., et al. (2016). Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet* **48**, 895-903.
Hoodless, P. A., Pye, M., Chazaud, C., Labbe, E., Attisano, L., Rossant, J. and Wrana, J. L. (2001). FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes Dev* **15**, 1257-1271.
Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R. and Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**, 205-212.
Izumi, N., Era, T., Akimaru, H., Yasunaga, M. and Nishikawa, S. (2007). Dissecting the molecular hierarchy for mesendoderm differentiation through a combination of embryonic stem cell culture and RNA interference. *Stem Cells* **25**, 1664-1674.
Izzi, L., Silvestri, C., von Both, I., Labbe, E., Zakin, L., Wrana, J. L. and Attisano, L. (2007). Foxh1 recruits Gsc to negatively regulate Mixl1 expression during early mouse development. *The EMBO journal* **26**, 3132-3143.
Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C. A., Schmitt, A. D., Espinoza, C. A. and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294.
Kartikasari, A. E., Zhou, J. X., Kanji, M. S., Chan, D. N., Sinha, A., Grapin-Botton, A., Magnuson, M. A., Lowry, W. E. and Bhushan, A. (2013). The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation. *The EMBO journal* **32**, 1393-1408.
Kim, S. W., Yoon, S. J., Chuong, E., Oyolu, C., Wills, A. E., Gupta, R. and Baker, J. (2011). Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev Biol* **357**, 492-504.
Klein, F. A., Pakozdi, T., Anders, S., Ghavi-Helm, Y., Furlong, E. E. and Huber, W. (2015). FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **31**, 3085-3091.
Kothary, R., Clapoff, S., Darling, S., Perry, M. D., Moran, L. A. and Rossant, J. (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707-714.
Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242.
Kwon, G. S. and Hadjantonakis, A. K. (2007). Eomes::GFP-a tool for live imaging cells of the trophoblast, primitive streak, and telencephalon in the mouse embryo. *Genesis* **45**, 208-217.
Kwon, G. S., Viotti, M. and Hadjantonakis, A. K. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev Cell* **15**, 509-520.
Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K. and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* **24**, 265-276.
Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550.

Mathelier, A., Fornes, O., Arenillas, D. J., Chen, C. Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic acids research* **44**, D110-115.

McGowan, S. J., Hughes, J. R., Han, Z. P. and Taylor, S. (2013). MIG: Multi-Image Genome viewer. *Bioinformatics* **29**, 2477-2478.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560.

Morrison, G. M., Oikonomopoulou, I., Migueles, R. P., Soneji, S., Livigni, A., Enver, T. and Brickman, J. M. (2008). Anterior definitive endoderm from ESCs reveals a role for FGF signaling. *Cell Stem Cell* **3**, 402-415.

Norris, D. P., Brennan, J., Bikoff, E. K. and Robertson, E. J. (2002). The Foxh1-dependent autoregulatory enhancer controls the level of Nodal signals in the mouse embryo. *Development* **129**, 3455-3468.

Norris, D. P. and Robertson, E. J. (1999). Asymmetric and node-specific nodal expression patterns are controlled by two distinct cis-acting regulatory elements. *Genes Dev* **13**, 1575-1588.

Nowotschin, S., Costello, I., Piliszek, A., Kwon, G. S., Mao, C. A., Klein, W. H., Robertson, E. J. and Hadjantonakis, A. K. (2013). The T-box transcription factor Eomesodermin is essential for AVE induction in the mouse embryo. *Genes Dev* **27**, 997-1002.

Pearce, E. L., Mullen, A. C., Martins, G. A., Krawczyk, C. M., Hutchins, A. S., Zediak, V. P., Banica, M., DiCioccio, C. B., Gross, D. A., Mao, C. A., et al. (2003). Control of effector CD8+ T cell function by the transcription factor Eomesodermin. *Science* **302**, 1041-1043.

Perea-Gomez, A., Shawlot, W., Sasaki, H., Behringer, R. R. and Ang, S. (1999). HNF3beta and Lim1 interact in the visceral endoderm to regulate primitive streak formation and anterior-posterior polarity in the mouse embryo. *Development* **126**, 4499-4511.

Pirrotta, V. and Li, H. B. (2012). A view of nuclear Polycomb bodies. *Curr Opin Genet Dev* **22**, 101-109.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283.

Robertson, E. J. (2014). Dose-dependent Nodal/Smad signals pattern the early mouse embryo. *Seminars in cell & developmental biology*.

Russ, A. P., Wattler, S., Colledge, W. H., Aparicio, S. A., Carlton, M. B., Pearce, J. J., Barton, S. C., Surani, M. A., Ryan, K., Nehls, M. C., et al. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95-99.

Sasaki, H. and Hogan, B. L. (1996). Enhancer analysis of the mouse HNF-3 beta gene: regulatory elements for node/notochord and floor plate are independent and consist of multiple sub-elements. *Genes to cells : devoted to molecular & cellular mechanisms* **1**, 59-72.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B. M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet* **47**, 1179-1186.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472.

Shawlot, W. and Behringer, R. R. (1995). Requirement for Lim1 in head-organizer function. *Nature* **374**, 425-430.

Stower, M. J. and Srinivas, S. (2014). Heading forwards: anterior visceral endoderm migration in patterning the mouse embryo. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**.

Teo, A. K., Arnold, S. J., Trotter, M. W., Brown, S., Ang, L. T., Chng, Z., Robertson, E. J., Dunn, N. R. and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev* **25**, 238-250.

van den Ameele, J., Tiberi, L., Bondue, A., Paulissen, C., Herpoel, A., Iacovino, M., Kyba, M., Blanpain, C. and Vanderhaeghen, P. (2012). Eomesodermin induces Mesp1 expression and cardiac differentiation from embryonic stem cells in the absence of Activin. *EMBO reports* **13**, 355-362.

Vincent, S. D., Dunn, N. R., Hayashi, S., Norris, D. P. and Robertson, E. J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes Dev* **17**, 1646-1662.

Wang, Q., Zou, Y., Nowotschin, S., Kim, S. Y., Li, Q. V., Soh, C. L., Su, J., Zhang, C., Shu, W., Xi, Q., et al. (2017). The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* **20**, 70-86.

Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R. S., Paquette, D., Dostie, J. and Bickmore, W. A. (2014). Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev* **28**, 2778-2791.

Williamson, I., Lettice, L. A., Hill, R. E. and Bickmore, W. A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143**, 2994-3001.

Yamamoto, M., Meno, C., Sakai, Y., Shiratori, H., Mochida, K., Ikawa, Y., Saijoh, Y. and Hamada, H. (2001). The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes Dev* **15**, 1242-1256.

Yoon, S. J., Foley, J. W. and Baker, J. C. (2015). HEB associates with PRC2 and SMAD2/3 to regulate developmental fates. *Nature communications* **6**, 6546.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B. D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355-364.

Zentner, G. E., Tesar, P. J. and Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **21**, 1273-1283.

Figures

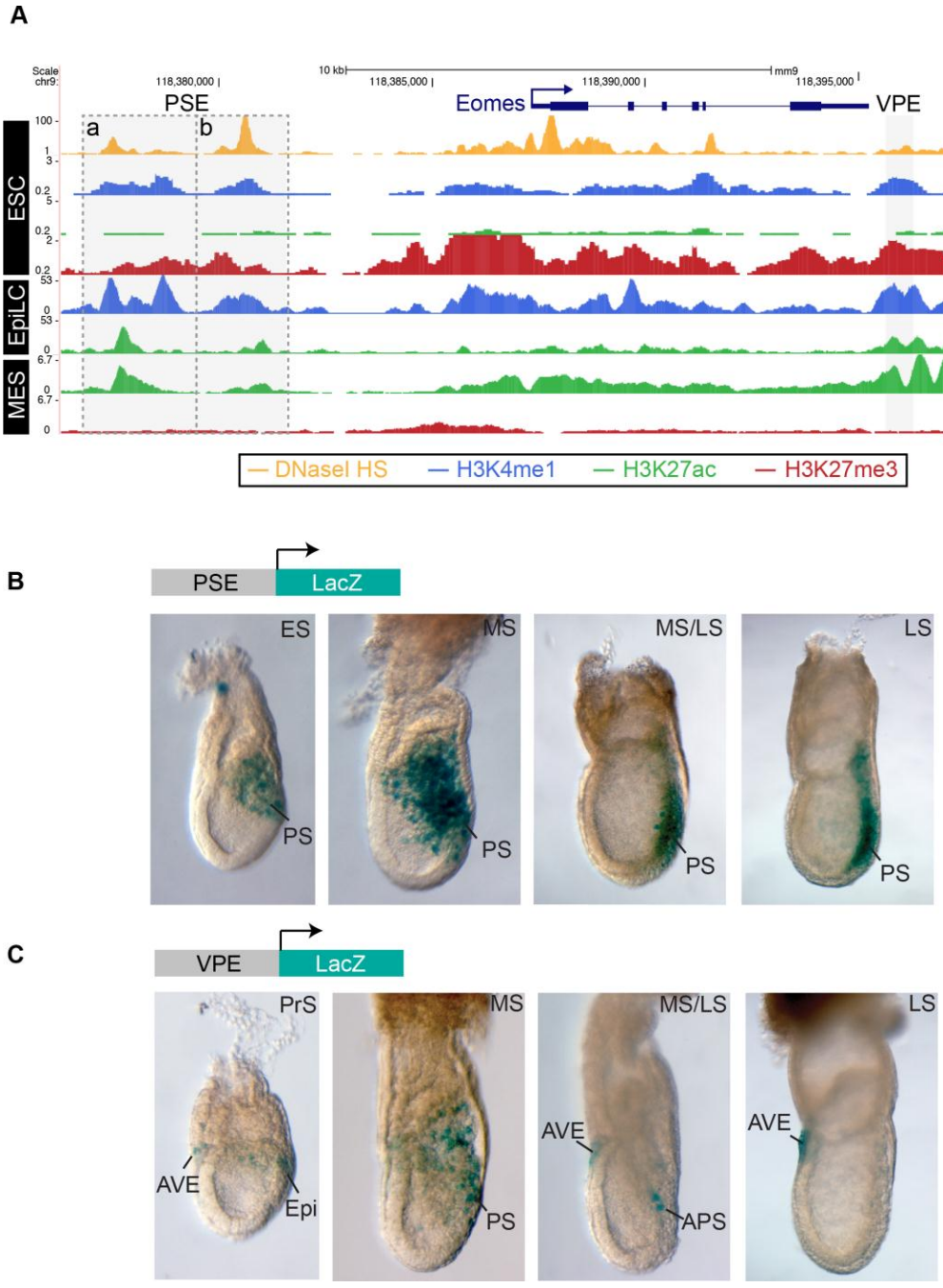


Figure 1. Mapping proximal *Eomes* enhancers active at gastrulation.

(A) ChIP-seq of H3K4me1, H3K27me3 and H3K27ac and DNaseI hypersensitivity (HS) in ESC, epiblast like cells (EpiLC) and mesoderm (MES) (Alexander et al., 2015; Buecker et al., 2014; Consortium, 2012) identify potential proximal *Eomes* enhancers activated during differentiation. The PSE cluster and VPE regions are highlighted in grey. (B-C) X-gal stained transgenic embryos expressing enhancer driven LacZ reporters. (B) PSE reporter activity is

confined to the primitive streak (PS) at early- (ES), mid- (MS) and late-streak (LS) stages of gastrulation (2/4 transgenic mouse lines). (C) VPE reporter activity detectable in the proximal posterior epiblast (Epi) at pre-streak (PrS) stage, the PS at MS stage, becomes restricted to the anterior PS (APS) and lost at LS stage. Between PrS stage and LS stage VPE activity is also detectable in the anterior visceral endoderm (AVE) (2/6 transgenic mouse lines).

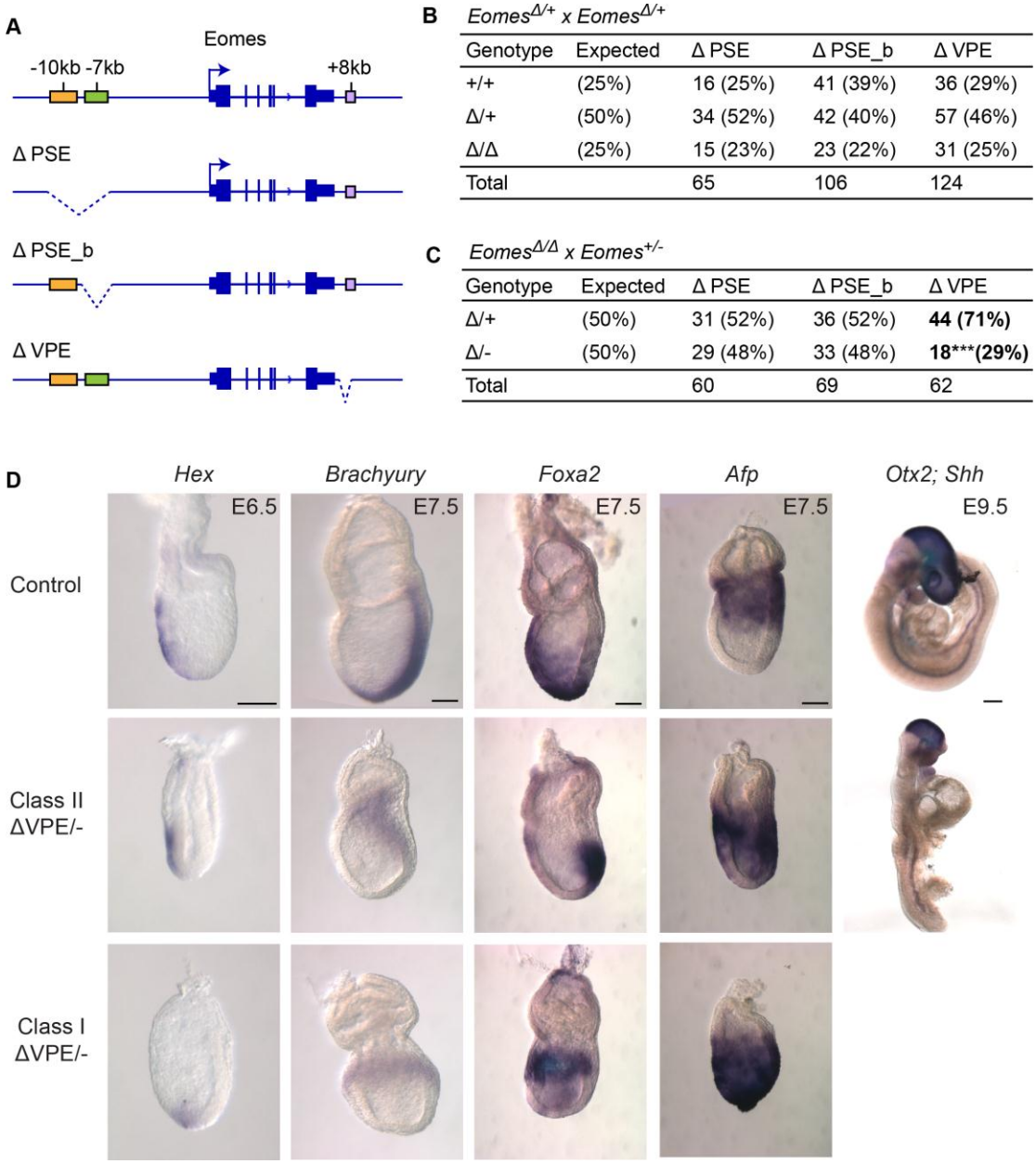


Figure 2. Targeted deletions of proximal enhancers show that only the VPE is required for proper gastrulation.

(A) Targeted deletions of the 5kb ΔPSE, 2kb ΔPSE_b, and 0.7kb ΔVPE generated by homologous recombination (Fig. S2-S4). (B) Genotypes of weanlings from heterozygous enhancer deletion intercrosses. Homozygous deletion of proximal enhancers does not affect viability (ΔPSE p=0.9, ΔPSE_b p=0.4, ΔVPE p=0.5, Chi-square test). (C) Genotypes of weanlings from heterozygous null and homozygous enhancer deletion matings. *Eomes*$^{\Delta VPE/-}$ animals are significantly underrepresented compared with littermate controls (ΔPSE p=0.87, ΔPSE_b p=0.7, ΔVPE ***p=0.001, Chi-square test). (D) Whole-mount *in-situ* hybridisation of *Eomes*$^{\Delta VPE/-}$ embryos. Class I mutants exhibit failure in A-P axis specification, and class II

display APS defects. At E6.5 in class I mutants expression of the AVE marker *Hex* is confined to the distal VE (n=4/10 $Eomes^{\Delta VPE/-}$ embryos analysed). At E7.5, the mesoderm marker *Brachyury* (n=2/5) and DE marker *Foxa2* (n=3/7) are mis-localised proximally. In class II mutants *Hex* marks the AVE, *Brachyury* expression fails to extend distally (n=3/5), while the *Foxa2* domain is confined to the APS and the DE domain is lost (n=3/7). Consistent with failure to specify DE in both mutant classes expression of *Afp*+ VE cells fails to disperse proximally (for Class I and Class II, n=2 and n=2 of 7 $Eomes^{\Delta VPE/-}$ embryos analysed, respectively). At E9.5 class II mutants display venture closure and neural tube defects, fused or malformed somites, loss of *Otx2*+ forebrain tissue, and an anterior truncation of the *Shh* midline (n=3/3 viable morphologically abnormal $Eomes^{\Delta VPE/-}$ embryos recovered). Scale bars: 100 μm.

