

## RESEARCH ARTICLE

# Mesogenin 1 is a master regulator of paraxial presomitic mesoderm differentiation

Ravindra B. Chalamalasetty<sup>1</sup>, Robert J. Garriock<sup>1</sup>, William C. Dunty, Jr<sup>1</sup>, Mark W. Kennedy<sup>1</sup>, Parthav Jailwala<sup>2</sup>, Han Si<sup>2</sup> and Terry P. Yamaguchi<sup>1,\*</sup>

## ABSTRACT

Neuromesodermal (NM) stem cells generate neural and paraxial presomitic mesoderm (PSM) cells, which are the respective progenitors of the spinal cord and musculoskeleton of the trunk and tail. The Wnt-regulated basic helix-loop-helix (bHLH) transcription factor mesogenin 1 (*Msgn1*) has been implicated as a cooperative regulator working in concert with T-box genes to control PSM formation in zebrafish, although the mechanism is unknown. We show here that, in mice, *Msgn1* alone controls PSM differentiation by directly activating the transcriptional programs that define PSM identity, epithelial-mesenchymal transition, motility and segmentation. Forced expression of *Msgn1* in NM stem cells *in vivo* reduced the contribution of their progeny to the neural tube, and dramatically expanded the unsegmented mesenchymal PSM while blocking somitogenesis and notochord differentiation. Expression of *Msgn1* was sufficient to partially rescue PSM differentiation in *Wnt3a*<sup>-/-</sup> embryos, demonstrating that *Msgn1* functions downstream of *Wnt3a* as the master regulator of PSM differentiation. Our data provide new insights into how cell fate decisions are imposed by the expression of a single transcriptional regulator.

**KEY WORDS:** bHLH transcription factor, Paraxial mesoderm, PSM, Differentiation, EMT, Motility, Wnt, Embryonic stem cell, Somite, Mouse

## INTRODUCTION

The process of vertebrate body formation provides an excellent opportunity to understand the mechanisms driving stem cell differentiation because the body axis extends progressively during development and is sustained by a stem cell population found at the posterior end of the embryo. Neuromesodermal (NM) stem cells initially reside near the node and in the caudal lateral epiblast adjacent to the primitive streak (PS), and contribute to the somites and neural tube of the entire trunk and tail (Nicolas et al., 1996; Cambay and Wilson, 2007; Tzouanacou et al., 2009; Wilson et al., 2009). On the path to formation of the trunk musculoskeleton, NM stem cells must undergo several necessary processes and developmental transitions. Proper entry into the presomitic mesoderm (PSM) cell fate requires that epithelial NM stem cells undergo ingressions in the PS, undertake epithelial-mesenchymal transition (EMT) to become mesoderm, adopt PSM cell identity, undergo lateral cell motility and finally repress neural cell fate. The segmentation clock is a characteristic trait of PSM cells as it is an essential component of a clock-and-wavefront mechanism that periodically divides the anterior PSM into the

discrete segmental units known as somites (Dequéant and Pourquié, 2008). In mice, the PSM is continuously maintained at the posterior end by the addition of new paraxial mesoderm progenitors furnished by the PS (or the tailbud at later stages) until ~65 pairs of somites have been generated. These somites serve as the building blocks of the trunk and tail, forming the muscle, bone, cartilage and dermis. As so many separate processes are associated with acquisition of a PSM cell fate, it is reasonable to expect complex and diverse transcriptional control over these processes; however, how these processes are coordinated remains unknown.

Wnt signaling plays a pivotal role in the formation of the PSM. Loss-of-function mutations in *Wnt3a* or *Ctnnb1* ( $\beta$ -catenin) lead to a loss of paraxial mesoderm progenitors and their derivatives, the PSM and somites, whereas conditional gain-of-function mutations that stabilize  $\beta$ -catenin in the mesoderm result in a large expansion of the PSM (Takada et al., 1994; Cano et al., 2000; Dunty et al., 2008). Several transcriptional regulators lie downstream of Wnt signaling in the PS and, therefore, are excellent candidates to regulate PSM formation (Dunty et al., 2014). These include the T-box transcription factors brachyury (hereafter referred to by its symbol *T*) and *Tbx6*, and the basic helix-loop-helix (bHLH) transcription factor mesogenin 1 (*Msgn1*) (Yamaguchi et al., 1999; Wittler et al., 2007; Dunty et al., 2008; Chalamalasetty et al., 2011). Combinatorial expression of these factors defines nascent paraxial mesoderm (*T*<sup>+</sup>*Tbx6*<sup>+</sup>*Msgn1*<sup>+</sup>) or differentiated PSM (*Tbx6*<sup>+</sup>*Msgn1*<sup>+</sup>). Null mutations in these genes lead to posterior phenotypes that bear similarities to the *Wnt3a*<sup>-/-</sup> phenotype, as they all display a loss of posterior somites and PSM (Fujimoto and Yanagisawa, 1979; Chapman and Papaioannou, 1998; Yoon et al., 2000; Yoon and Wold, 2000). However, gain-of-function studies on *Tbx6* in mice (Wehn and Chapman, 2010) and *T* in embryonic stem cells (ESCs) (Lindsley et al., 2008) do not reveal strong mesoderm-inducing activity, raising the question as to what is the major positive regulator(s) of PSM fate.

In zebrafish, the core pathway known to initiate PSM identity involves co-regulation by *Msgn1* and the *Tbx6*-related T-box transcription factor spadetail (Fior et al., 2012; Yabe and Takada, 2012), indicating a role for multiple transcription factors in PSM cell fate. However, we present evidence that, in mice, *Msgn1* is necessary to maintain the PSM and might be sufficient for PSM specification. Using molecular, genetic and genomic approaches, we show that *Msgn1* directly activates a gene regulatory network that orchestrates PSM differentiation, EMT and motility. Forced expression of *Msgn1* in ESCs and *in vivo*, resulted in cells adopting PSM fates, demonstrating that *Msgn1* is a master regulator of PSM differentiation.