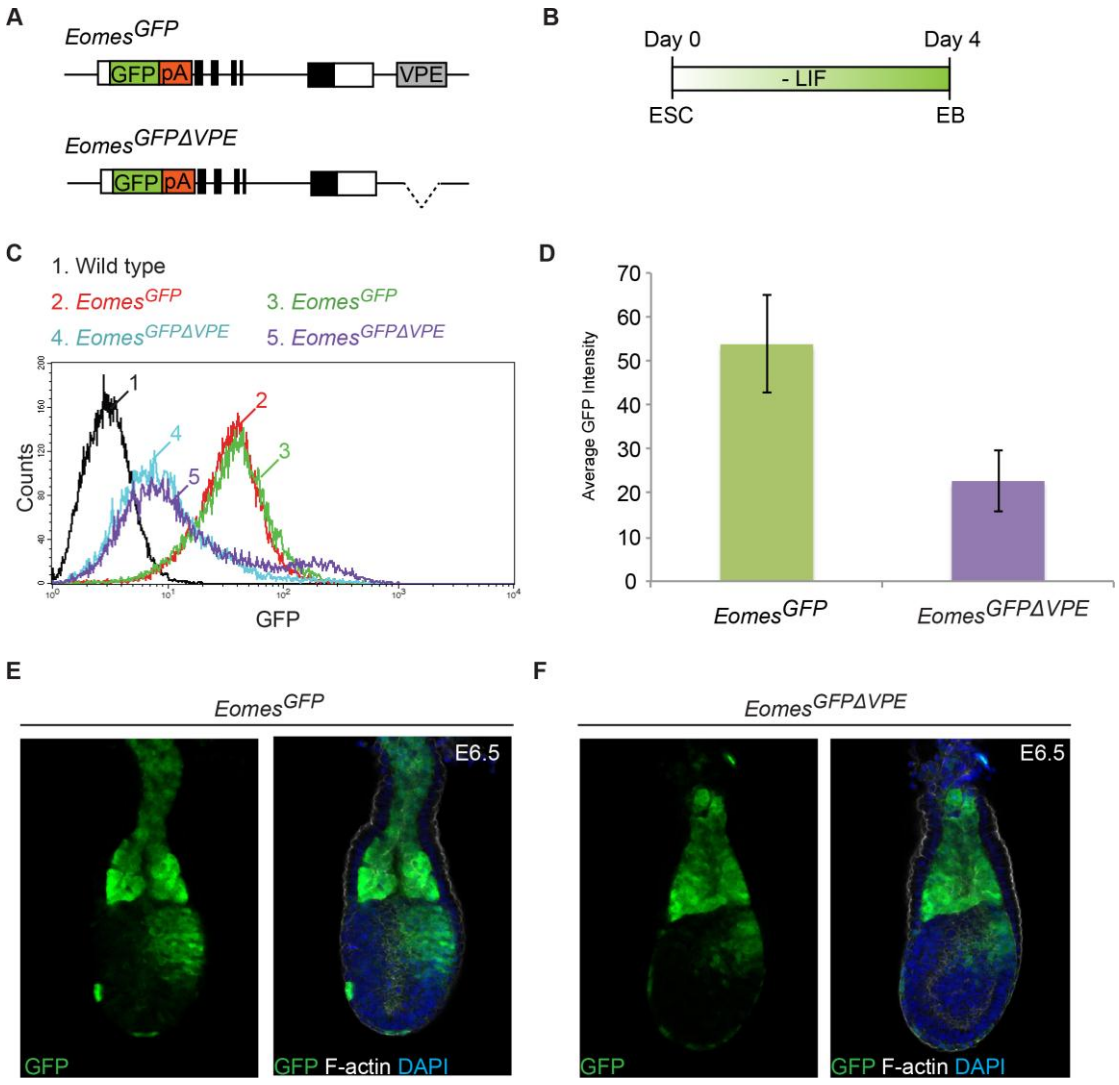


Figure 3. VPE deletion profoundly reduces the level of *Eomes*^{GFP} reporter expression.

(A) Configuration of the *Eomes*^{GFP} and *Eomes*$^{GFP\Delta VPE}$ alleles (Fig. S5). (B) Schematic of embryoid body (EB) differentiation protocol. (C-D) Flow cytometry analysis of wild type, *Eomes*$^{GFP/+}$ and *Eomes*$^{GFP\Delta VPE/+}$ day 4 EBs. (C) Representative histograms showing wild type, two independently targeted *Eomes*$^{GFP/+}$ and two *Eomes*$^{GFP\Delta VPE/+}$ clones. (D) Average GFP intensity in *Eomes*$^{GFP/+}$ (n=4) and *Eomes*$^{GFP\Delta VPE/+}$ (n=4) cultures. Deletion of the VPE significantly reduces expression to 42% of the intact *Eomes*^{GFP} reporter (p=0.05, Student's t-test). Error bars represent ± s.e.m. (E-F) Confocal images of *Eomes*^{GFP} and *Eomes*$^{GFP\Delta VPE}$ reporter expression in E6.5 embryos stained with anti-GFP antibody, DAPI (DNA) and Phalloidin (F-actin). Domains of reporter expression are not perturbed by VPE deletion.

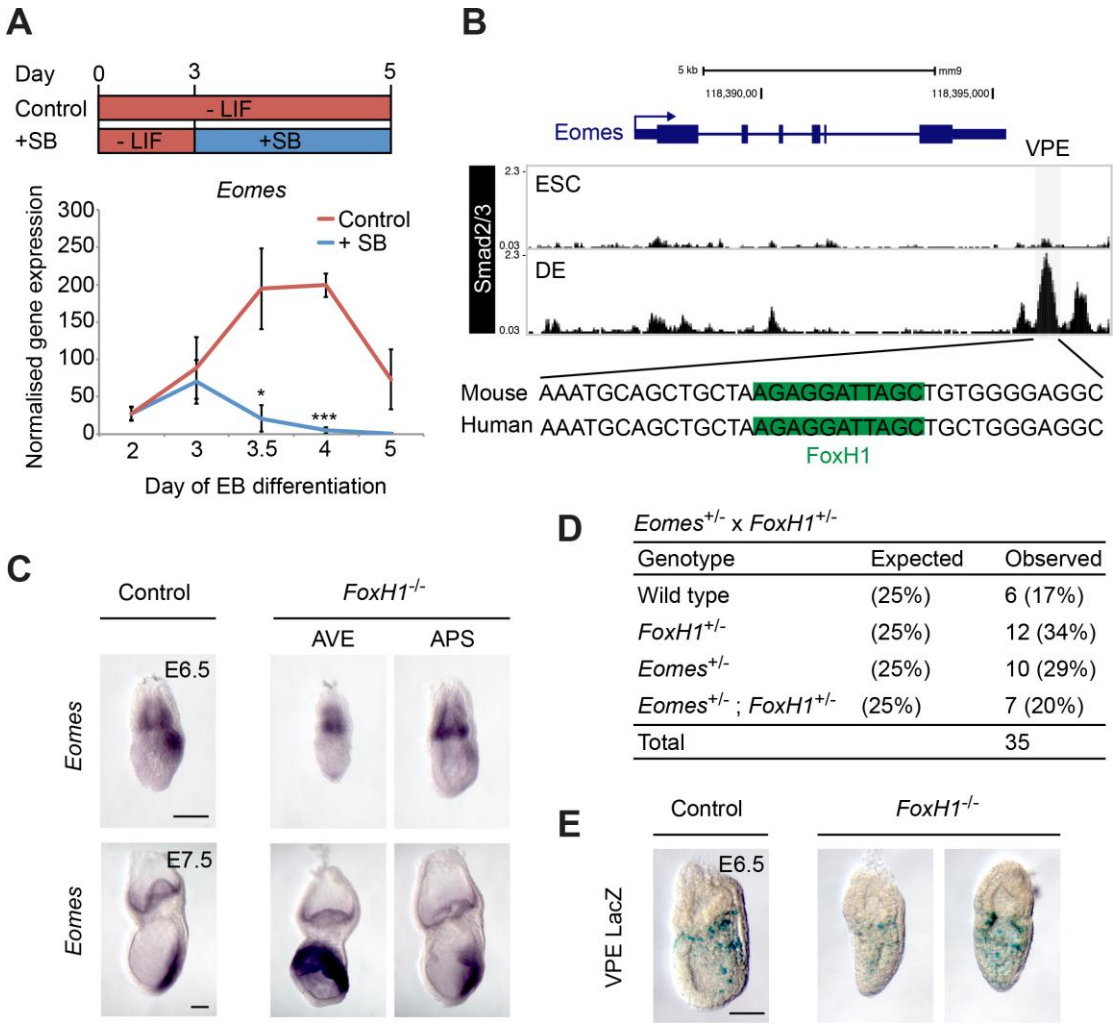


Figure 4. VPE expression is regulated by Smad2 and independently of FoxH1.

(A) RT-qPCR analysis of *Eomes* mRNA expression during EB differentiation. SB-431542 (SB) inhibition of Nodal/Smad2 signaling from day 3 onwards significantly reduces *Eomes* expression at d3.5 and d4 of differentiation (*p<0.05, ***p<0.001, Student's t-test, n=3). Error bars represent ±s.e.m. (B) ChIP-seq of Smad2/3 in definitive endoderm (DE) reveals binding to the VPE (Yoon et al., 2015), overlapping a predicted and conserved binding site for FoxH1, identified with JASPAR at >80% confidence (Mathelier et al., 2016). (C) Whole-mount *in situ* hybridisation of *Eomes* mRNA in control and *FoxH1* null embryos. *Eomes* is expressed in both AVE and APS defective FoxH1 mutant subtypes at E6.5 and E7.5. (D) Genotypes of weanlings from *Eomes* and *FoxH1* intercross matings. Compound heterozygotes do not show reduced viability (p=0.5, Chi Square test). (E) VPE-LacZ reporter activity both in the VE and epiblast is retained in *FoxH1* mutant embryos at E6.5. Scale bars: 100 μm.

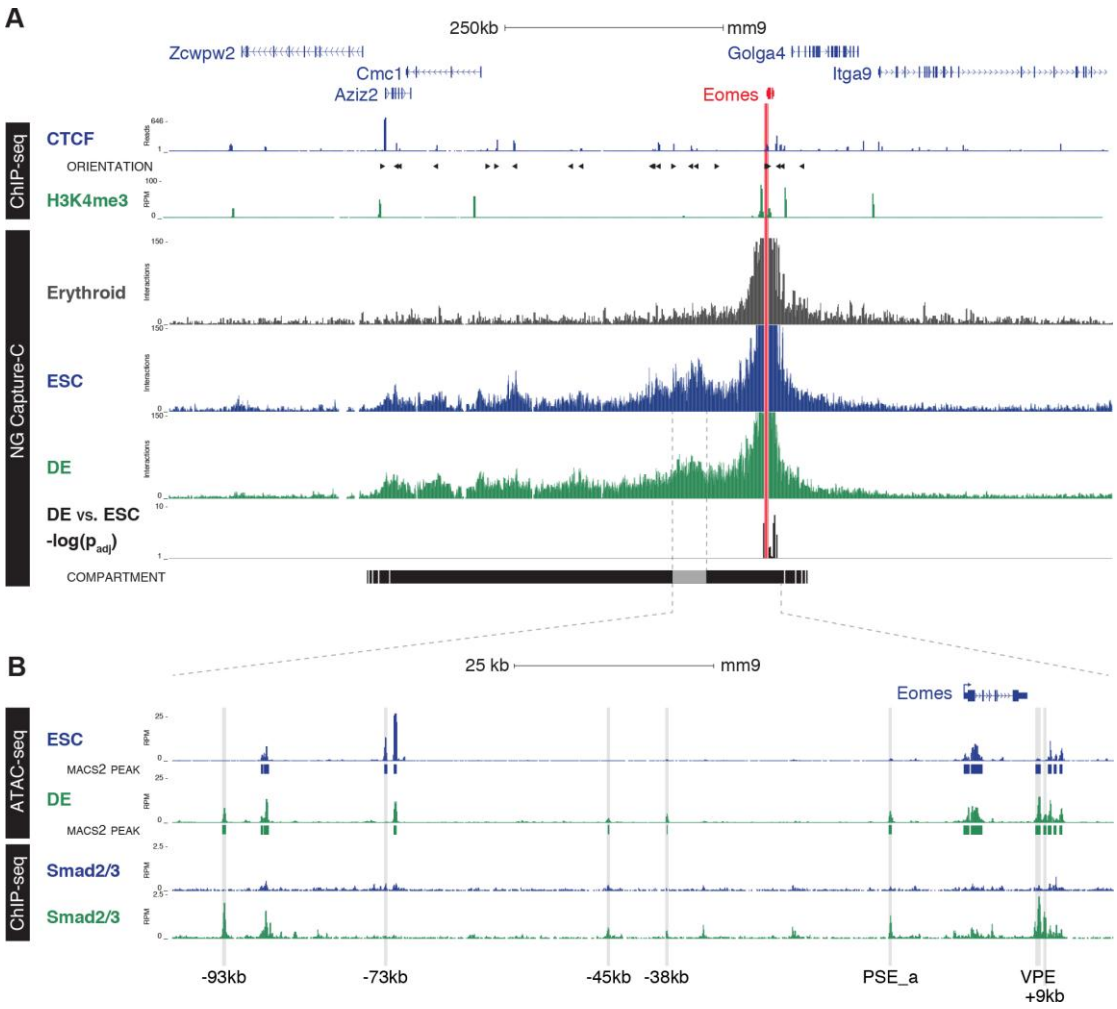


Figure 5. *Eomes* is regulated by Smad2/3 binding in a preformed compartment.

(A) NG Capture-C interaction profiles of the *Eomes* promoter (chr9:117,683,476-118,771,067) from erythrocytes (grey), ESC (blue) and DE (green). Tracks show mean interactions of normalized biological replicates (n=3), and DESeq2 significant differences between DE and ESC (-log(P_{adj}); p≤0.05). The *Eomes* compartment as determined by boundaries of strong promoter interactions with CTCF orientation (arrowheads) based upon binding in ESC (Handoko et al., 2011). Histone modifications for H3K4me3 (DE, n=3) shows promoter regions. (B) Zoom in panel of the region of the Eomes compartment showing highest association with the promoter, from chr9: 118,252,500-118,405,500. Open chromatin was generated using ATAC-seq in ESC and DE (n=3), with the addition of MACS2 called peaks annotated beneath each ATAC-seq track and Smad2/3 ChIP-seq in ESC (blue) and DE (green) (Yoon et al., 2015). Regions of chromatin accessibility unique to ESC (-73kb) and those associated with Smad2/3 occupancy in DE (-93kb, -45kb, -38kb, PSE_a, VPE and +9kb) are highlighted.

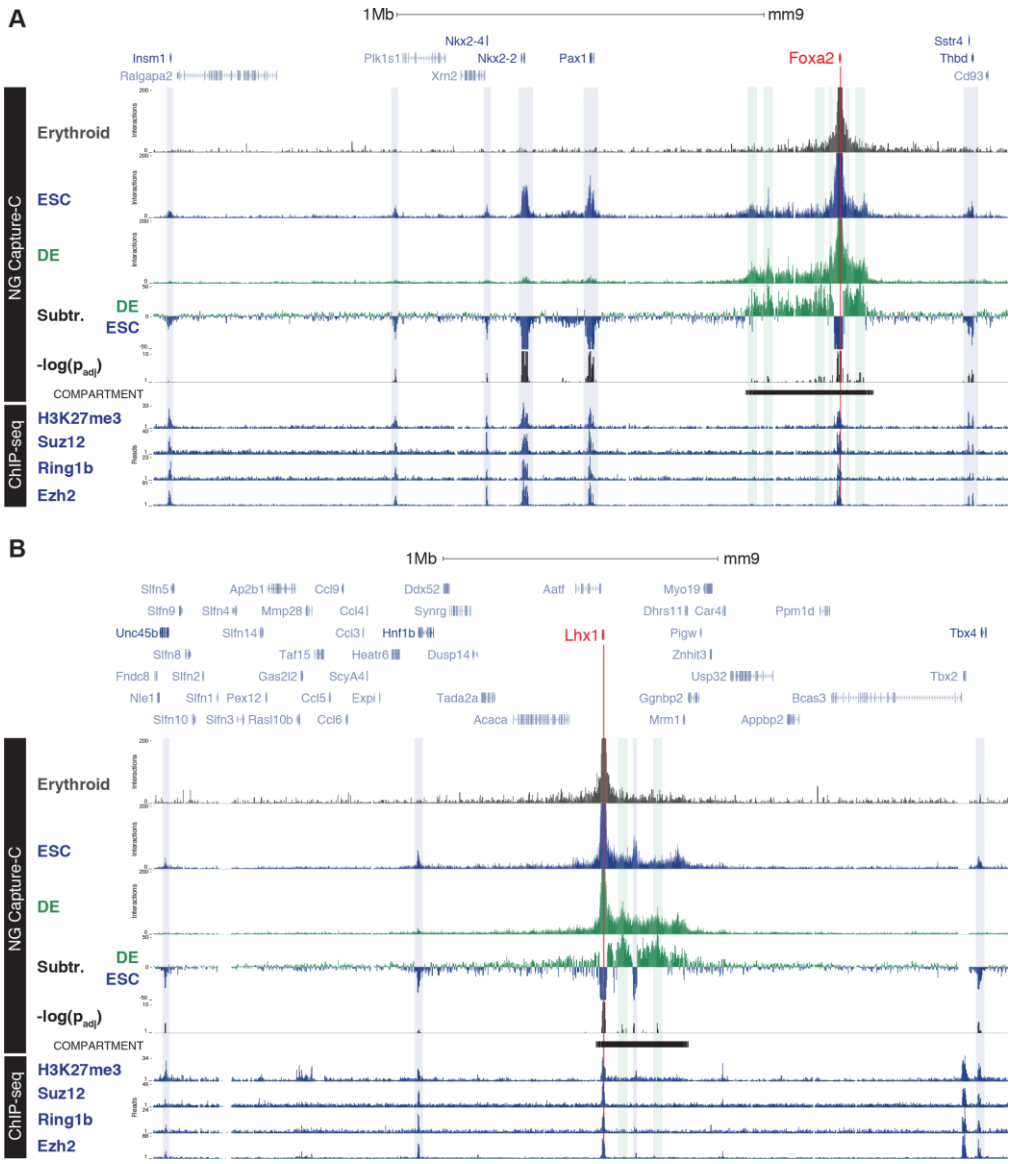


Figure 6. *Foxa2* and *Lhx1* form long-range interactions with polycomb repressed promoters.

NG Capture-C interaction profiles of the *Foxa2* (A) and *Lhx1* (B) promoters from erythrocytes (grey), ESC (blue) and DE (green) with chr2: 146,001,500-148,328,000 and chr11: 82,700,000-85,808,000 shown respectively. Tracks show mean interactions of normalized biological replicates (n=3), subtraction of ESC from DE (Subtr.) and DESeq2 significant differences between DE and ESC (-log(P_{adj}); $p \leq 0.05$). Peaks of strongest interactions in ESC (shaded boxes) were manually identified and highlighted. Compartments were determined by boundaries of strong (continuous) promoter interactions. Location of the Polycomb Repressor Complexes components (Ezh2, Suz12, Ring1b) and associated histone modification (H3K27me3) in ESC are shown (Ku et al., 2008; Mikkelsen et al., 2007).