## RESULTS

### *Msgn1* is necessary and sufficient for PSM differentiation

Previous studies of *Msgn1*<sup>-/-</sup> embryos that analyzed the expression of the PSM marker *Tbx6* have been contradictory (Yoon and Wold, 2000; Nowotschin et al., 2012). Given that *Tbx6* is expressed in two

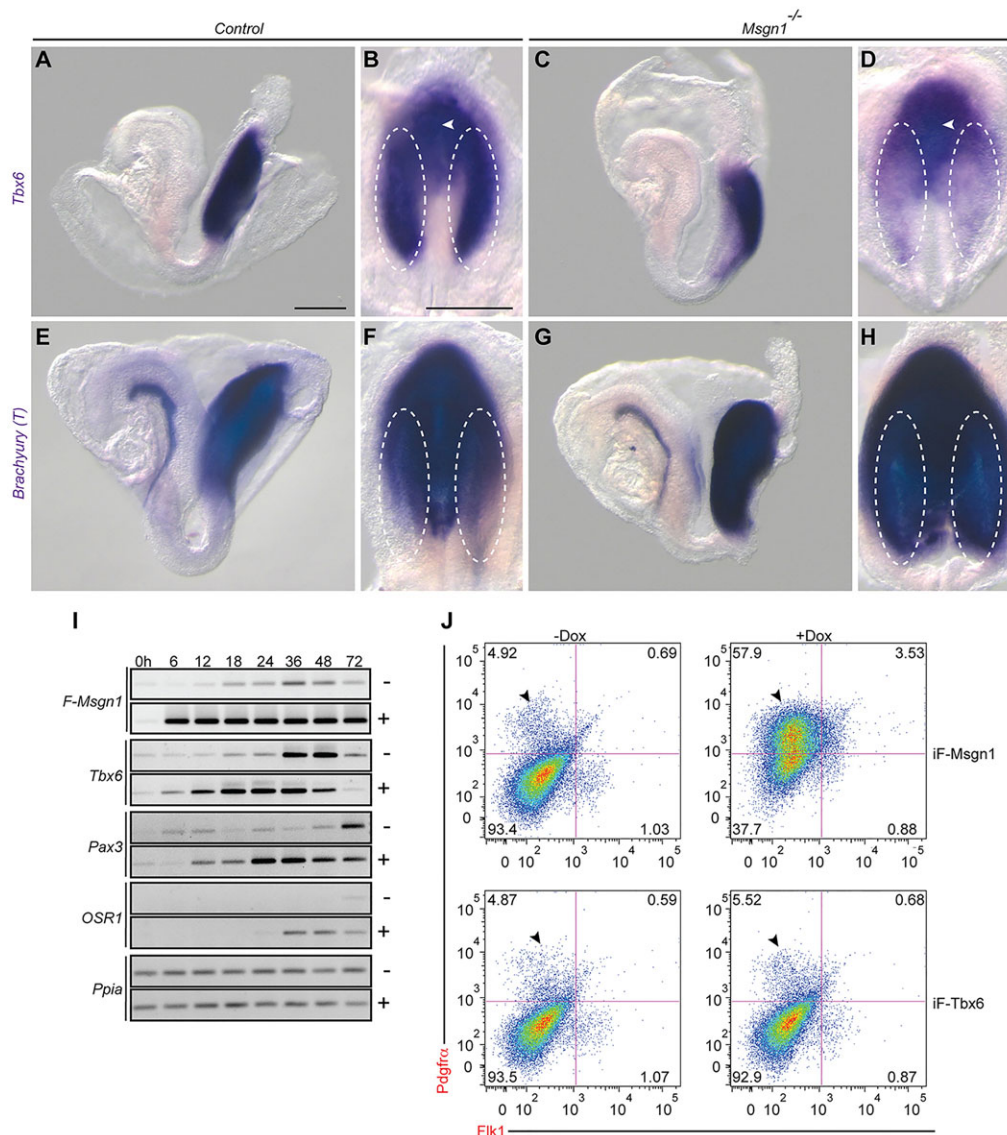
<sup>1</sup>Cancer and Developmental Biology Laboratory, Center for Cancer Research, NCI-Frederick, NIH, Frederick, MD 21702, USA. <sup>2</sup>CCRIFX Bioinformatics Core, Leidos Biomedical Research, FNLCR, Frederick, MD 21702, USA.

\*Author for correspondence (yamagute@mail.nih.gov)

related populations of posterior mesoderm – the nascent paraxial mesoderm emerging from the PS (Fig. 1A,B, arrowhead) and the more anteriorly located differentiated PSM (dashed ovals, Fig. 1B) – we re-examined *Tbx6* expression in *Msgn1*<sup>-/-</sup> embryos at a stage [E8.5, 4–8 somite stage (ss)] that allowed us to distinguish these two populations (Fig. 1C,D). We found that *Tbx6* expression was dramatically reduced in *Msgn1*<sup>-/-</sup> embryos (Fig. 1C), particularly in the PSM domain (cf. dashed ovals, Fig. 1B,D). By contrast, *T* expression in immature paraxial mesoderm progenitors was expanded in the *Msgn1*<sup>-/-</sup> mesoderm compared to wild-type embryos (Fig. 1E–H). These results demonstrate that *Tbx6* expression is *Msgn1*-dependent and suggest that both *Msgn1* and *Tbx6* are necessary for PSM differentiation.

To determine whether *Msgn1* or *Tbx6* is sufficient to drive PSM differentiation, we generated inducible Tet-ON ESC lines to conditionally express *Msgn1* or *Tbx6*. Recombinant ESCs expressing a doxycycline (Dox)-inducible FLAG epitope-tagged *Msgn1* (iF-*Msgn1*) were cultured in suspension for 2 days to initiate embryoid body (EB) formation and epiblast differentiation, before addition of Dox. Dox-mediated induction of *Msgn1* at this timepoint represents premature, ectopic expression because endogenous *Msgn1* is not expressed until day 4 of differentiation (supplementary material

Fig. S1A–C). EBs were treated with Dox for 6–72 h before cells were harvested for RT-PCR analysis. Expression of *Msgn1* in differentiating EBs resulted in the rapid activation of the PSM gene *Tbx6*, and the anterior PSM and somite marker *Pax3*, but not the intermediate mesoderm marker *Osr1* (Mugford et al., 2008) (Fig. 1I). *Pdgfra* is a cell surface marker of mesenchymal paraxial mesoderm progenitors (Sakurai et al., 2006; Takebe et al., 2006), and, together with *Flk1* (*Kdr* – Mouse Genome Informatics), marks progenitors with broad mesodermal potential (Ema et al., 2006; Lindsley et al., 2008). Fluorescence-activated cell sorting (FACS) analysis of iF-*Msgn1* EBs treated with Dox demonstrated that the induction of *Msgn1* resulted in the rapid and preferential cell surface expression of *Pdgfra*. Whereas only 1.36% of control cells were *Pdgfra*<sup>+</sup>, 62.8% were *Pdgfra*<sup>+</sup> after only 12 h of Dox treatment (arrowheads, supplementary material Fig. S1D). A similar percentage of *Msgn1*-expressing cells were *Pdgfra*<sup>+</sup> after 24 h of Dox administration, compared with 4.92% in untreated cells (arrowheads, Fig. 1J). A fivefold increase in *Pdgfra*<sup>+</sup>/*Flk1*<sup>+</sup> double-positive cells was also detected in *Msgn1*-expressing cells (3.53%, compared with 0.69% in controls). Importantly, overexpression of FLAG-tagged *Tbx6* (F-*Tbx6*) in EBs (supplementary material Fig. S1E–G) did not induce

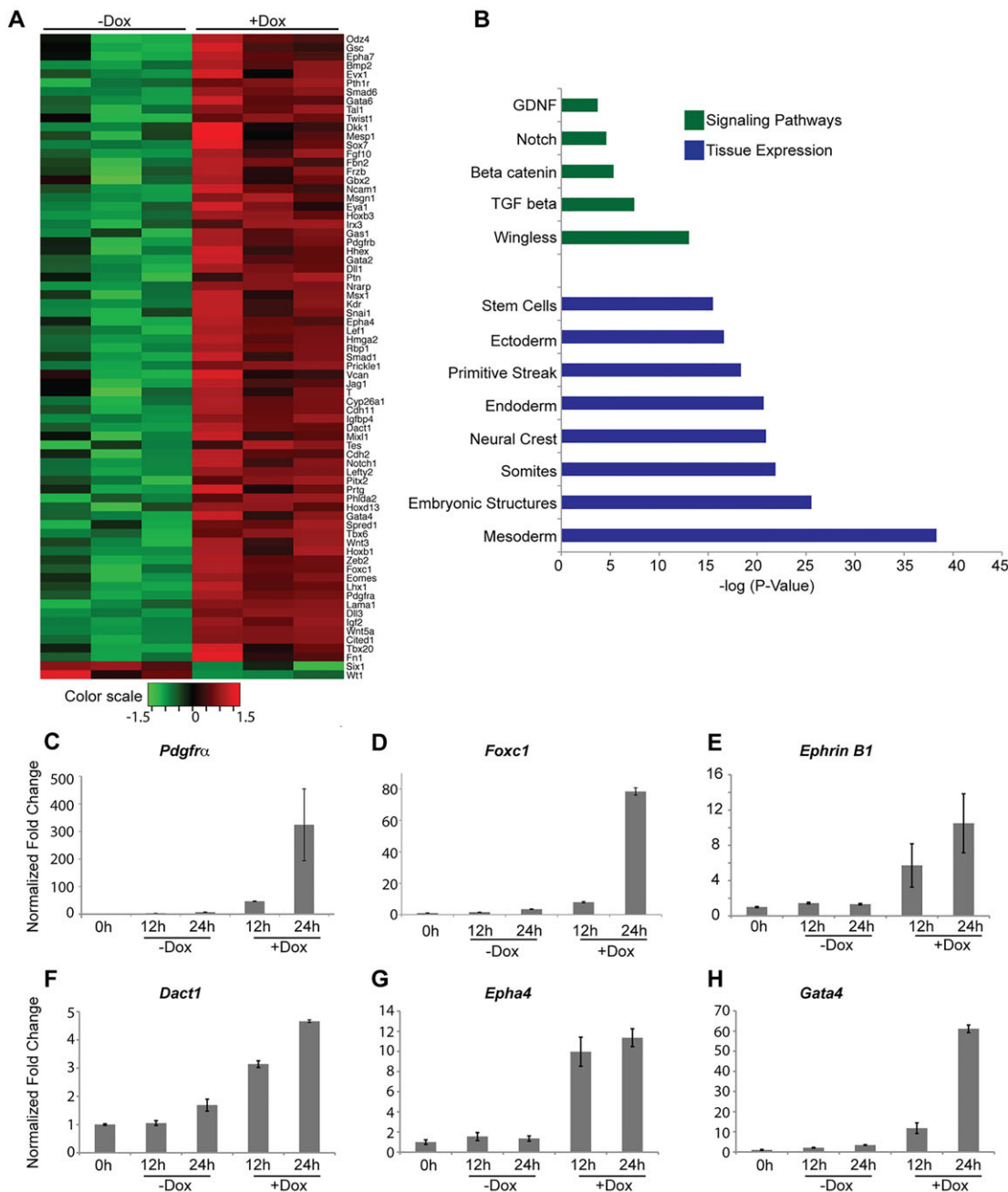


these mesoderm markers (Fig. 1J). These results suggest that *Msgn1* is sufficient to elicit PSM differentiation from ESCs.

### Genome-wide characterization of *Msgn1* target genes

Microarrays were used to assess the effects of *Msgn1* overexpression on global gene expression in EBs. Transcriptional profiles of iF-*Msgn1* EBs at 12, 24 and 48 h after Dox treatment revealed that *Msgn1* rapidly activated gene expression (NCBI GEO dataset GSE29848). After 24 h of Dox treatment, 85% (280/332,  $P \leq 0.05$ ) of the unique, differentially expressed genes were upregulated >1.5 fold (supplementary material Fig. S2A and Table S1). Pathway analysis revealed that a highly

significant number of the genes activated by *Msgn1* in EBs were expressed in mesoderm (73/332,  $P \leq 4.94 \times 10^{-39}$ ) (Fig. 2A, Table 1; supplementary material Table S1), somites, the PS and stem cells, and were associated with developmental signaling pathways including the Wnt/ $\beta$ -catenin pathway (Fig. 2B, Table 1). Remarkably, *Msgn1* activated genes crucial for the development of multiple aspects of PSM development. *Msgn1* activated key mesodermal progenitor markers, including *Tbx6*, *Pdgfra* and *Gata4* (Fig. 1I, Table 1), as well as genes that control EMT (*Snai1*, *Twist1*, *Zeb1*, *Zeb2* and *Foxc1*), segmentation and Notch signaling (*Epha4*, *Dll1*, *Dll3*, *Notch1*, *Lfng* and *Nrarp*), and planar cell polarity (*Wnt5a*, *Prickle1* and *Dact1*) (Table 1).



**Fig. 2. Identification of the *Msgn1* transcriptome in differentiating iF-*Msgn1* EBs.** (A) Hierarchical clustering of differentially expressed mesoderm-associated genes induced by F-*Msgn1* (+Dox) for 24 h ( $P \leq 0.05$ ; fold change  $\geq 1.5$ ). The color bar represents  $\log_2$  intensity values. (B) Pathway and tissue expression analysis of all differentially expressed genes ( $P \leq 0.05$ , fold change  $\geq 1.5$ ) induced by F-*Msgn1* for 24 h. The x-axis shows the  $P$ -value ( $-\log_{10}$ ). (C-H) qPCR analysis of select differentially expressed mesoderm genes in iF-*Msgn1* EBs treated with (+) or without (-) Dox for 12 h and 24 h. The y-axis shows the normalized fold change. Results are mean  $\pm$  s.e.m.