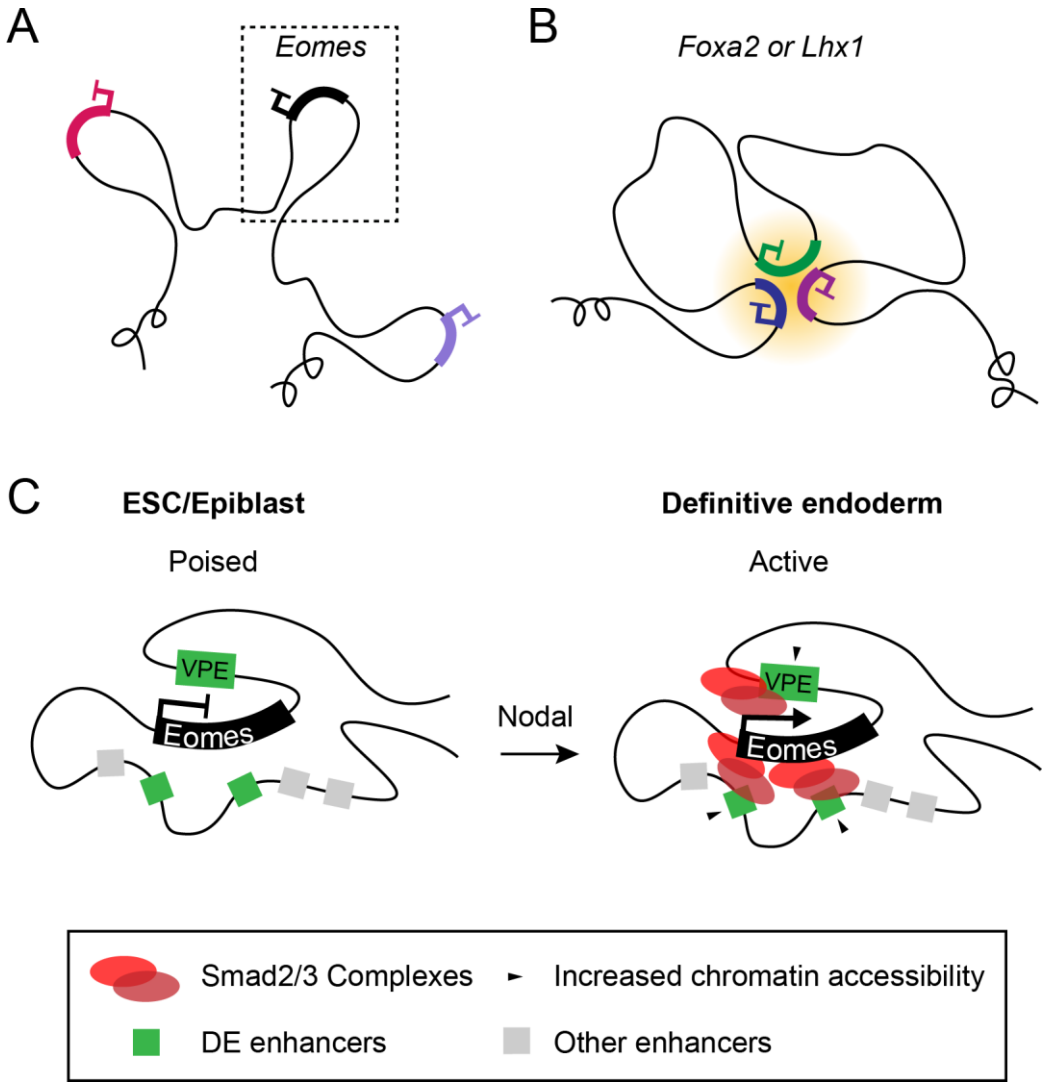


Figure 7. *Eomes*, *Foxa2*, and *Lhx1* exhibit distinct modes of 3D chromatin organisation during differentiation.

(A) In ESC *Eomes*, *Foxa2* and *Lhx1* are organised into pre-formed chromatin compartments. (B) Unlike *Eomes*, both *Foxa2* and *Lhx1* promoters form extra-compartmental contacts with other polycomb-repressed gene promoters. (C) Model for *Eomes* activation. The poised chromatin architecture at the *Eomes* locus is permissive for rapid transcriptional induction in response to localised Nodal signaling during gastrulation, primarily via enhancer binding of Smad2/3 complexes.

Supplemental Material

Supplementary Methods

Generation of targeted alleles

Targeting vectors containing 5' and 3' arms homologous to the *Eomes* locus, a FLP recognition target (FRT) flanked PGK.Neomycin selection cassette and a PGK.DTA (diphtheria toxin A) cassette for negative selection. The ΔVPE targeting vector was generated by recombineering using oligos listed in Table S1, designed to delete 656bp of the VPE. The ΔPSE_b vector includes a 5' 5.8kb SpeI-EcoRV fragment and a 3' 5kb KpnI-EcoRI fragment of the *Eomes* locus, and deletes 2019bp of PSE_b. The PSE vector comprises a 5' 5.6kb AatII-Bsu36I fragment, where the upstream AatII site was introduced by PCR (Table S1), and a 3' 5kb KpnI-EcoRI fragment of the *Eomes* locus, resulting in deletion of 4775bp of the PSE. XhoI (PSE, PSE_b) or ApaLI (VPE) linearized vectors (15ug) were electroporated into CCE ES cells, and $Eomes^{GFP/+}$ cells. Screening of drug resistant ESC clones was carried out by Southern blot analysis with the restriction enzymes and probes summarised in Fig. S2, S3, S4 and S5 using standard protocols (Behringer et al., 2013).

Supplemental References

Alexander, J. M., Hota, S. K., He, D., Thomas, S., Ho, L., Pennacchio, L. A. and Bruneau, B. G. (2015). Brg1 modulates enhancer activation in mesoderm lineage commitment. *Development* **142**, 1418-1430.

Arnold, S. J., Sugnaseelan, J., Groszer, M., Srinivas, S. and Robertson, E. J. (2009). Generation and analysis of a mouse line harboring GFP in the Eomes/Tbr2 locus. *Genesis* **47**, 775-781.

Behringer, R., Gertsenstein, M., Nagy, K. and Nagy, A. (2013). *Manipulating the mouse embryo : a laboratory manual, fourth edition* (Fourth edition. edn): Cold Spring Harbor Laboratory Press.

Brown, S., Teo, A., Pauklin, S., Hannan, N., Cho, C. H., Lim, B., Vardy, L., Dunn, N. R., Trotter, M., Pedersen, R., et al. (2011). Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* **29**, 1176-1185.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T. and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838-853.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

Costello, I., Nowotschin, S., Sun, X., Mould, A. W., Hadjantonakis, A. K., Bikoff, E. K. and Robertson, E. J. (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes Dev* **29**, 2108-2122.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W., Ye, C., Ping, J. L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43**, 630-638.

Kim, S. W., Yoon, S. J., Chuong, E., Oyolu, C., Wills, A. E., Gupta, R. and Baker, J. (2011). Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev Biol* **357**, 492-504.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242.

Lolas, M., Valenzuela, P. D., Tjian, R. and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. *Proc Natl Acad Sci U S A* **111**, 4478-4483.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560.

Teo, A. K., Arnold, S. J., Trotter, M. W., Brown, S., Ang, L. T., Chng, Z., Robertson, E. J., Dunn, N. R. and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev* **25**, 238-250.

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N. A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386-399.

Wang, Q., Zou, Y., Nowotschin, S., Kim, S. Y., Li, Q. V., Soh, C. L., Su, J., Zhang, C., Shu, W., Xi, Q., et al. (2017). The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* **20**, 70-86.

Supplemental Figures

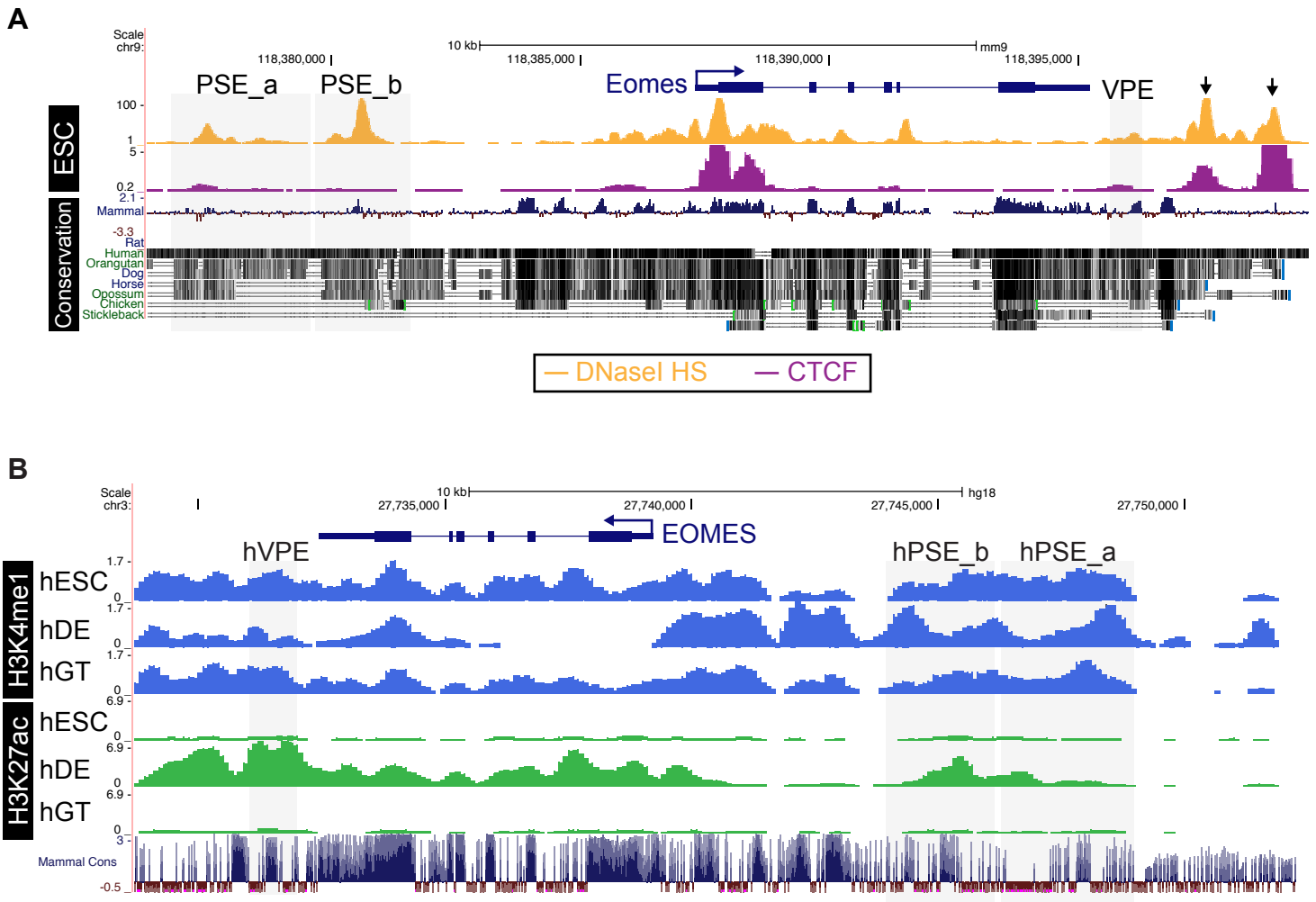


Figure S1: PSE and VPE enhancers are conserved in human

(A) DNaseI hypersensitivity (HS) and ChIP-seq of CTCF in ESC (Consortium, 2012). Conservation at the Eomes locus across vertebrates (UCSC browser, mm9). Boxes indicate PSE_a, PSE_b, and VPE enhancer regions, highly conserved amongst mammals. Arrows indicate CTCF bound regions downstream of the VPE. (B) ChIP-seq of H3K27ac and H3K4me1 histone modifications at the Eomes locus in human ESC (hESC), definitive endoderm (hDE) and human gut tube (hGT) (UCSC browser, hg18) (Wang et al., 2015). Homologous regions to the mouse VPE and PSE are associated with these active enhancer marks and are highlighted in grey. Human VPE, PSE_a and PSE_b (hVPE, hPSE_a, hPSE_b)

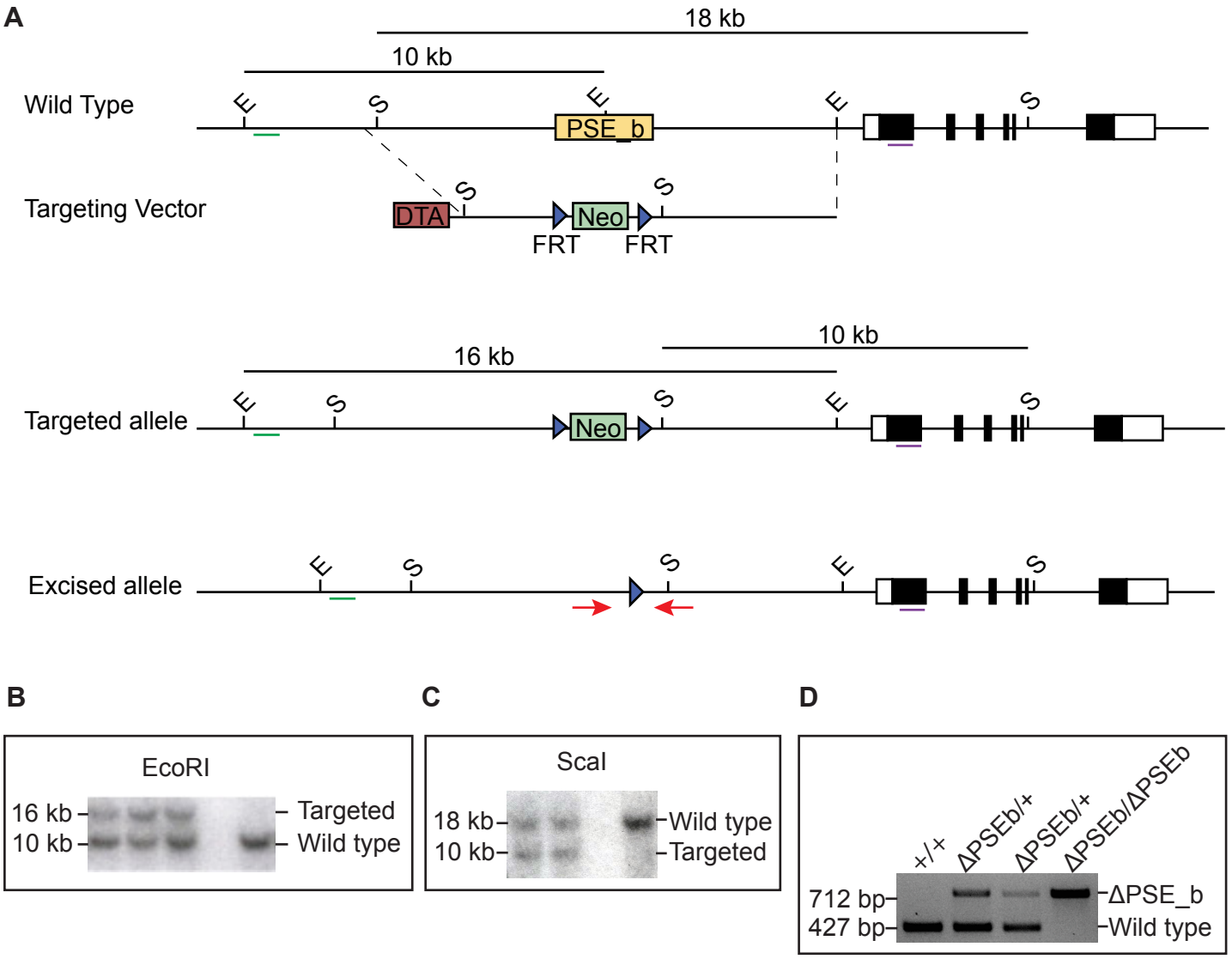


Figure S2: Targeted deletion of the PSE_b sub-region

(A) Targeting strategy to delete the 2kb PSE_b region (chr9:118379552-118381570; mm9) by homologous recombination. Southern blot restriction digest used for screening are indicated together with the probes (green and blue bars) and expected fragment sizes for the correctly targeted allele. EcoRI (E), ScaI (S), FLP-recombinase recognition site (FRT) site, Neomycin resistance cassette (Neo), Diphtheria toxin A cassette (DTA). Red arrows indicate primers for verifying FLP excision. (B,C) Southern blot of successfully targeted ESC clones. (D) PCR genotyping of $Eomes^{\Delta PSE\_b}$ mice.

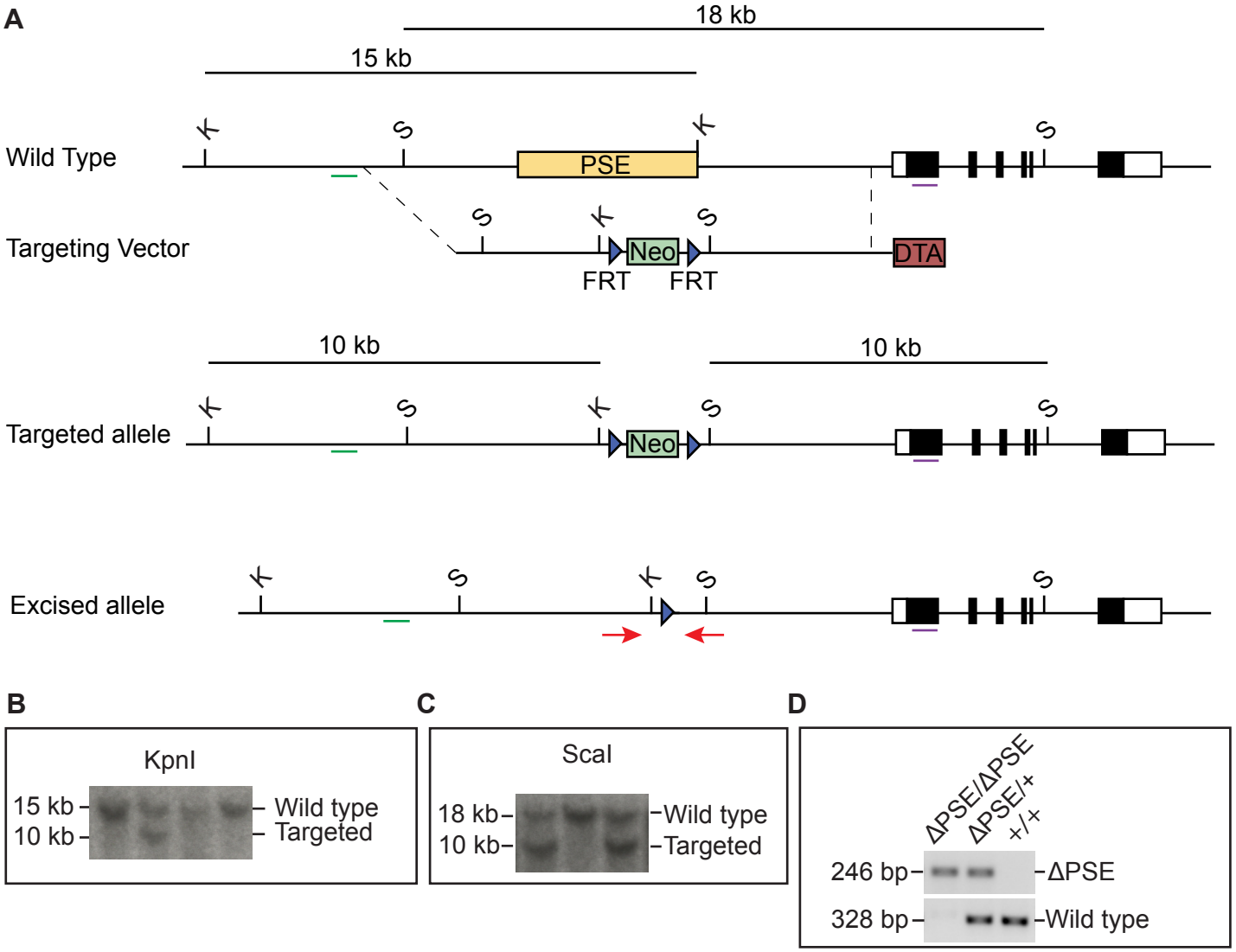
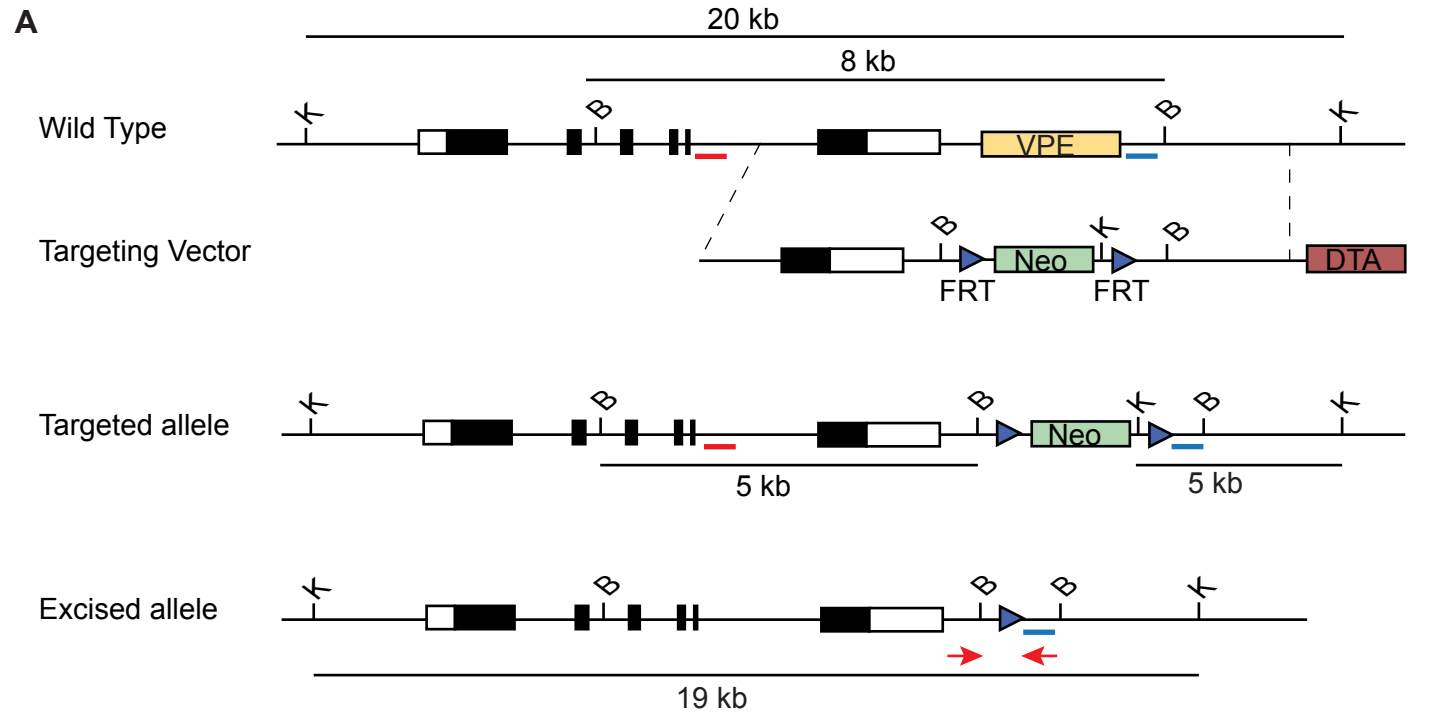
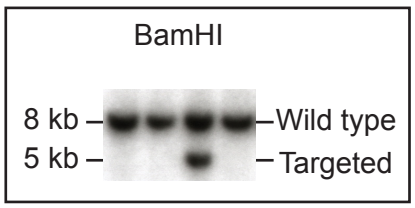
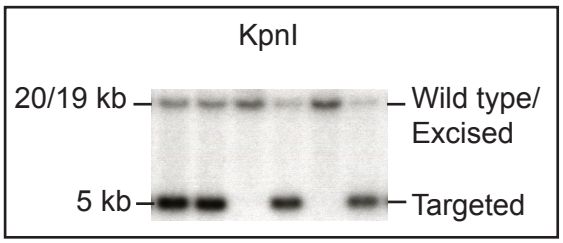