**Table 1. Pathway analysis of differentially expressed genes upon induction of *Msgn1* for 24 h**

Category	Differentially expressed genes
Epithelial to mesenchymal transition	<i>Twist1</i> (1.9), <i>Lef1</i> (1.8), <i>Zeb2</i> (2.5), <i>Zeb1</i> (1.6), <i>Snai1</i> (2.8), <i>Notch1</i> (2.0), <i>Wnt5a</i> (7.4), <i>Mmp2</i> (1.6), <i>Pdgfra</i> (6.2), <i>Foxc1</i> (1.8), <i>Fn1</i> (1.5), <i>Cdh2</i> (1.5), <i>Pdgfrb</i> (1.5)
Mesoderm expression	<i>Igfbp4</i> (3.2), <i>Gata4</i> (2.1), <i>Lefty2</i> (1.5), <i>Prickle1</i> (2.7), <i>Eomes</i> (2.9), <i>Smad1</i> (1.8), <i>Dact1</i> (2.5), <i>Dkk1</i> (1.6), <i>Eya1</i> (1.5), <i>Sox7</i> (2.2), <i>Wnt3</i> (1.8), <i>Kdr</i> (2.7), <i>Cdh11</i> (3.9), <i>Cyp26a1</i> (3.5), <i>Phlda2</i> (1.52), <i>Snai1</i> (2.8), <i>Hhex</i> (1.5), <i>Jag1</i> (1.6), <i>Fbn2</i> (2.4), <i>Tbx6</i> (1.9), <i>Nrarp</i> (2.3), <i>Wnt5a</i> (7.4), <i>Gbx2</i> (1.6), <i>Gas1</i> (1.7), <i>Foxc1</i> (1.8), <i>Smad6</i> (2.0), <i>Prtg</i> (1.8), <i>Evx1</i> (2.0), <i>Zeb2</i> (2.5), <i>Hoxb3</i> (2.1), <i>Lama1</i> (4.0), <i>Pth1r</i> (2.0), <i>Frzb</i> (2.0), <i>Vcan</i> (1.8), <i>Hoxd13</i> (1.7), <i>Gata2</i> (1.5), <i>Odz4</i> (1.7), <i>Dll1</i> (2.0), <i>Lef1</i> (1.8), <i>Bmp2</i> (1.9), <i>Hmga2</i> (2.0), <i>T</i> (1.6), <i>Cdh2</i> (1.5), <i>Pitx2</i> (1.5), <i>Ncam1</i> (1.8), <i>Wt1</i> (−1.8), <i>Epha4</i> (2.9), <i>Tbx20</i> (1.7), <i>Fn1</i> (1.5), <i>Rbp1</i> (1.8), <i>Igf2</i> (3.3), <i>Six1</i> (−1.6), <i>Lhx1</i> (4.2), <i>Mixl1</i> (2.8), <i>Notch1</i> (2.0), <i>Mesp1</i> (1.5), <i>Msx1</i> (1.6), <i>Msgn1</i> (1.8), <i>Tes</i> (1.5), <i>Tal1</i> (2.0), <i>Pdgfra</i> (6.2), <i>Ptn</i> (1.6), <i>Epha7</i> (1.7), <i>Irx3</i> (1.7), <i>Hoxb1</i> (2.4), <i>Gsc</i> (1.9), <i>Twist1</i> (1.9), <i>Gata6</i> (2.1), <i>Spred1</i> (1.7), <i>Pdgfrb</i> (1.5), <i>Dll3</i> (6.3), <i>Fgf10</i> (2.4), <i>Cited1</i> (5.6)
Skeletal system development	<i>Csrp1</i> (−1.5), <i>Smad1</i> (1.8), <i>Mmp2</i> (1.6), <i>Eya1</i> (1.5), <i>Pkdcc</i> (1.7), <i>Hoxb2</i> (1.7), <i>Fbn2</i> (2.4), <i>Lgals3</i> (−1.7), <i>Wnt5a</i> (7.4), <i>Gas1</i> (1.7), <i>Foxc1</i> (1.8), <i>Mgp</i> (−1.5), <i>Hoxb3</i> (2.1), <i>Pth1r</i> (2.0), <i>Frzb</i> (2.0), <i>Hoxd13</i> (1.7), <i>Smad3</i> (1.6), <i>Bmp2</i> (1.9), <i>Col1a2</i> (1.6), <i>Six1</i> (−1.5), <i>Msx1</i> (1.6), <i>Pdgfra</i> (6.2), <i>Hoxb1</i> (2.4), <i>Gsc</i> (1.9), <i>Twist1</i> (1.9), <i>Pdgfrb</i> (1.5), <i>Zeb1</i> (1.6), <i>Dll3</i> (6.3)
Wnt signaling pathway	<i>Igfbp4</i> (3.2), <i>Sp5</i> (2.5), <i>Prickle1</i> (2.7), <i>Lgr5</i> (1.7), <i>Dact1</i> (2.5), <i>Dkk1</i> (1.6), <i>Sox7</i> (2.2), <i>Wnt3</i> (1.8), <i>Cyp26a1</i> (3.5), <i>Snai1</i> (2.8), <i>Hhex</i> (1.5), <i>Tbx6</i> (1.9), <i>Nrarp</i> (2.3), <i>Lgals3</i> (−1.7), <i>Wnt5a</i> (7.4), <i>Gbx2</i> (1.6), <i>Prf2c2</i> (−2.0), <i>Foxc1</i> (1.8), <i>Peg3</i> (1.7), <i>Frzb</i> (2.0), <i>Olfm1</i> (1.9), <i>Lef1</i> (1.8), <i>Nkd1</i> (1.7), <i>Bmp2</i> (1.9), <i>T</i> (1.6), <i>Cdh2</i> (1.5), <i>Pitx2</i> (1.5), <i>Hprt</i> (5.3), <i>Wt1</i> (−1.8), <i>Nr0b1</i> (−1.9), <i>Six1</i> (−1.6), <i>Lhx1</i> (4.2), <i>Mixl1</i> (2.8), <i>Mesp1</i> (1.5), <i>Msx1</i> (1.6), <i>Irx3</i> (1.7), <i>Dixdc1</i> (1.5), <i>Nkd2</i> (1.6), <i>Pdzrn3</i> (1.7), <i>Rspo3</i> (2.6), <i>Gata6</i> (2.1), <i>Sp8</i> (1.9), <i>AW551984</i> (1.7), <i>Fgf10</i> (2.4)
TGFβ signaling pathway	<i>Igfbp4</i> (3.2), <i>Gata4</i> (2.1), <i>Serpinh1</i> (1.6), <i>Lefty2</i> (1.5), <i>Lmo1</i> (1.9), <i>Eomes</i> (2.9), <i>Smad1</i> (1.8), <i>Bmper</i> (2.0), <i>Snai1</i> (2.8), <i>Hhex</i> (1.5), <i>H2-Ab1</i> (1.6), <i>Fbn2</i> (2.4), <i>Foxc1</i> (1.8), <i>Smad6</i> (2.0), <i>Id1</i> (1.8), <i>Mgp</i> (−1.5), <i>Fli1</i> (4.8), <i>Zeb2</i> (2.5), <i>Fstl1</i> (2.1), <i>Vcan</i> (1.8), <i>Smad3</i> (1.6), <i>Ets1</i> (1.8), <i>Bmp2</i> (1.9), <i>Hmga2</i> (2.0), <i>T</i> (1.6), <i>Tbx20</i> (1.7), <i>Fn1</i> (1.5), <i>Col1a2</i> (1.6), <i>Six1</i> (−1.6), <i>Mixl1</i> (2.8), <i>Msx1</i> (1.6), <i>Twist1</i> (1.9), <i>Gata6</i> (2.1), <i>Zeb1</i> (1.6), <i>Zfp521</i> (1.7), <i>Fgf10</i> (2.4), <i>Cited1</i> (5.6)
Notch signaling pathway	<i>Dll1</i> (2.0), <i>Dll3</i> (6.3), <i>Notch1</i> (2.0), <i>Nrarp</i> (2.3), <i>Epha4</i> (2.9), <i>Lfng</i> , <i>Maml3</i> (Chalamalasetty et al., 2011)
Planar cell polarity pathway	<i>Wnt5a</i> (7.4), <i>Prickle</i> (2.7), <i>Dact1</i> (2.5)

Differentially expressed genes with  $P \leq 0.05$  and a fold change  $\geq 1.5$  are listed. Values in parentheses indicate fold changes. Where there is no value, there is no change.

*Msgn1*-dependent gene activation was validated by qPCR analysis for select genes (Fig. 2C–H; Chalamalasetty et al., 2011). These data show that *Msgn1* regulates the expression of key regulators of PSM fate.

To determine which of these genes represent direct targets of *Msgn1*, genomic *Msgn1*-binding sites in differentiating EBs were determined by chromatin immunoprecipitation sequencing (ChIP-Seq). A total of 4087 identified *Msgn1*-binding sites or ‘peaks’ were assigned to 1860 genes because they were found within or adjacent [−10 kb to Transcription Start Site (TSS); +5 kb to Transcription Termination Site (TTS)] to a gene. A significant number of peaks were within 2 kb upstream of the TSS, suggesting that *Msgn1* predominantly bound to promoters (Fig. 3A,B; supplementary material Table S2). Integration of the ChIP-Seq and microarray data revealed that one-third (121/332 genes) of the genes that were differentially expressed upon *Msgn1* induction were directly bound by *Msgn1* (Fig. 3C; supplementary material Table S3). Gene ontology (GO) term analysis using the GREAT algorithm (McLean et al., 2010) showed enrichment for genes that are expressed in the PS and mesoderm, and those that regulate mesoderm differentiation and morphogenesis, germ layer formation and EMT, among others (Fig. 3D; supplementary material Fig. S2B and Table S2). Analysis of all 4087 *Msgn1*-binding site sequences for common sequence motifs identified an E-box motif (CANNTG), specifically CCATTTGT (Fig. 3E; supplementary material Fig. S2C), which is a well-characterized binding site for bHLH transcription factors (Jones, 2004). This *Msgn1*-binding site was located centrally within identified peaks (supplementary material Fig. S2D) and bound to *Msgn1* in electrophoretic mobility shift assays (EMSA) (supplementary material Fig. S3; Chalamalasetty et al., 2011). Several other interesting motifs were associated with *Msgn1*-binding sites, including binding sites for the Ets, T-box, Gata and Sp transcription factors (supplementary material Fig. S2E).

### ***Msgn1* directly activates key regulators of PSM fate**

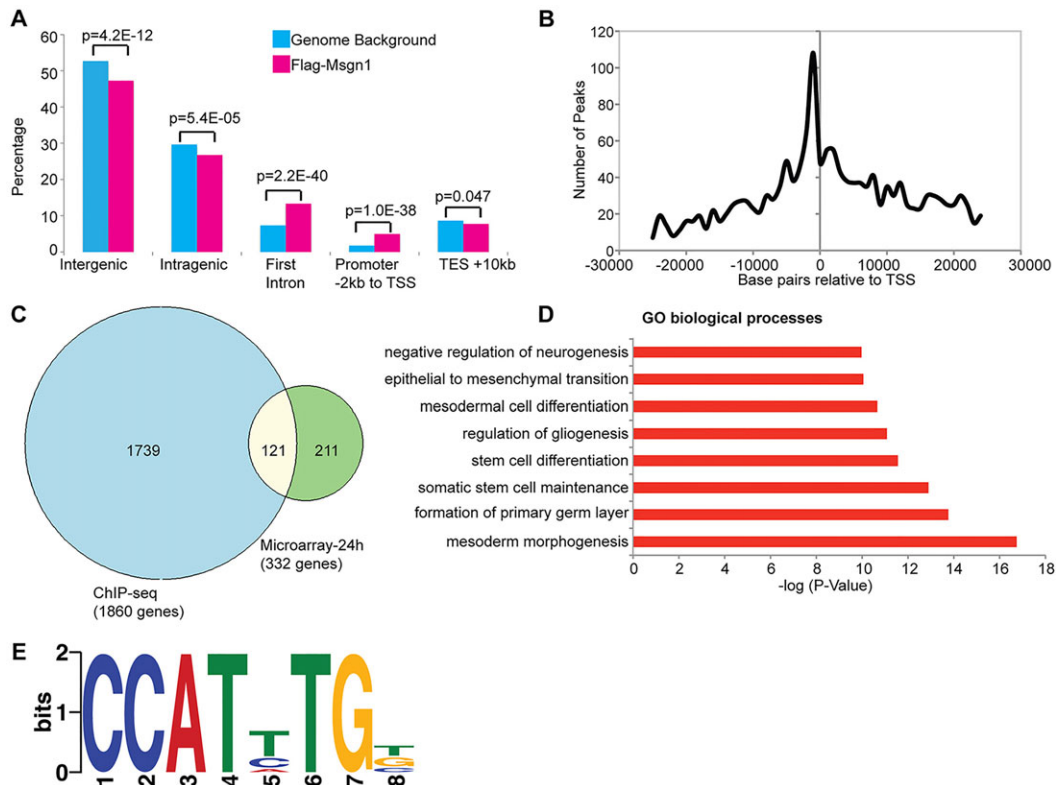
The rapid expression of *Tbx6* and *Pdgfra* mRNA and protein in EBs upon *Msgn1* induction (Fig. 1I; supplementary material Fig. S1D) is

consistent with *Msgn1* directly activating these genes. Indeed, ChIP-Seq revealed that *Msgn1* bound to regulatory elements in both *Tbx6* (Fig. 4A) and *Pdgfra* (Fig. 4D). The *Tbx6+3913* 3′ enhancer contained four E-boxes (Fig. 4A), including two embedded consensus *Msgn1*-binding sites (E1 and E3) that bound (supplementary material Fig. S3A), and were activated by, *Msgn1* *in vitro* (Fig. 4B). The *Tbx6* ChIP-seq peak was validated by ChIP-qPCR in iF-*Msgn1* EBs (Fig. 4C). Similarly, six E-boxes (E1–6) were discerned in the *Pdgfra* −2140 peak (Fig. 4D), including two consensus *Msgn1*-binding sites (E2 and E4) that bound *Msgn1* (supplementary material Fig. S3B). Mutation of E2 showed that it was necessary for *Msgn1*-dependent activation of a *Pdgfra* −2140 enhancer-luciferase reporter (Fig. 4E). Importantly, both *Tbx6* and *Pdgfra* were downregulated in the *Msgn1*<sup>−/−</sup> PSM (Fig. 1A–D, Fig. 4F–I), clearly demonstrating their dependence upon *Msgn1* *in vivo*. These data show that *Msgn1* positively regulates key mediators of the PSM cell fate.