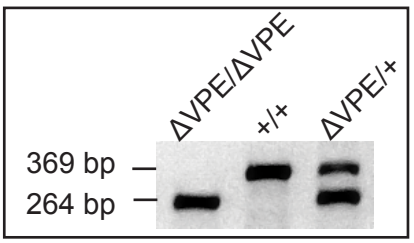
Figure S3: Targeted deletion of the PSE region

(A) Targeting strategy to delete the 5kb PSE region (chr9:118376796-118381570; mm9) by homologous recombination. Southern blot restriction digest used for screening are indicated together with the probes (green and blue bars) and expected fragment sizes for the correctly targeted allele. KpnI (K), ScaI (S), FLP-recombinase recognition site (FRT) site, Neomycin resistance cassette (Neo), Diphtheria toxin A cassette (DTA). Red arrows indicate primers for verifying FLP excision. (B,C) Southern blot of successfully targeted ESC clones. (D) PCR genotyping of Eomes$^{\Delta PSE}$ mice.









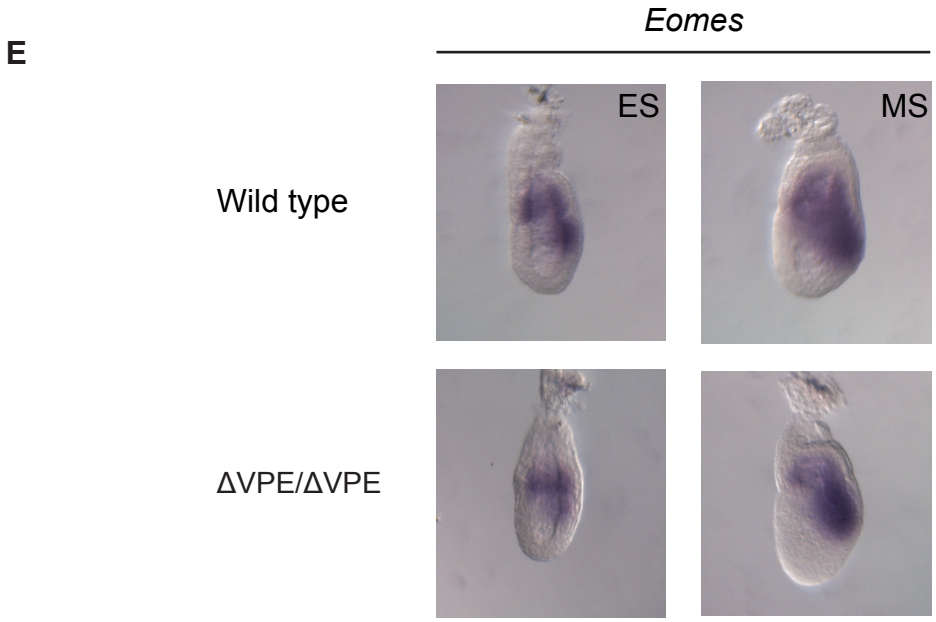


Figure S4: Targeted deletion of the VPE region

(A) Targeting strategy to delete the 0.7kb VPE region (chr9:118395625-118396280; mm9) by homologous recombination. Southern blot probes (red and blue bars), restriction digests and expected fragment sizes are indicated for the targeted and excised alleles. BamHI (B), KpnI (K), FLP-recombinase recognition site (FRT) site, Neomycin resistance cassette (Neo), Diphtheria toxin A cassette (DTA). Red arrows indicate primers for verifying FLP excision. (B) Southern blot of targeted ESC clones. (C) Southern blot to identify excision of Neo cassette in targeted ESC clones. (D) PCR genotyping ΔVPE allele in mice derived from $Eomes^{\Delta VPE/+}$ intercrosses. (E) Whole-mount *in situ* hybridisation of Eomes transcripts at early mid-streak stages shows Eomes expression domains are unaltered in $Eomes^{\Delta VPE/\Delta VPE}$ compared to wild type embryos.

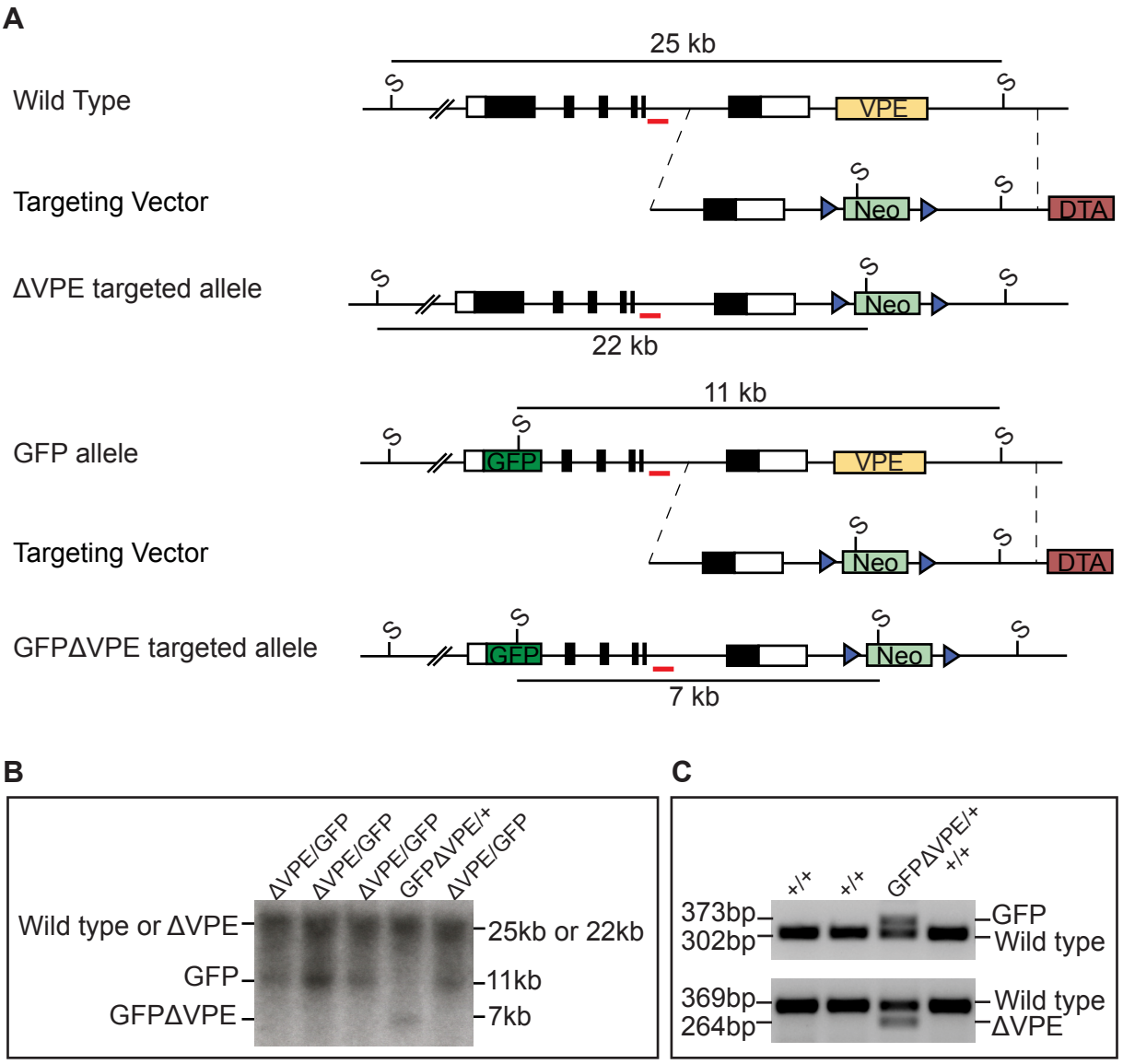


Figure S5: Generating Eomes^{GFP} allele lacking the VPE region

(A) Heterozygous *Eomes*$^{GFP/+}$ (Arnold et al., 2009) ESC were re-targeted using the same construct and primary screening strategy as used to delete the VPE. Southern blot strategy used to distinguish targeting the VPE region in either the GFP or wild type alleles, and expected fragment sizes are indicated. SpeI (S). (B) Southern blot showing two different genotypes of successfully targeted clones; *Eomes*$^{GFP\Delta VPE/+}$ and *Eomes*$^{GFP/\Delta VPE}$. (C) PCR genotyping of *Eomes*$^{GFP\Delta VPE/+}$ mice.

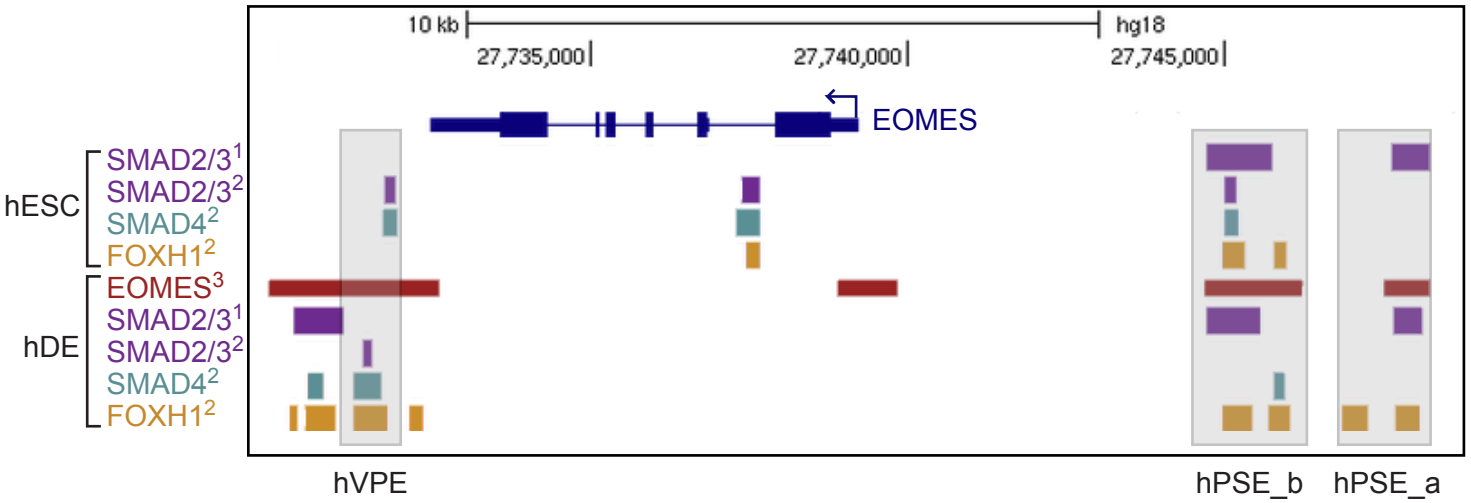


Figure S6: Regulation of the VPE by Nodal signaling

(A) Homologous human regions of the mouse VPE and PSE are bound by EOMES and mediators of the Nodal signaling pathway in hESCs and hDE. 1=(Brown et al., 2011) 2=(Kim et al., 2011), 3=(Teo et al., 2011). ChIP-seq data showing regions bound by SMAD2/3 (purple), SMAD4 (green), FOXH1 (orange) and EOMES (red) are represented by coloured bars and were aligned to the EOMES locus on the UCSC Genome browser Human Mar. 2006 (NCBI36/hg18) Assembly (http://genome.ucsc.edu/). Homologous regions to the mouse VPE and PSE are highlighted in grey. Human VPE, PSE_a and PSE_b (hVPE, hPSE_a, hPSE_b). FoxH1 binds the conserved FoxH1 binding site at the VPE in hDE.

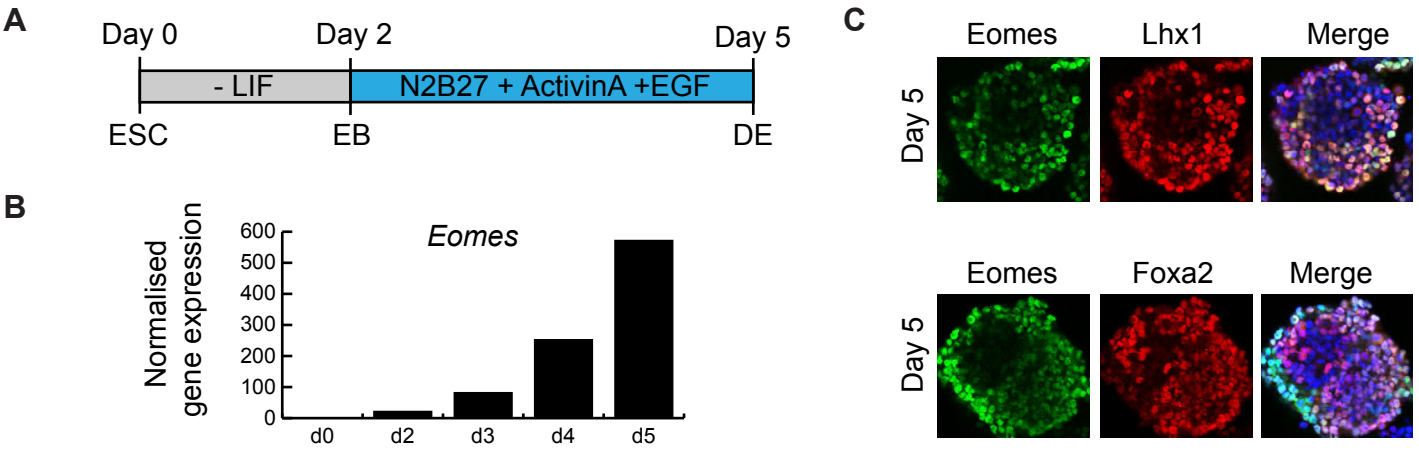


Figure S7: Definitive endoderm differentiation

(A) Schematic of protocol to differentiate ESC to definitive endoderm (DE) fate. ESC were grown in the absence of LIF for 2 days to form embryoid bodies (EB) and then differentiated in N2B27 medium, 20ng/ml ActivinA and 20ng/ml EGF for a further 3 days. (B) qPCR of *Eomes* mRNA demonstrates a dramatic increase in expression over the course of the 5 day differentiation regime. Gene expression is normalised to *Gapdh*. (C) 2D confocal images of d5 DE EBs stained with antibodies against definitive endoderm markers Eomes, Lhx1 or Foxa2, and counterstained with DAPI.

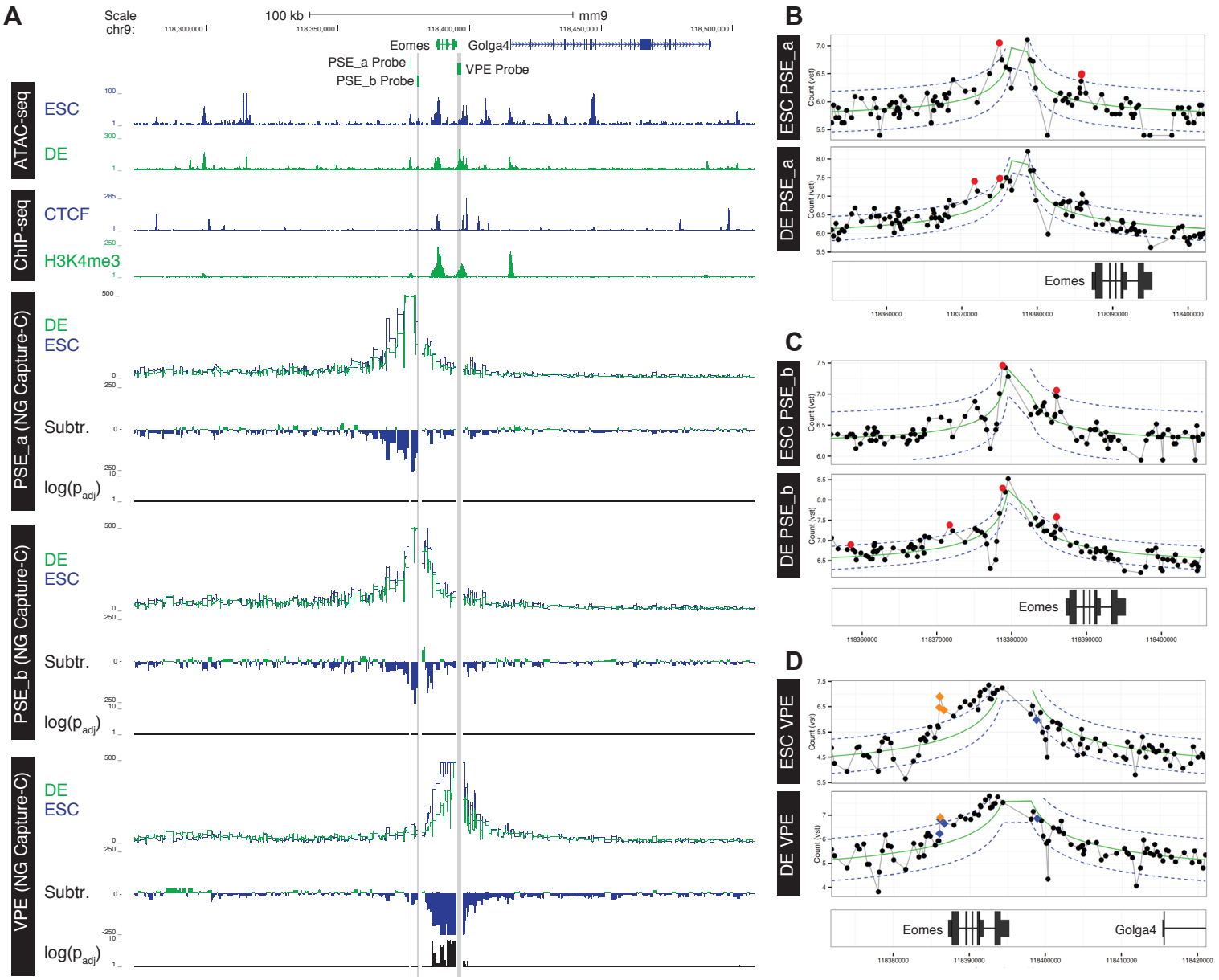


Figure S8: NG Capture-C from the *Eomes* enhancers.

(A) NG Capture-C interaction profiles of the PSE_a, PSE_b, and VPE from ESC (blue) and DE (green). Tracks show mean interactions of normalized biological replicates (n=3), subtraction of ESC from DE (Subtr.) and DESeq2 significant differences between DE and ESC (-log(P_{adj}); p≤0.05). Open chromatin was determined by ATAC-seq in both ESC and DE, ChIP-seq of the boundary element CTCF in ESC is from published data (Handoko et al., 2011) and H3K4me3 ChIP-seq was generated in triplicate from DE. FourCSeq comparison of NG Capture-C between DE and ESC from the PSE_a (B), PSE_b (C), and VPE (D). Red circles mark fragments with more interactions than expected based upon proximity to the promoter (green line), Blue Diamonds show fragments with significantly different interactions between the two conditions (P≤0.05), Orange Diamonds show fragments with enriched reactions that are significantly different between the two conditions.

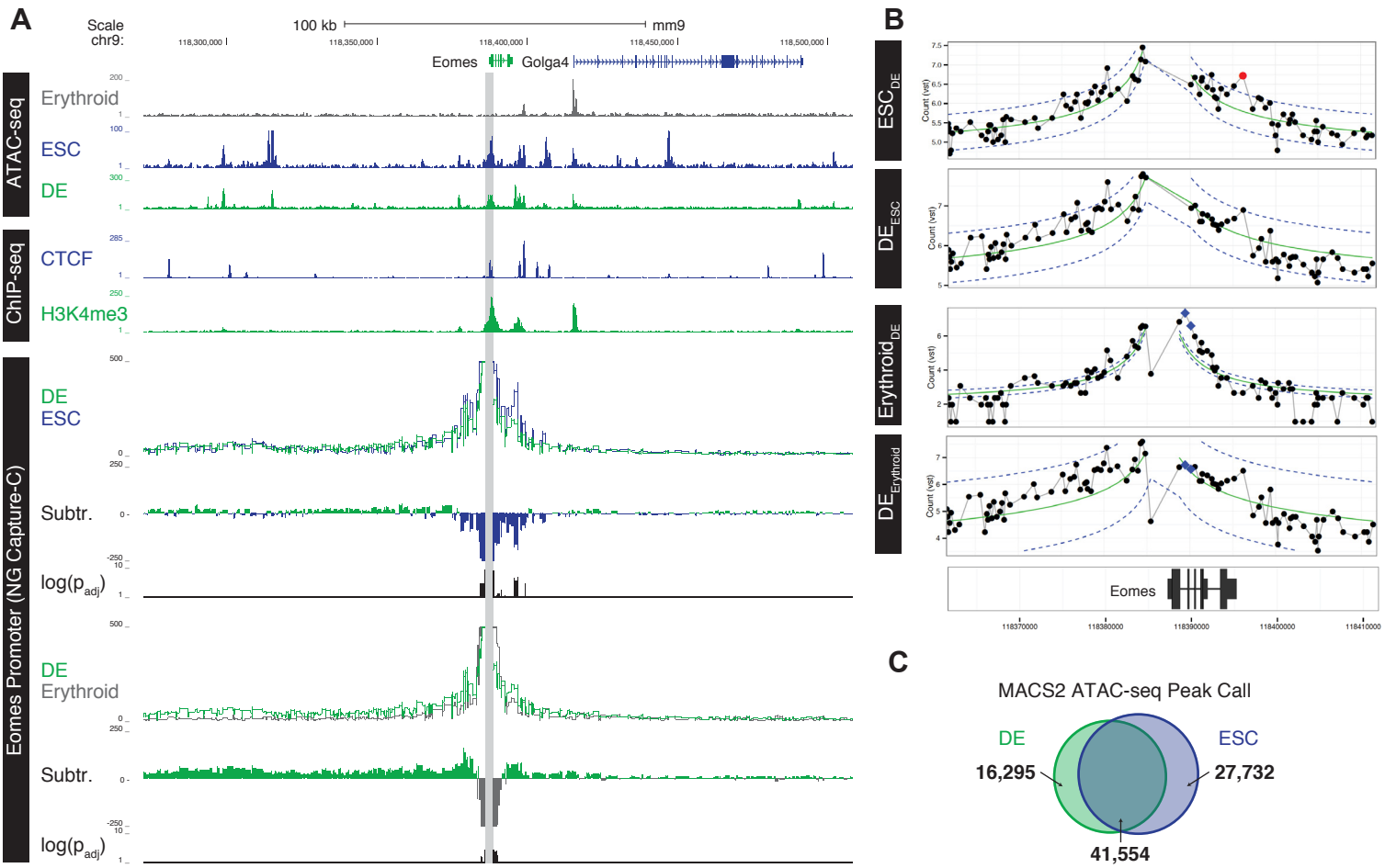


Figure S9: NG Capture-C from the *Eomes* promoter.

(A) NG Capture-C interaction profiles of the *Eomes* promoter from terminally differentiated erythrocytes (Ery, grey), ESC (blue) and DE (green). Tracks show mean interactions of normalized biological replicates (n=3), subtraction of ESC and PHS from DE (Subtr.) and DESeq2 significant differences between the cell types (-log(P_{adj}); p≤0.05). Open chromatin was determined by ATAC-seq in all three cell types (n=3), ChIP-seq of the boundary element CTCF in ESC is from published data (Handoko et al., 2011) and H3K4me3 ChIP-seq was generated in triplicate from DE. (B) FourCSeq comparison of NG Capture-C of the *Eomes* promoter between DE, ESC and Ery. Comparison condition is shown in subscript. Red circles mark fragments with more interactions than expected based upon proximity to the promoter (green line), Blue Diamonds show fragments with significantly different interactions between the two conditions (P≤0.05), Orange Diamonds show fragments with enriched reactions that are significantly different between the two conditions. (C) Comparison of MACS2 peak call for ATAC-seq from DE and ESC.

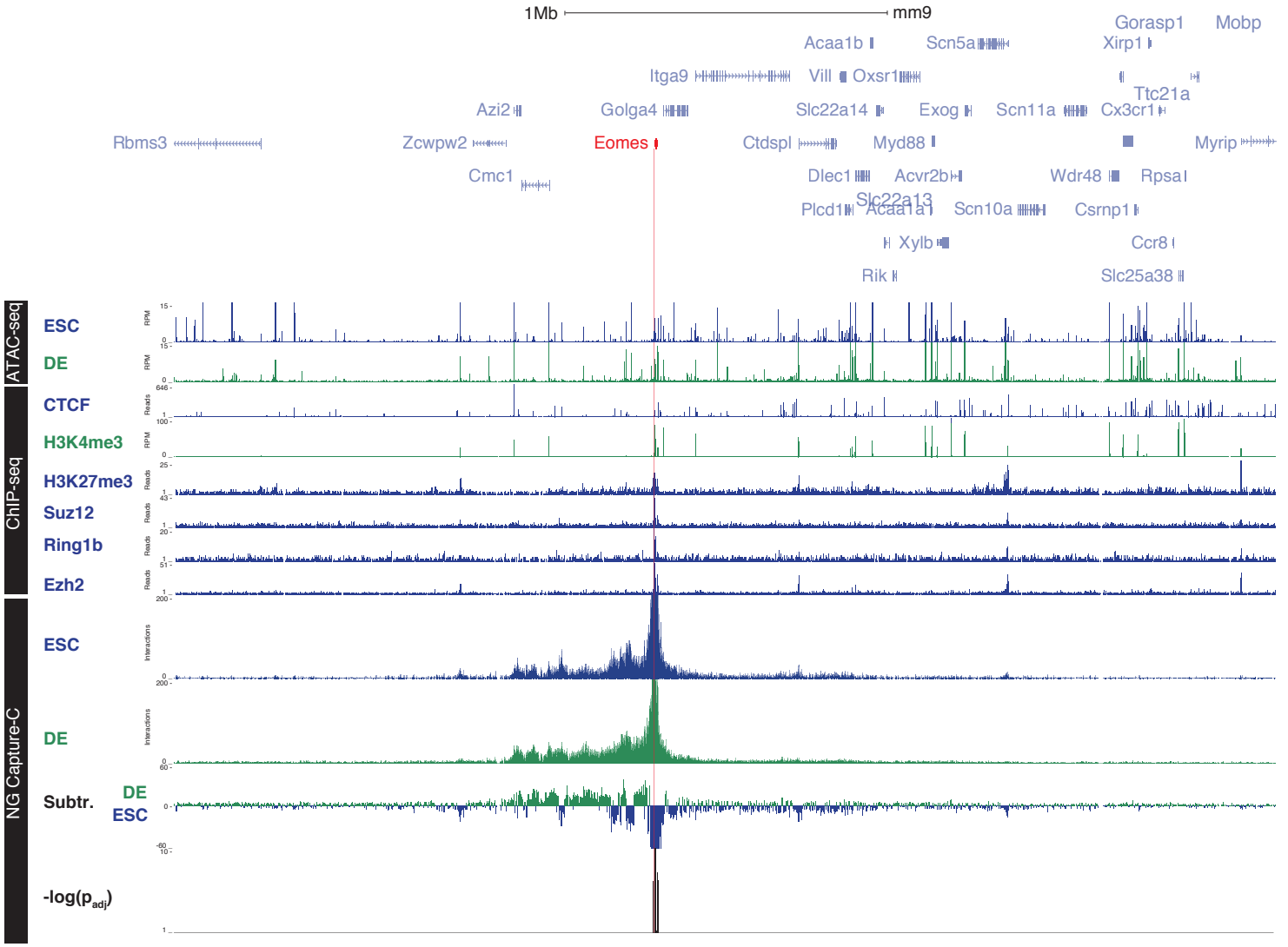


Figure S10: Long-range NG Capture-C from the *Eomes* promoter.

NG Capture-C interaction profiles of the *Eomes* promoter (chr9:116890604-120321539) from erythrocytes (grey), ESC (blue) and DE (green). Tracks show mean interactions of normalized biological replicates (n=3), subtraction of ESC from DE (Subtr.) and DESeq2 significant differences between DE and ESC (-log(P_{adj}); p≤0.05). Location of the Polycomb Repressor Complexes components (Ezh2, Suz12, Ring1b) and associated histone modification (H3K27me3) in ESC are shown (Ku et al., 2008; Mikkelsen et al., 2007). Open chromatin was determined by ATAC-seq in all three cell types (n=3), ChIP-seq of the boundary element CTCF in ESC is from published data (Handoko et al., 2011) and H3K4me3 ChIP-seq was generated in triplicate from DE.

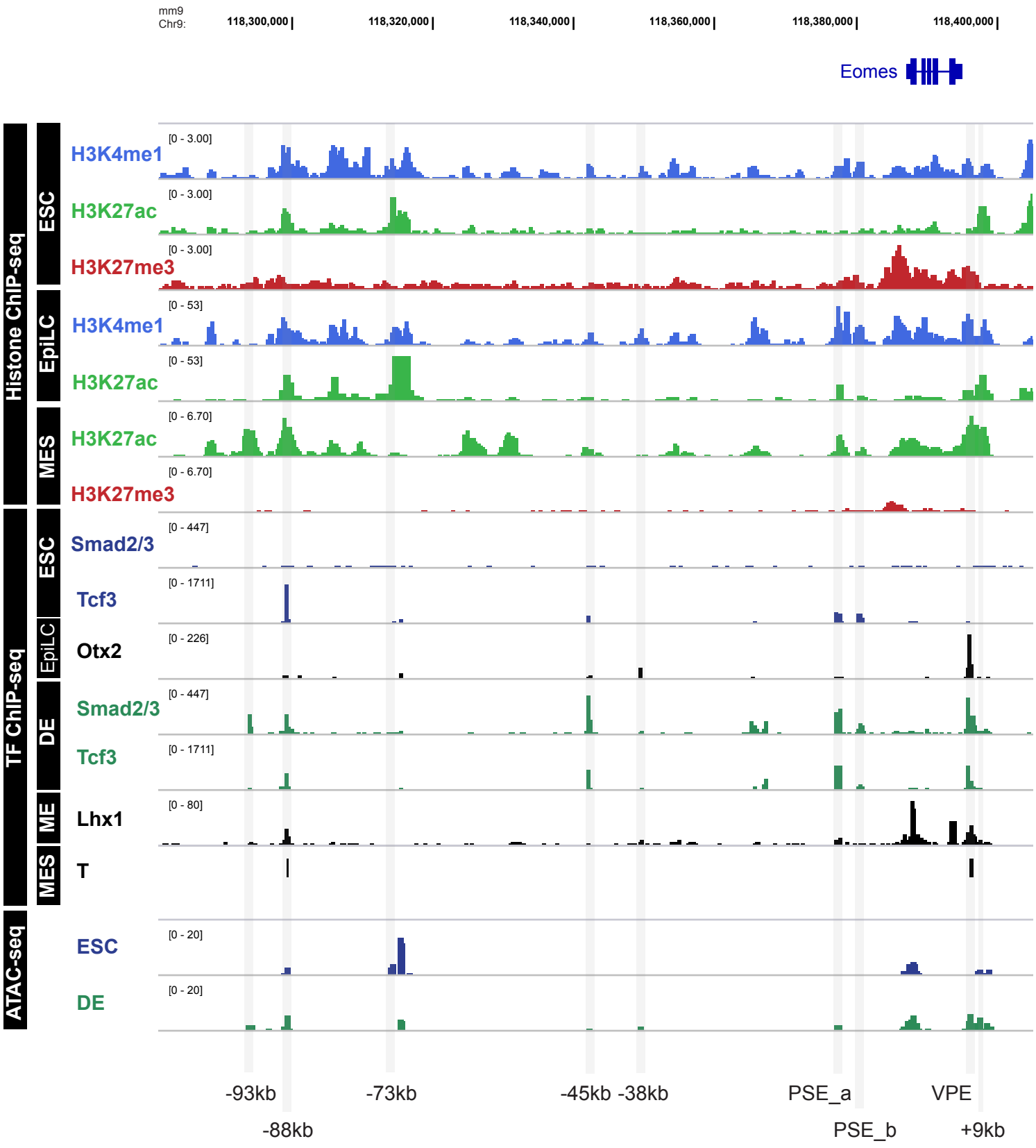


Figure S11: Mapping enhancers within the *Eomes* compartment.

ChIP-seq of histone modifications H3K4me1 (light blue), H3K27me3 (red) and H3K27ac (light green) in ESC, epiblast like cells (EpiLC) and mesoderm (MES) (Alexander et al., 2015; Buecker et al., 2014; Consortium, 2012). Open chromatin was generated using ATAC-seq in ESC and DE (n=3). ChIP-seq of TFs involved in endoderm and anterior mesendoderm specification. Smad2/3 and Tcf3 in ESC (blue) and DE (green) (Wang et al., 2017). Otx2 in EpiLC (Buecker et al., 2014), Lhx1 in P19 mesendoderm (ME) (Costello et al., 2015), and Brachyury (T) (Lolas et al., 2014) in MES. Regions of increased chromatin accessibility unique to ESC (-73kb) and those associated with Smad2/3 occupancy uniquely in DE (-93kb, -45kb, -38kb, PSE_a, VPE and +9kb) are highlighted as in Fig. 5B. In addition, a TF binding hotspot accessible in both ESC and DE (-88kb), and the PSE_b, are also highlighted.

Table S1: Primers used in this study

Primer name	Forward sequence	Reverse sequence	Product
Targeting vectors			
VPE Recombineering	GGCTGGGGTTGGG GAAGGAGTGTTTGC CCTGGAGATGCAAG ATTGTGCTCGGATC CAATTAACCCTCAC TAAAGGGC	GGTCCCAGAAGTTTG GAGGACGGGAAAGA CTGTCCACAGCTCAG GTATATCGAAGTTAT AAGCTTGAAGTTCCT ATACTTTC	n/a
PSE AatII	TGACGTCTGTGTTC AAAAGCACGAGGG	ACCAGAGACCGTATG TTCCC	2.7kb
Transgenic reporter			
VPE LacZ	GCCCTGGAGATGC AAGATTG	CAGCTCAGGTATATC TTCTGGC	696bp
Genotyping			
VPE WT	TCGTTGAGTGGTGA GCAGGGAG	AGCGAGGACATCCA CGGAAAAC	369bp
VPE Δ	TCGTTGAGTGGTGA GCAGGGAG	TTTGGAGGACGGGA AAGACTG	264bp
PSE WT	AGGGTGGCTCTATA CAGGTG	GCATTGGAGTTGAAG GTGGG	328bp
PSE Δ	AGGGTGGCTCTATA CAGGTG	TCACAAGTCTCTCCT GGCAC	246bp
PSE_b WT	TTGCGTTTGTTGGG TTTTGG	CCATCACTGGGAGA GTAGGC	427bp
PSE_b Δ	GGCTATTGCCTCCA TACAGC	CCATCACTGGGAGA GTAGGC	712bp
LacZ	TTACCAGGCCGAAG CAGCGTTGTTG	GCGGCAGTAAGGCG GTCGGGATAGT	300bp
RT-PCR			
Gapdh	CAATGACCCCTTCA TTGACC	GATCTCGCTCCTGGA AGATG	145bp
Eomes	TGTTTTCGTGGAAG TGGTTCTGGC	AGGTCTGAGTCTTGG AAGGTTCATTC	323bp

Table S2: Antibodies used in this study

Name	Catalog number	Company
Foxa2	sc-6554	Santa Cruz
Lhx1	sc-19341	Santa Cruz
TBR2/Eomes	ab23345	Abcam
GFP AlexaFluor 488	A21311	Invitrogen
Goat IgG AlexaFluor 594	A11058	Invitrogen
Rabbit IgG AlexaFluor 488	A21206	Invitrogen
Anti-H3K4me3	07-473	Millipore

Table S3. Long-range *Foxa2* and *Lhx1* promoter interactions identified by NG Capture-C

Click here to Download Table S3

Table S4. Probes used for NG Capture-C.