### ***Msgn1* is a direct regulator of EMT**

Given that *Snai1* and other EMT genes are induced by *Msgn1* (Table 1), and that *Snai1* is considered a master regulator of the EMT program (Batlle et al., 2000; Barrallo-Gimeno and Nieto, 2005), we examined the mechanism of *Snai1* activation by *Msgn1*. RT-PCR analysis validated the microarray data, showing that *Snai1* was rapidly and robustly activated in EBs within 6 h of *Msgn1* induction (Fig. 5A). ChIP-seq analysis revealed that *Msgn1* bound a previously described *Snai1* 3′ enhancer (Palmer et al., 2007) that contains two E-boxes (Fig. 5B). Co-transfection of *Msgn1* expression vectors with *Snai1*+5148 luciferase reporter constructs showed that *Msgn1* activated the reporter sevenfold (Fig. 5C). Moreover, binding of *Msgn1* to the *Snai1* enhancer was confirmed by ChIP assays in both iF-*Msgn1* EBs and PSM extracts using anti-*Msgn1* antibody (Fig. 5D).

To determine whether *Snai1* expression *in vivo* depends upon *Msgn1* and the Wnt3a/β-catenin signaling pathway, we examined *Snai1* expression in *Wnt3a*<sup>−/−</sup>, conditional *Ctnnb1* loss-of-function



**Fig. 3.** ChIP-seq analysis of iF-Msgn1 EBs. (A) Genomic distribution of 4087 Msgn1 peaks (pink) compared with those from genome background (blue). For genome background, percentages of peaks from random sampling of similar length reads across the genome background are used. A binomial exact test was used to calculate the  $P$ -values. TSS, transcription start site; TES, transcription end site. (B) Msgn1 peaks within 25 kb upstream or downstream of the TSS are plotted. (C) Venn diagram depicting the intersection of 1860 genes bound by Msgn1 (blue circle) with 332 differentially expressed genes at the 24-h time point (green circle). The overlap is statistically significant ( $P \leq 4.6 \times 10^{-80}$ , hypergeometric probability test). (D) GO biological processes associated with genes bound by F-Msgn1 and analyzed by GREAT. The  $x$ -axis shows the  $P$ -value ( $-\log_{10}$ ). (E) Discriminative DNA motif discovery (DREME) analysis of Msgn1 peaks identifies an E-box (underlined), CCATHTGB, as the top motif ( $E$ -value =  $7.6 \times 10^{-279}$ ). The motif logo displays nucleotide conservation at that position (measured in bits) and height of the symbol reflect the relative frequency at that position.

and gain-of-function mutants, and *Msgn1*<sup>-/-</sup> embryos. Whole-mount *in situ* hybridization (WISH) confirmed that *Snail*, like *Msgn1* (Chalamalasetty et al., 2011), depended upon Wnt3a and  $\beta$ -catenin signaling for expression in the PSM (supplementary material Fig. S3C–F). Consistent with *Snail* being a direct target of Msgn1, examination of *Msgn1* mutants showed that *Snail* expression was absent from the *Msgn1*<sup>-/-</sup> PSM, but, as expected, was maintained in the allantois and head mesenchyme where Msgn1 is not normally expressed (Fig. 5E–H; supplementary material Fig. S3G).

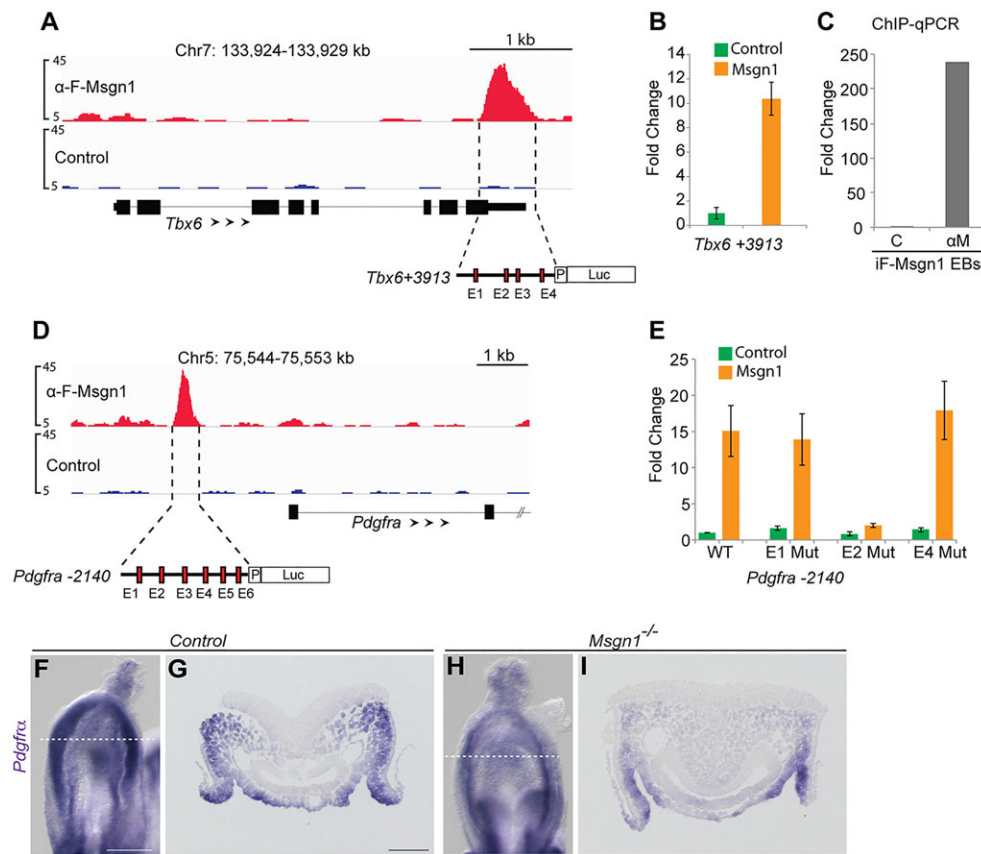
Additionally, the ChIP-seq data showed that six additional genes implicated in the regulation of EMT [*Zeb2*, *Foxc1*, *Lef1*, *Noch1*, fibronectin (*Fn1*) and *Pdgfra*, see above] were directly bound by Msgn1. Quantitative real-time PCR (qPCR) analysis confirmed that Msgn1 expression in EBs resulted in the upregulation of several mesenchymal markers including N-cadherin (*Cdh2*), *Fn1*, *Foxc1*, *Mmp2*, *Twist1* and vimentin (*Vim*), whereas epithelial and junctional markers, such as E-cadherin (*Cdh1*), claudin 3 and claudin 7 (*Cldn3*, *Cldn7*), and occludin (*Ocln*) were downregulated (Fig. 6A; supplementary material Fig. S3H). As many of the targets of Msgn1 are also *Snail* targets, the results show that Msgn1 activates a redundant mechanism to initiate the EMT program (Cano et al., 2000; Guaita et al., 2002; Barrallo-Gimeno and Nieto, 2005).