Click here to Download Table S4

Table S5. Accession codes used in this study.

Click here to Download Table S5

Supplemental Material

Supplementary Methods

Generation of targeted alleles

Targeting vectors containing 5' and 3' arms homologous to the *Eomes* locus, a FLP recognition target (FRT) flanked PGK.Neomycin selection cassette and a PGK.DTA (diphtheria toxin A) cassette for negative selection. The ΔVPE targeting vector was generated by recombineering using oligos listed in Table S1, designed to delete 656bp of the VPE. The ΔPSE_b vector includes a 5' 5.8kb SpeI-EcoRV fragment and a 3' 5kb KpnI-EcoRI fragment of the *Eomes* locus, and deletes 2019bp of PSE_b. The PSE vector comprises a 5' 5.6kb AatII-Bsu36I fragment, where the upstream AatII site was introduced by PCR (Table S1), and a 3' 5kb KpnI-EcoRI fragment of the *Eomes* locus, resulting in deletion of 4775bp of the PSE. XhoI (PSE, PSE_b) or ApaLI (VPE) linearized vectors (15ug) were electroporated into CCE ES cells, and $Eomes^{GFP/+}$ cells. Screening of drug resistant ESC clones was carried out by Southern blot analysis with the restriction enzymes and probes summarised in Fig. S2, S3, S4 and S5 using standard protocols (Behringer et al., 2013).

Supplemental Figures

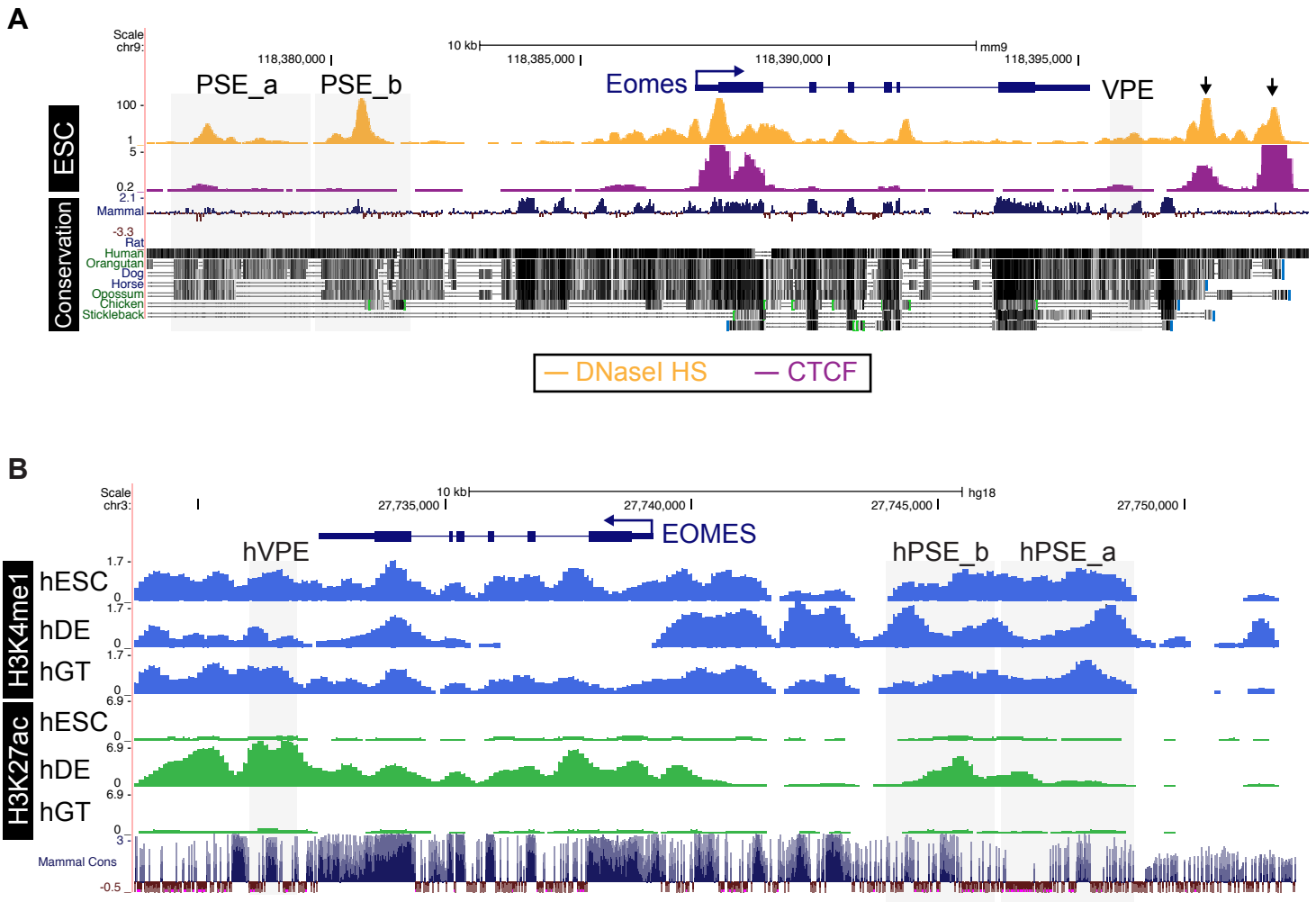


Figure S1: PSE and VPE enhancers are conserved in human

(A) DNaseI hypersensitivity (HS) and ChIP-seq of CTCF in ESC (Consortium, 2012). Conservation at the Eomes locus across vertebrates (UCSC browser, mm9). Boxes indicate PSE_a, PSE_b, and VPE enhancer regions, highly conserved amongst mammals. Arrows indicate CTCF bound regions downstream of the VPE. (B) ChIP-seq of H3K27ac and H3K4me1 histone modifications at the Eomes locus in human ESC (hESC), definitive endoderm (hDE) and human gut tube (hGT) (UCSC browser, hg18) (Wang et al., 2015). Homologous regions to the mouse VPE and PSE are associated with these active enhancer marks and are highlighted in grey. Human VPE, PSE_a and PSE_b (hVPE, hPSE_a, hPSE_b)

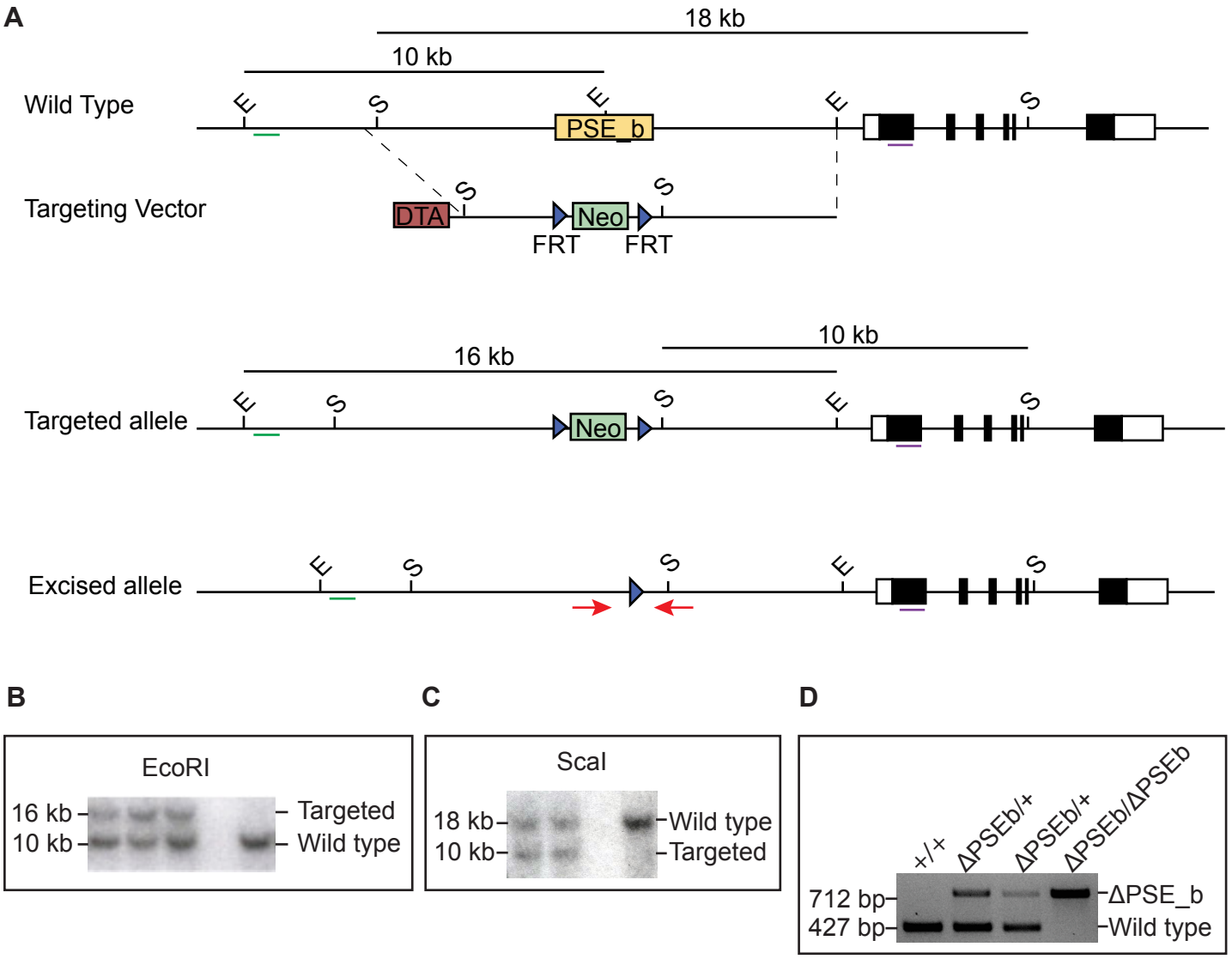


Figure S2: Targeted deletion of the PSE_b sub-region

(A) Targeting strategy to delete the 2kb PSE_b region (chr9:118379552-118381570; mm9) by homologous recombination. Southern blot restriction digest used for screening are indicated together with the probes (green and blue bars) and expected fragment sizes for the correctly targeted allele. EcoRI (E), ScaI (S), FLP-recombinase recognition site (FRT) site, Neomycin resistance cassette (Neo), Diphtheria toxin A cassette (DTA). Red arrows indicate primers for verifying FLP excision. (B,C) Southern blot of successfully targeted ESC clones. (D) PCR genotyping of $Eomes^{\Delta PSE\_b}$ mice.

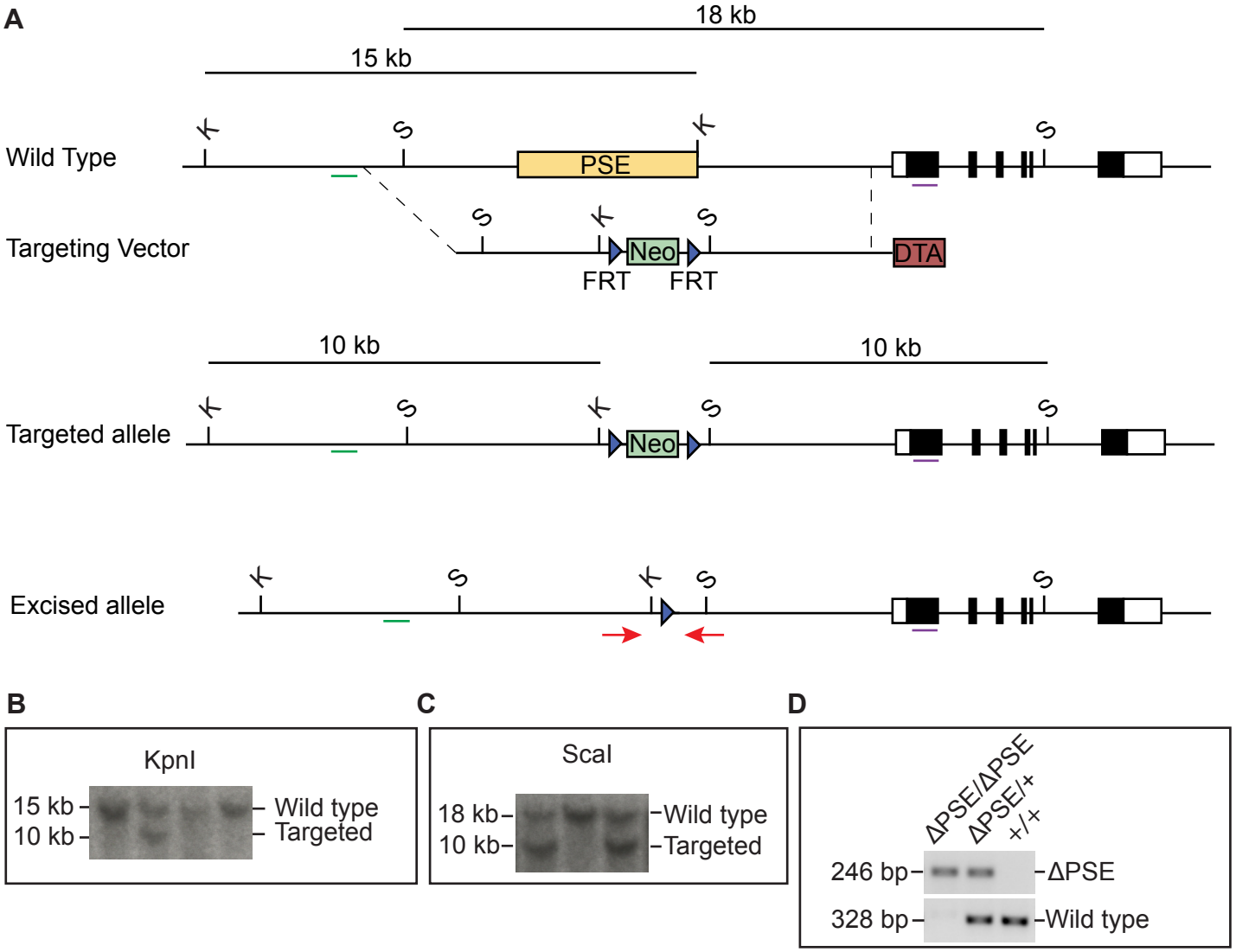
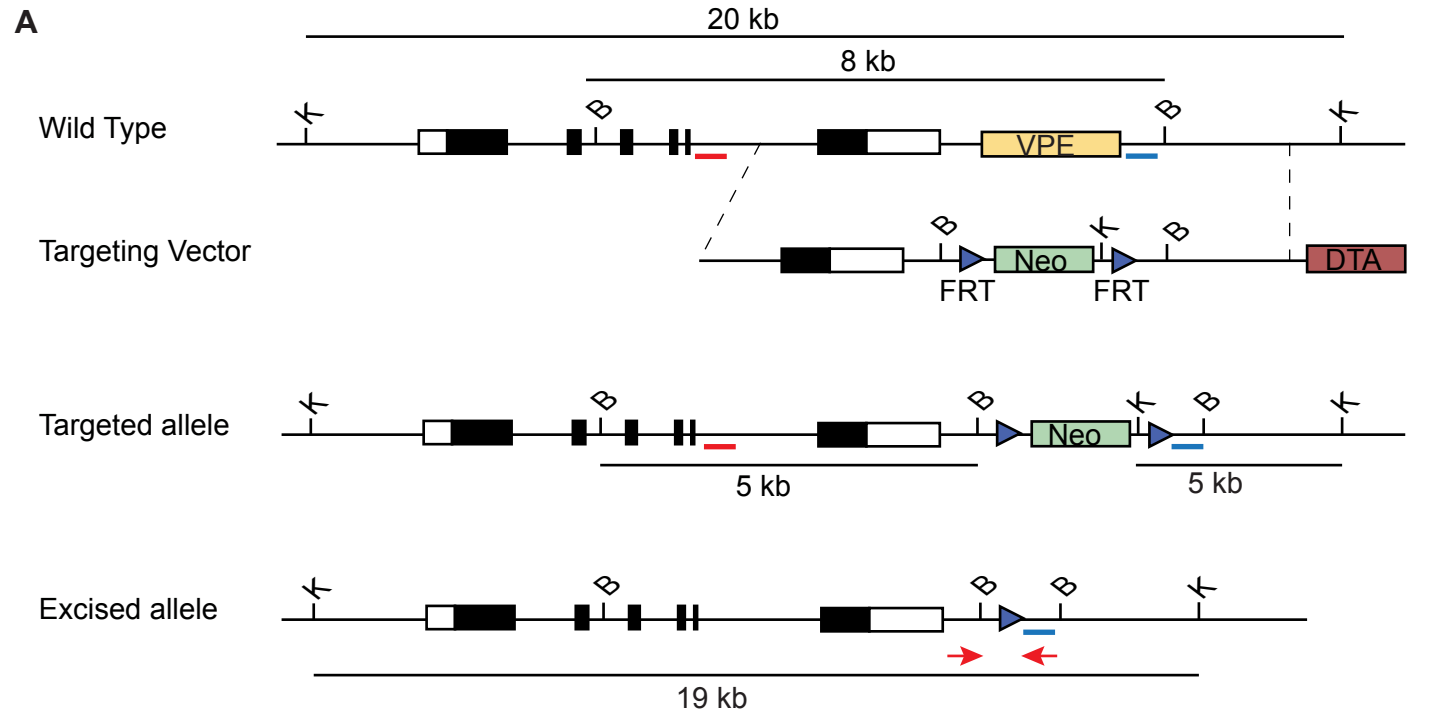
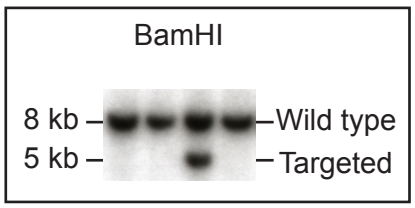
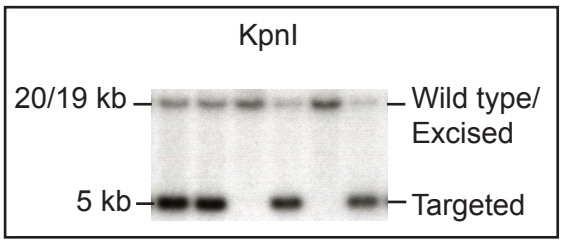


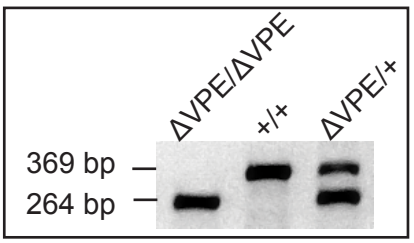
Figure S3: Targeted deletion of the PSE region

(A) Targeting strategy to delete the 5kb PSE region (chr9:118376796-118381570; mm9) by homologous recombination. Southern blot restriction digest used for screening are indicated together with the probes (green and blue bars) and expected fragment sizes for the correctly targeted allele. KpnI (K), ScaI (S), FLP-recombinase recognition site (FRT) site, Neomycin resistance cassette (Neo), Diphtheria toxin A cassette (DTA). Red arrows indicate primers for verifying FLP excision. (B,C) Southern blot of successfully targeted ESC clones. (D) PCR genotyping of Eomes$^{\Delta PSE}$ mice.









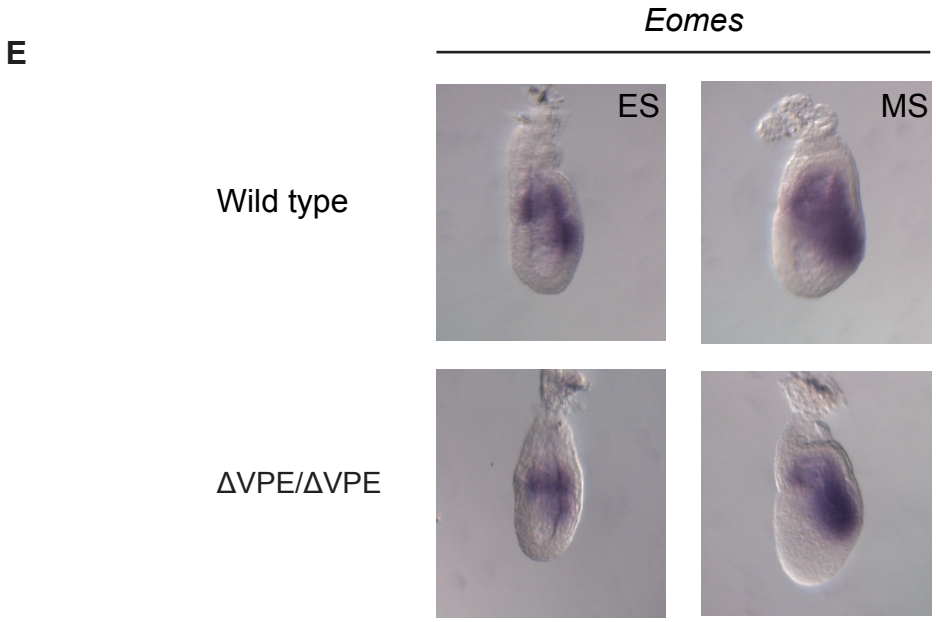


Figure S4: Targeted deletion of the VPE region

(A) Targeting strategy to delete the 0.7kb VPE region (chr9:118395625-118396280; mm9) by homologous recombination. Southern blot probes (red and blue bars), restriction digests and expected fragment sizes are indicated for the targeted and excised alleles. BamHI (B), KpnI (K), FLP-recombinase recognition site (FRT) site, Neomycin resistance cassette (Neo), Diphtheria toxin A cassette (DTA). Red arrows indicate primers for verifying FLP excision. (B) Southern blot of targeted ESC clones. (C) Southern blot to identify excision of Neo cassette in targeted ESC clones. (D) PCR genotyping ΔVPE allele in mice derived from $Eomes^{\Delta VPE/+}$ intercrosses. (E) Whole-mount *in situ* hybridisation of Eomes transcripts at early mid-streak stages shows Eomes expression domains are unaltered in $Eomes^{\Delta VPE/\Delta VPE}$ compared to wild type embryos.

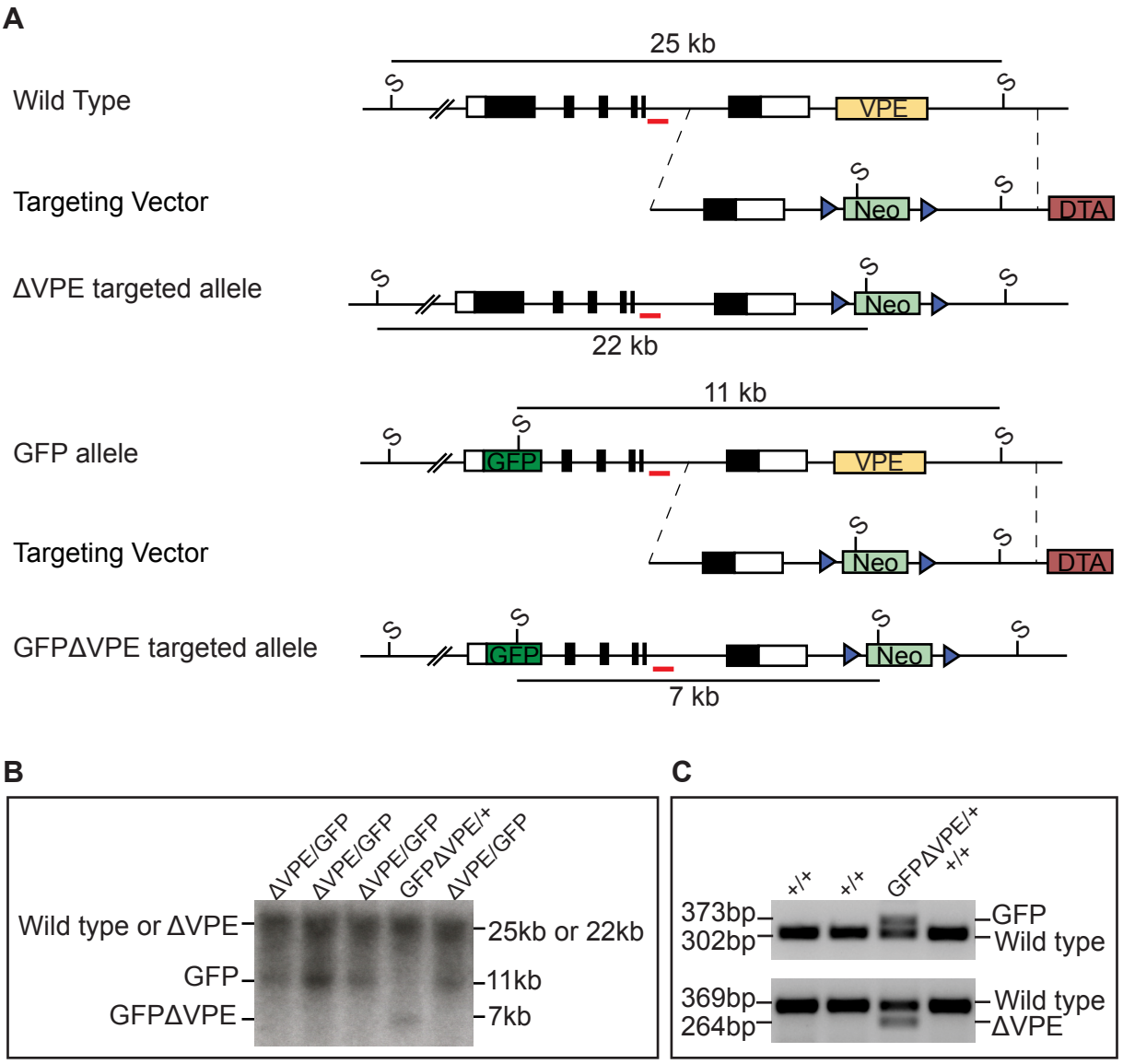


Figure S5: Generating Eomes^{GFP} allele lacking the VPE region

(A) Heterozygous *Eomes*$^{GFP/+}$ (Arnold et al., 2009) ESC were re-targeted using the same construct and primary screening strategy as used to delete the VPE. Southern blot strategy used to distinguish targeting the VPE region in either the GFP or wild type alleles, and expected fragment sizes are indicated. SpeI (S). (B) Southern blot showing two different genotypes of successfully targeted clones; *Eomes*$^{GFP\Delta VPE/+}$ and *Eomes*$^{GFP/\Delta VPE}$. (C) PCR genotyping of *Eomes*$^{GFP\Delta VPE/+}$ mice.

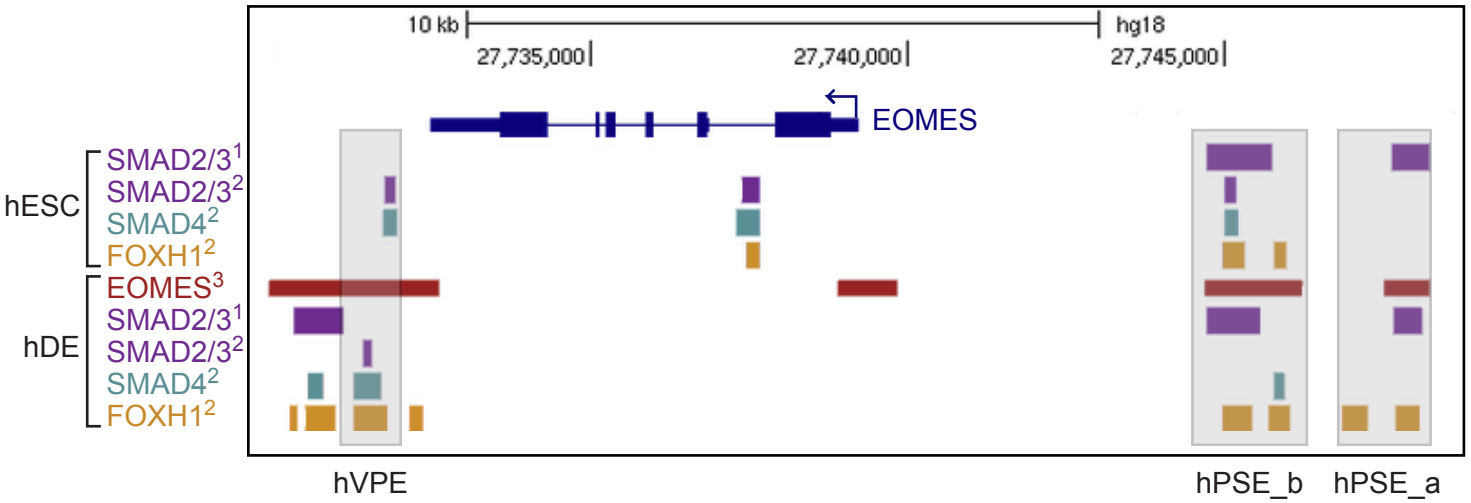


Figure S6: Regulation of the VPE by Nodal signaling

(A) Homologous human regions of the mouse VPE and PSE are bound by EOMES and mediators of the Nodal signaling pathway in hESCs and hDE. 1=(Brown et al., 2011) 2=(Kim et al., 2011), 3=(Teo et al., 2011). ChIP-seq data showing regions bound by SMAD2/3 (purple), SMAD4 (green), FOXH1 (orange) and EOMES (red) are represented by coloured bars and were aligned to the EOMES locus on the UCSC Genome browser Human Mar. 2006 (NCBI36/hg18) Assembly (http://genome.ucsc.edu/). Homologous regions to the mouse VPE and PSE are highlighted in grey. Human VPE, PSE_a and PSE_b (hVPE, hPSE_a, hPSE_b). FoxH1 binds the conserved FoxH1 binding site at the VPE in hDE.

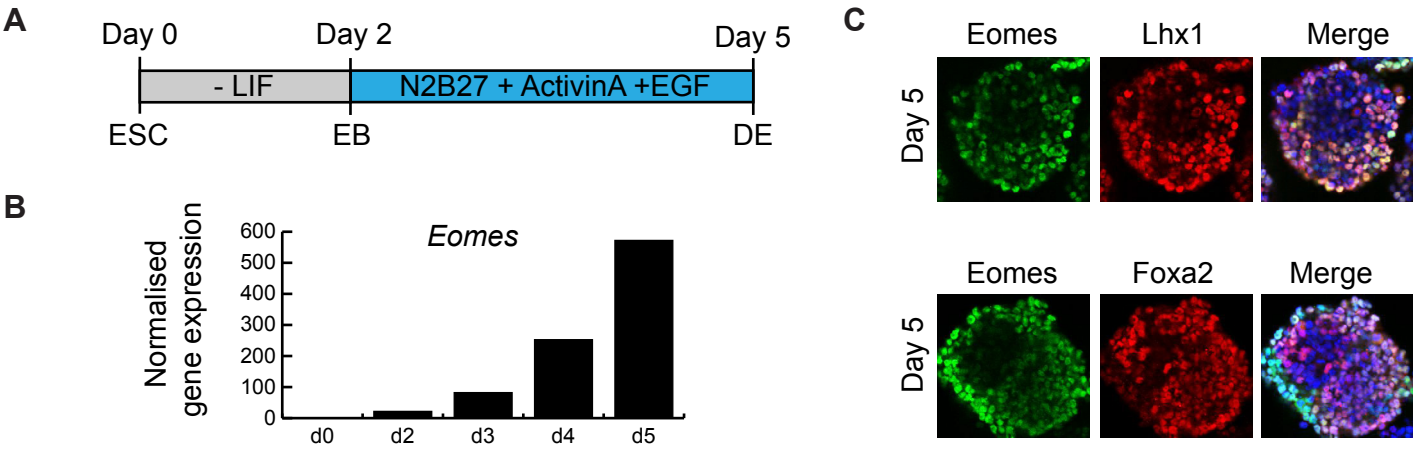


Figure S7: Definitive endoderm differentiation

(A) Schematic of protocol to differentiate ESC to definitive endoderm (DE) fate. ESC were grown in the absence of LIF for 2 days to form embryoid bodies (EB) and then differentiated in N2B27 medium, 20ng/ml ActivinA and 20ng/ml EGF for a further 3 days. (B) qPCR of *Eomes* mRNA demonstrates a dramatic increase in expression over the course of the 5 day differentiation regime. Gene expression is normalised to *Gapdh*. (C) 2D confocal images of d5 DE EBs stained with antibodies against definitive endoderm markers Eomes, Lhx1 or Foxa2, and counterstained with DAPI.

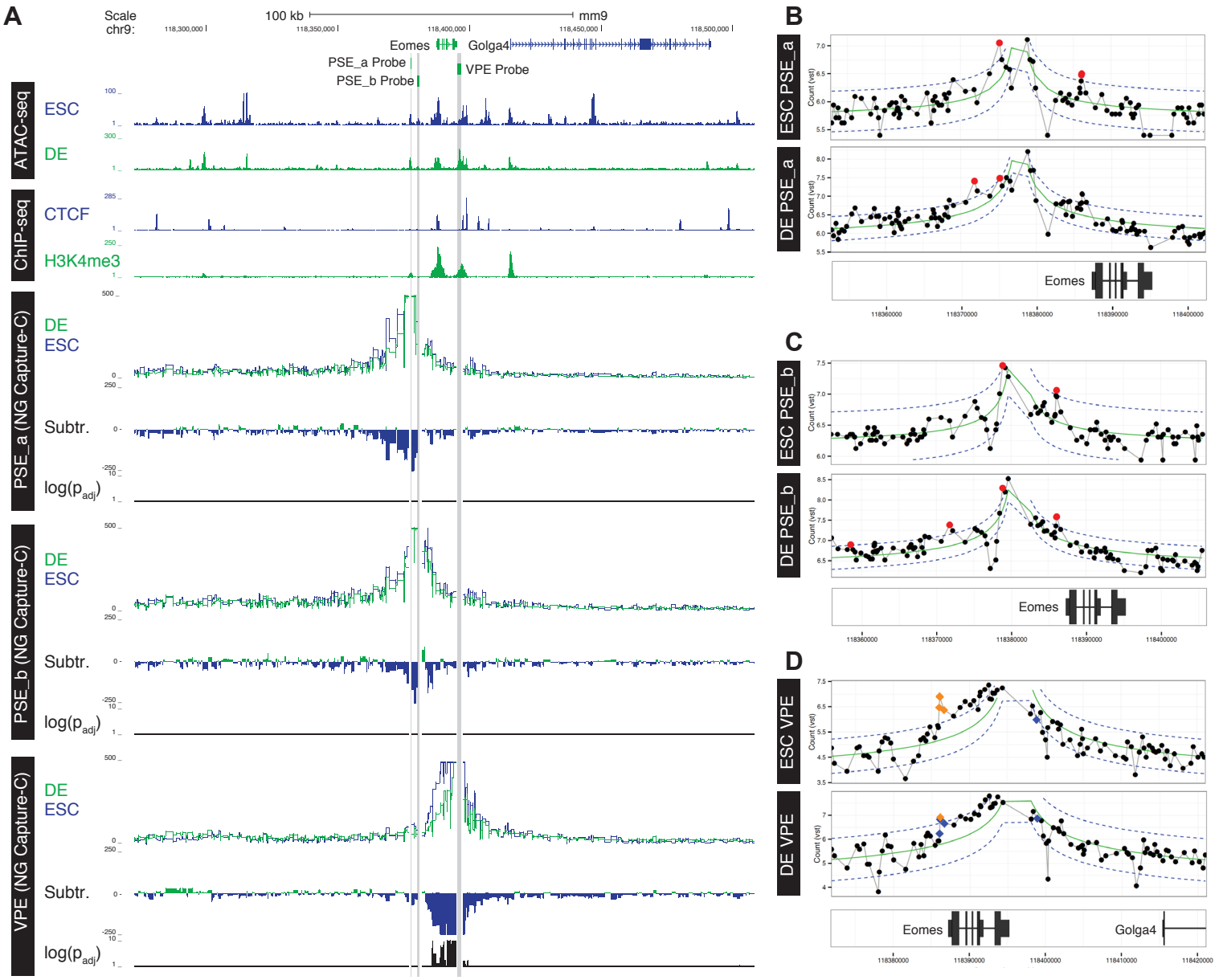


Figure S8: NG Capture-C from the *Eomes* enhancers.

(A) NG Capture-C interaction profiles of the PSE_a, PSE_b, and VPE from ESC (blue) and DE (green). Tracks show mean interactions of normalized biological replicates (n=3), subtraction of ESC from DE (Subtr.) and DESeq2 significant differences between DE and ESC (-log(P_{adj}); p≤0.05). Open chromatin was determined by ATAC-seq in both ESC and DE, ChIP-seq of the boundary element CTCF in ESC is from published data (Handoko et al., 2011) and H3K4me3 ChIP-seq was generated in triplicate from DE. FourCSeq comparison of NG Capture-C between DE and ESC from the PSE_a (B), PSE_b (C), and VPE (D). Red circles mark fragments with more interactions than expected based upon proximity to the promoter (green line), Blue Diamonds show fragments with significantly different interactions between the two conditions (P≤0.05), Orange Diamonds show fragments with enriched reactions that are significantly different between the two conditions.