The lack of *Snail* expression in the *Msgn1*<sup>-/-</sup> PSM prompted us to examine embryo sections for defects in EMT, which could manifest as abnormal cell morphology, ingression or movement in

the PS. Surprisingly, cells with mesenchymal morphology had clearly ingressed in the mutant PS (Fig. 6B,C). The morphology of these *Msgn1*<sup>-/-</sup> cells in the midline PS was indistinguishable from similarly located wild-type mesenchymal cells that had ingressed but had not yet moved from the PS, suggesting that the mutant cells display a cell motility defect. To directly test this, we assessed cell motility in PS explant cultures (Fig. 6D,E). In 18 of 21 control explants, motile cells migrated out of the explant, traveling distances of up to 1100  $\mu$ m within 48 h of culture (Fig. 6F). These motile cells expressed the mesenchymal marker vimentin, and not the epithelial marker E-cadherin (Fig. 6G–J). By contrast, motile, vimentin-positive cells were much less abundant in the periphery of *Msgn1*<sup>-/-</sup> explants ( $n=14$ ; Fig. 6K–N) and displayed significantly reduced migration, not traveling for more than  $\sim 300$   $\mu$ m (Fig. 6F). These results suggest that Msgn1 is required for the initiation of a motile mesenchymal state in PSM cells.

#### Msgn1 specifies PSM fates *in vivo*

To test whether Msgn1 can specify PSM fate *in vivo*, we generated a ‘Tet-ON’ *Msgn1*-expressing transgenic mouse line to be used in conjunction with Cre recombinase-expressing transgenic lines to precisely control the spatiotemporal expression of Msgn1 (supplementary material Fig. S4A). *T-Cre* mice were used to express Msgn1 in the trunk region, presumably in the NM stem cell that gives rise to both the neural tube and mesodermal lineages (Perantoni et al., 2005). Administration of Dox to *T-Cre*<sup>tg/+</sup>; *R26-flox-*



**Fig. 4. The paraxial mesoderm regulators *Tbx6* and *Pdgfra* are direct target genes of *Msgn1*.** (A) *Msgn1*-binding peaks (red) and input control (blue) at the *Tbx6* locus for the indicated genomic intervals (mm9 coordinates). The schematic (below) illustrates the four E-boxes (E1–4, red rectangles) identified in the *Tbx6*+3913 enhancer peak. Arrowheads reflect gene orientation and the y-axis indicates the peak height corresponding to the normalized fold enrichment. (B) Luciferase reporter assay of *Tbx6*+3913 enhancer co-transfected with pCS2 Control (green) or pCS2-*Msgn1* (orange) expression constructs. The y-axis shows the normalized fold change; error bars show the s.d.;  $n=3$ . (C) ChIP-qPCR analysis of F-Msgn1 binding to *Tbx6*+3913 enhancer using control 'C' and anti-*Msgn1* 'αM' ascites in iF-Msgn1 EBs. (D) *Msgn1* binds the *Pdgfra* locus at -2140 bp, relative to TSS. (E) Luciferase reporter assay of the *Pdgfra* -2140 wild-type (WT) and E-box mutant enhancers co-transfected with control or *Msgn1* expression constructs. The y-axis shows the normalized fold change; error bars show the s.d.;  $n=3$ . (F–I) WISH analysis of *Pdgfra* mRNA expression in control and *Msgn1*<sup>-/-</sup> embryos. G and I are cross sections through the PS [dashed white lines in F and H (ventral view)]. Scale bars: 200  $\mu$ m (F, H), 50  $\mu$ m (G, I).

*rtTA-Ires-EGFP*<sup>+</sup>; *lacZ-TRE-F-Msgn1*<sup>tg/+</sup> (hereafter referred to as *Tcre-F-Msgn1*<sup>GOF</sup>, where GOF stands for gain-of-function) transgenic embryos led to the highly elevated expression of *Msgn1* and the *Msgn1* reporter  $\beta$ -galactosidase ( $\beta$ -gal) (Fig. 7A–F; supplementary material Fig. S4B–D) throughout the trunk neural tube and paraxial mesoderm.

Using EGFP (*rtTA-Ires-EGFP*) expression as a permanent cell lineage tracer, the ectopic expression of *Msgn1* in NM stem cells and their progeny resulted in a reduction in the number of EGFP-expressing cells contributing to the neural tube. Although 62% ( $n=5$ ) of neural cells were EGFP<sup>+</sup> in controls, only 22% ( $n=5$ ) were EGFP<sup>+</sup> when *Msgn1* was induced (Fig. 7A–G), suggesting that *Msgn1* expression diverted NM progeny away from a neural fate. Nevertheless, F-Msgn1 was still detected in the cells that remained in the neural tube. These cells were Sox2<sup>+</sup> and *Tbx6*<sup>-</sup>, suggesting that they retained neural character (supplementary material Fig. S4E–L). To investigate whether the reduced number of EGFP<sup>+</sup> neural cells was caused by *Msgn1*-induced cell death, we examined *Tcre-F-Msgn1*<sup>GOF</sup> embryos for expression of the apoptosis initiator protein caspase-3. Although caspase-3 was detected in 2/5 and 3/3 mutants at E9.5 and E10.5, respectively, it was mostly observed in mesoderm cells and did not colocalize with F-Msgn1 (supplementary material Fig. S5), suggesting that the cell death is a secondary consequence of *Msgn1* expression.