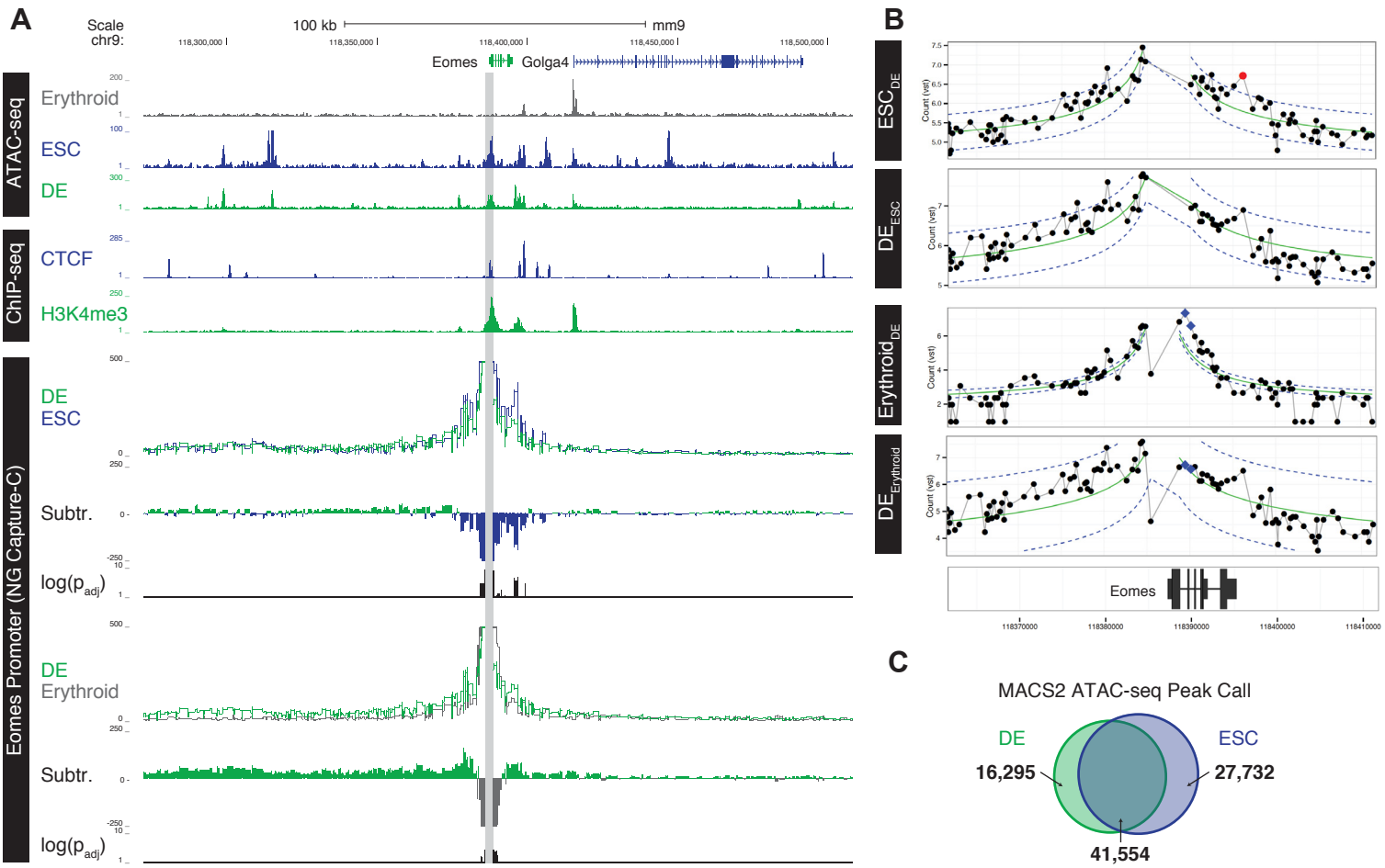


Figure S9: NG Capture-C from the *Eomes* promoter.

(A) NG Capture-C interaction profiles of the *Eomes* promoter from terminally differentiated erythrocytes (Ery, grey), ESC (blue) and DE (green). Tracks show mean interactions of normalized biological replicates (n=3), subtraction of ESC and PHS from DE (Subtr.) and DESeq2 significant differences between the cell types (-log(P_{adj}); p≤0.05). Open chromatin was determined by ATAC-seq in all three cell types (n=3), ChIP-seq of the boundary element CTCF in ESC is from published data (Handoko et al., 2011) and H3K4me3 ChIP-seq was generated in triplicate from DE. (B) FourCSeq comparison of NG Capture-C of the *Eomes* promoter between DE, ESC and Ery. Comparison condition is shown in subscript. Red circles mark fragments with more interactions than expected based upon proximity to the promoter (green line), Blue Diamonds show fragments with significantly different interactions between the two conditions (P≤0.05), Orange Diamonds show fragments with enriched reactions that are significantly different between the two conditions. (C) Comparison of MACS2 peak call for ATAC-seq from DE and ESC.

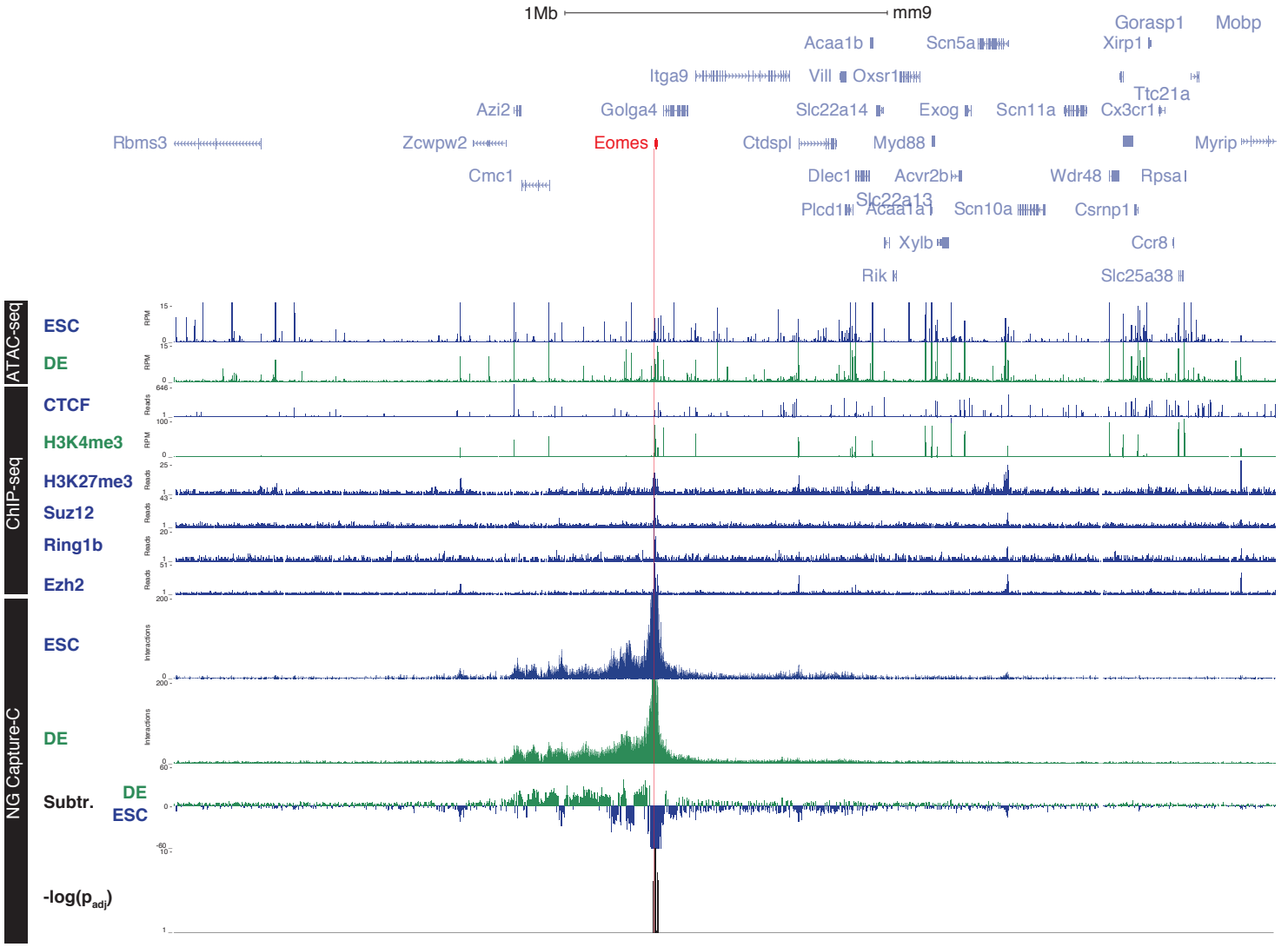


Figure S10: Long-range NG Capture-C from the *Eomes* promoter.

NG Capture-C interaction profiles of the *Eomes* promoter (chr9:116890604-120321539) from erythrocytes (grey), ESC (blue) and DE (green). Tracks show mean interactions of normalized biological replicates (n=3), subtraction of ESC from DE (Subtr.) and DESeq2 significant differences between DE and ESC (-log(P_{adj}); p≤0.05). Location of the Polycomb Repressor Complexes components (Ezh2, Suz12, Ring1b) and associated histone modification (H3K27me3) in ESC are shown (Ku et al., 2008; Mikkelsen et al., 2007). Open chromatin was determined by ATAC-seq in all three cell types (n=3), ChIP-seq of the boundary element CTCF in ESC is from published data (Handoko et al., 2011) and H3K4me3 ChIP-seq was generated in triplicate from DE.

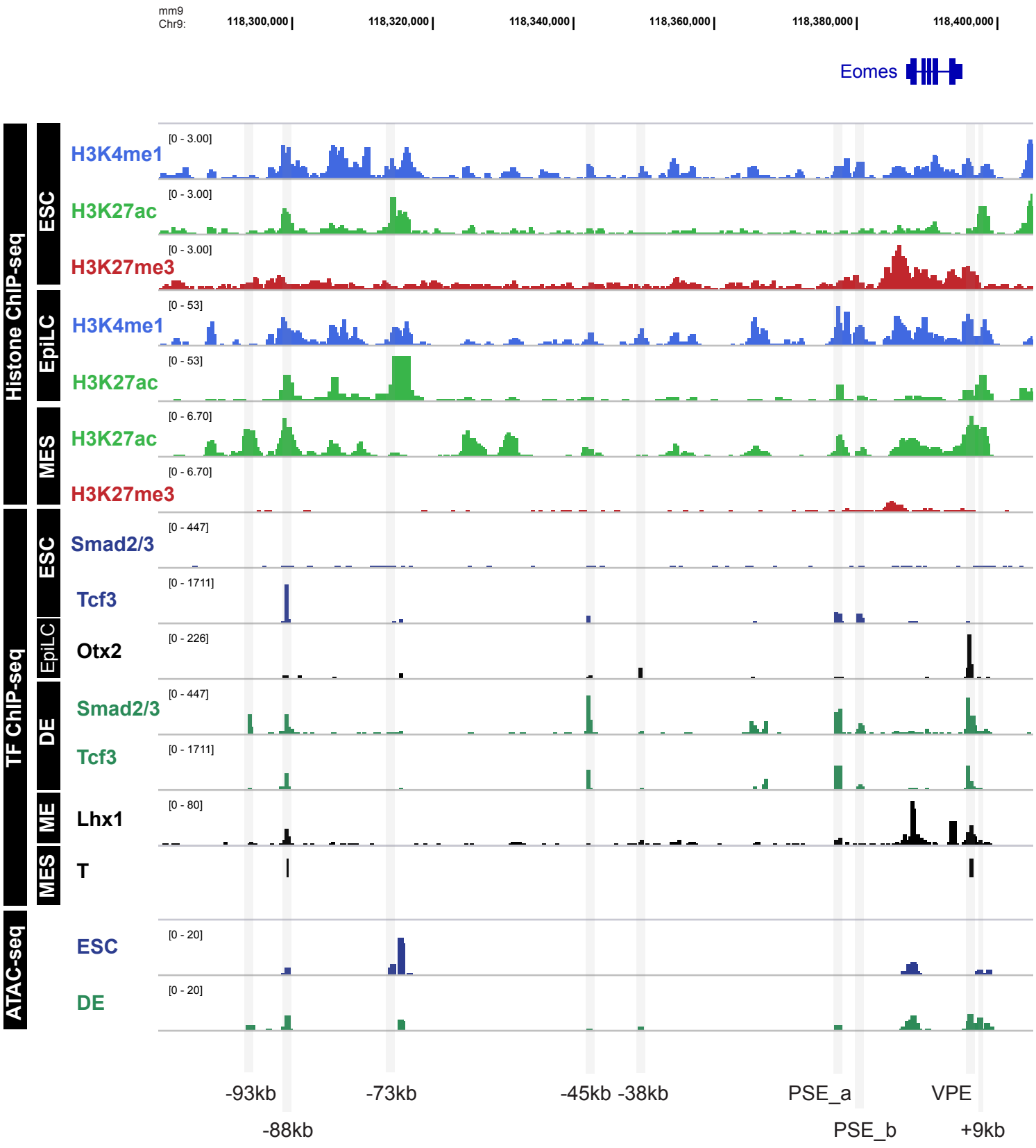


Figure S11: Mapping enhancers within the *Eomes* compartment.

ChIP-seq of histone modifications H3K4me1 (light blue), H3K27me3 (red) and H3K27ac (light green) in ESC, epiblast like cells (EpiLC) and mesoderm (MES) (Alexander et al., 2015; Buecker et al., 2014; Consortium, 2012). Open chromatin was generated using ATAC-seq in ESC and DE (n=3). ChIP-seq of TFs involved in endoderm and anterior mesendoderm specification. Smad2/3 and Tcf3 in ESC (blue) and DE (green) (Wang et al., 2017). Otx2 in EpiLC (Buecker et al., 2014), Lhx1 in P19 mesendoderm (ME) (Costello et al., 2015), and Brachyury (T) (Lolas et al., 2014) in MES. Regions of increased chromatin accessibility unique to ESC (-73kb) and those associated with Smad2/3 occupancy uniquely in DE (-93kb, -45kb, -38kb, PSE_a, VPE and +9kb) are highlighted as in Fig. 5B. In addition, a TF binding hotspot accessible in both ESC and DE (-88kb), and the PSE_b, are also highlighted.

Supplemental References

Alexander, J. M., Hota, S. K., He, D., Thomas, S., Ho, L., Pennacchio, L. A. and Bruneau, B. G. (2015). Brg1 modulates enhancer activation in mesoderm lineage commitment. *Development* **142**, 1418-1430.

Arnold, S. J., Sugnaseelan, J., Groszer, M., Srinivas, S. and Robertson, E. J. (2009). Generation and analysis of a mouse line harboring GFP in the Eomes/Tbr2 locus. *Genesis* **47**, 775-781.

Behringer, R., Gertsenstein, M., Nagy, K. and Nagy, A. (2013). *Manipulating the mouse embryo : a laboratory manual, fourth edition* (Fourth edition. edn): Cold Spring Harbor Laboratory Press.

Brown, S., Teo, A., Pauklin, S., Hannan, N., Cho, C. H., Lim, B., Vardy, L., Dunn, N. R., Trotter, M., Pedersen, R., et al. (2011). Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* **29**, 1176-1185.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T. and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838-853.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

Costello, I., Nowotschin, S., Sun, X., Mould, A. W., Hadjantonakis, A. K., Bikoff, E. K. and Robertson, E. J. (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes Dev* **29**, 2108-2122.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W., Ye, C., Ping, J. L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43**, 630-638.

Kim, S. W., Yoon, S. J., Chuong, E., Oyolu, C., Wills, A. E., Gupta, R. and Baker, J. (2011). Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev Biol* **357**, 492-504.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242.

Lolas, M., Valenzuela, P. D., Tjian, R. and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. *Proc Natl Acad Sci U S A* **111**, 4478-4483.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560.

Teo, A. K., Arnold, S. J., Trotter, M. W., Brown, S., Ang, L. T., Chng, Z., Robertson, E. J., Dunn, N. R. and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev* **25**, 238-250.

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N. A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386-399.

Wang, Q., Zou, Y., Nowotschin, S., Kim, S. Y., Li, Q. V., Soh, C. L., Su, J., Zhang, C., Shu, W., Xi, Q., et al. (2017). The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* **20**, 70-86.

Table S1: Primers used in this study

Primer name	Forward sequence	Reverse sequence	Product
Targeting vectors			
VPE Recombineering	GGCTGGGGTTGGG GAAGGAGTGTTTGC CCTGGAGATGCAAG ATTGTGCTCGGATC CAATTAACCCTCAC TAAAGGGC	GGTCCCAGAAGTTTG GAGGACGGGAAAGA CTGTCCACAGCTCAG GTATATCGAAGTTAT AAGCTTGAAGTTCCT ATACTTTC	n/a
PSE AatII	TGACGTCTGTGTTC AAAAGCACGAGGG	ACCAGAGACCGTATG TTCCC	2.7kb
Transgenic reporter			
VPE LacZ	GCCCTGGAGATGC AAGATTG	CAGCTCAGGTATATC TTCTGGC	696bp
Genotyping			
VPE WT	TCGTTGAGTGGTGA GCAGGGAG	AGCGAGGACATCCA CGGAAAAC	369bp
VPE Δ	TCGTTGAGTGGTGA GCAGGGAG	TTTGGAGGACGGGA AAGACTG	264bp
PSE WT	AGGGTGGCTCTATA CAGGTG	GCATTGGAGTTGAAG GTGGG	328bp
PSE Δ	AGGGTGGCTCTATA CAGGTG	TCACAAGTCTCTCCT GGCAC	246bp
PSE_b WT	TTGCGTTTGTTGGG TTTTGG	CCATCACTGGGAGA GTAGGC	427bp
PSE_b Δ	GGCTATTGCCTCCA TACAGC	CCATCACTGGGAGA GTAGGC	712bp
LacZ	TTACCAGGCCGAAG CAGCGTTGTTG	GCGGCAGTAAGGCG GTCGGGATAGT	300bp
RT-PCR			
Gapdh	CAATGACCCCTTCA TTGACC	GATCTCGCTCCTGGA AGATG	145bp
Eomes	TGTTTTCGTGGAAG TGGTTCTGGC	AGGTCTGAGTCTTGG AAGGTTCATTC	323bp

Table S2: Antibodies used in this study

Name	Catalog number	Company
Foxa2	sc-6554	Santa Cruz
Lhx1	sc-19341	Santa Cruz
TBR2/Eomes	ab23345	Abcam
GFP AlexaFluor 488	A21311	Invitrogen
Goat IgG AlexaFluor 594	A11058	Invitrogen
Rabbit IgG AlexaFluor 488	A21206	Invitrogen
Anti-H3K4me3	07-473	Millipore

Table S3. Long-range *Foxa2* and *Lhx1* promoter interactions identified by NG Capture-C

Click here to Download Table S3

Table S4. Probes used for NG Capture-C.

Click here to Download Table S4

Table S5. Accession codes used in this study.

Click here to Download Table S5