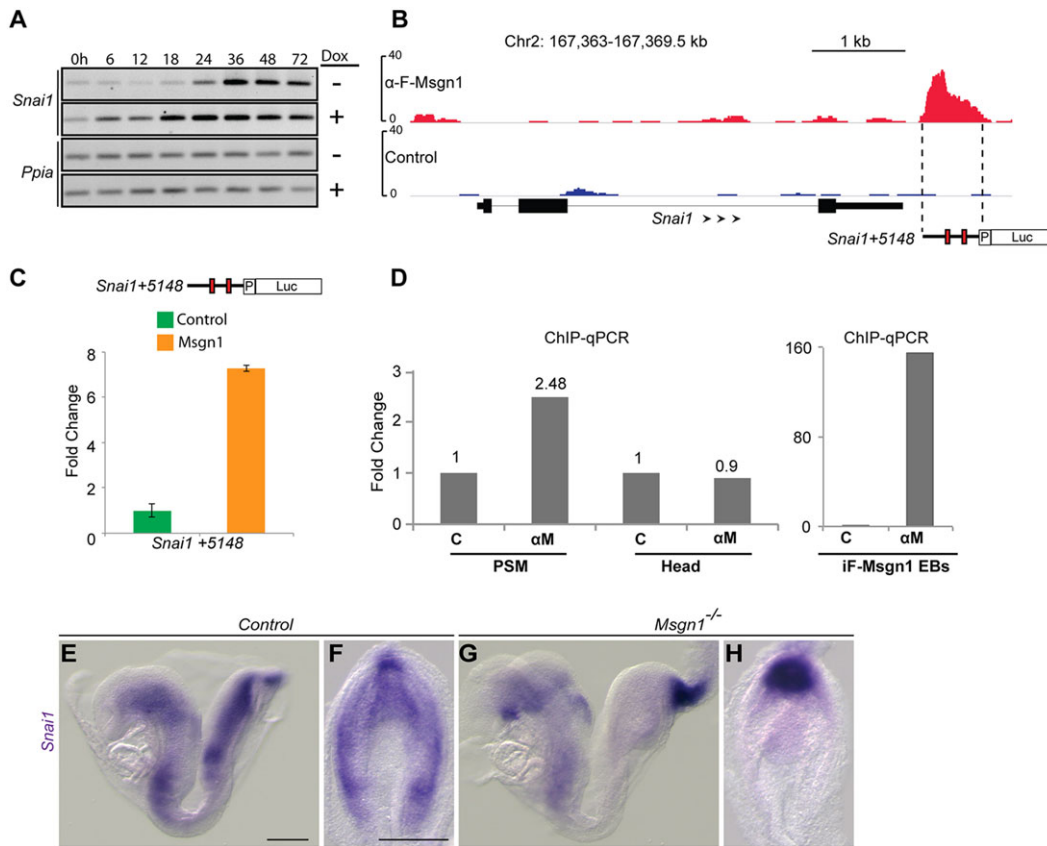
The reduced number of NM stem cell progeny contributing to the neural tube was accompanied by a correspondingly dramatic increase in the number of *Tcre-F-Msgn1*<sup>GOF</sup> PSM cells (Fig. 7H–N), and by the enlarged mRNA and protein expression domains of the PSM markers *Dll3* and *Tbx6* (Fig. 7O–T; supplementary material Fig. S4I–L). The expansion of the PSM came at the expense of somites because segment boundaries were absent and the expression of the somite polarity marker *Uncx4.1* (*Uncx* – Mouse Genome

Informatics) was reduced or absent (Fig. 7O–T). *T* expression in the germ layer progenitors of the PS was also expanded (Fig. 7U, X). Thus, forced expression of *Msgn1* in NM stem cells resulted in fewer progeny becoming neural, and more cells becoming PSM, which is consistent with *Msgn1* specifying the PSM fate.

We took advantage of *T-Cre* expression in the posterior notochord (Perantoni et al., 2005) to test whether ectopic expression of *Msgn1* could induce notochord progenitors to take up a PSM fate. *Msgn1* expression in notochord progenitors severely disrupted the expression of the notochord marker *T* (Fig. 7U–Z) in the posterior midline of *Tcre-F-Msgn1*<sup>GOF</sup> embryos treated with Dox (Fig. 7X–Z). Sections revealed that the T<sup>+</sup> midline notochord, which normally divides the adjacent *Tbx6*<sup>+</sup> PSM into two paraxial domains (Fig. 7H–J, V, W), was absent in *Tcre-F-Msgn1*<sup>GOF</sup> embryos (Fig. 7Y, Z) and was replaced by a large swath of *Msgn1*<sup>+</sup> *Tbx6*<sup>+</sup> PSM (Fig. 7K–M). Sections further revealed that the *Tcre-F-Msgn1*<sup>GOF</sup> neural tube was smaller and displayed abnormal ventral morphology (cf. Fig. 7V, Y), consistent with the absence of a floor plate due to a lack of notochord-derived ventralizing signals (Dodd et al., 1998). These results suggest that the ectopic expression of *Msgn1* suppressed notochord fate and promoted a PSM fate.

#### ***Msgn1* functions downstream of Wnt in mesoderm formation**

Wnt3a initiates PSM formation in the trunk region of the embryo. As *Msgn1* lies downstream of Wnt3a, we tested whether *Msgn1* was sufficient to induce PSM differentiation independent of Wnt activity. iF-Msgn1 and control iCre EBs were treated with the secreted Wnt antagonist Dkk1 to inhibit endogenous Wnt signaling (Fig. 8A). Dkk1 suppressed the number of *Pdgfra*<sup>+</sup> cells in uninduced cells (12.5% in samples without Dox compared with 5% in samples without Dox and with Dkk1) but had no effect on



**Fig. 5. The EMT master regulator *Snai1* is a direct target of *Msgn1*.** (A) Semi-quantitative PCR analysis of *Snai1* expression in iF-*Msgn1* EBs treated with (+) or without (-) Dox from 0–72 h. Cyclophilin A (*Ppia*) is used as a loading control. (B) *Msgn1* binds to the +5148 bp enhancer of *Snai1*. The schematic (below) shows the two E-boxes (red rectangles) identified in the *Snai1* +5148 enhancer peak. Arrowheads indicate orientation and y-axis denotes peak height, corresponding to the normalized fold enrichment. (C) Luciferase reporter assay of *Snai1* +5148 enhancer co-transfected with control or *Msgn1* expression constructs. The y-axis shows the normalized fold change; error bars show the s.d.;  $n=3$ . (D) ChIP-qPCR analysis of *Msgn1* enrichment on *Snai1* +5148 locus using control 'C' and anti-*Msgn1* 'αM' ascites in E9.5 PSM and head extracts and in iF-*Msgn1* EBs. (E–H) WISH analysis of *Snai1* mRNA in E8.5 control and *Msgn1*<sup>-/-</sup> embryos (E,G, lateral view; F,H, ventral view). Scale bars: 200 μm.

*Msgn1*-induced samples (cf. 57% in samples treated only with Dox and 55.8% in samples treated with Dox and *Dkk1*). In controls, *Dkk1* treatment suppressed *Pdgfra* irrespective of Dox treatment. Thus, *Msgn1* is capable of promoting PSM differentiation in ESCs even in the absence of Wnt activity.

To examine whether *Msgn1* was sufficient to restore PSM differentiation *in vivo* and rescue the *Wnt3a*<sup>-/-</sup> mutant phenotype, we expressed F-*Msgn1* in *Wnt3a*<sup>-/-</sup> NM stem cells using T-Cre (Fig. 8B–K). *Wnt3a*<sup>-/-</sup> mutants lack posterior paraxial mesoderm (Fig. 8B,C,E,F), do not express the somite and PSM marker *Meox1* in the posterior embryo (Fig. 8H,J) and instead display a duplicate neural tube (Fig. 8F). Forced expression of *Msgn1* restored paraxial mesoderm to *Wnt3a*<sup>-/-</sup> embryos (cf. Fig. 8F,G) and posteriorly extended the *Meox1* expression domain (cf. Fig. 8J,K), suggesting that *Msgn1* is sufficient to drive PSM differentiation in the absence of *Wnt3a*. *Msgn1* expression did not fully rescue posterior axis extension, presumably owing to a requirement for *Wnt3a* to maintain trunk progenitors. Interestingly, the duplicate neural tube was still observed despite the forced expression of *Msgn1* (Fig. 8G), suggesting that the expression of *Msgn1* alone is insufficient to suppress neural specification.

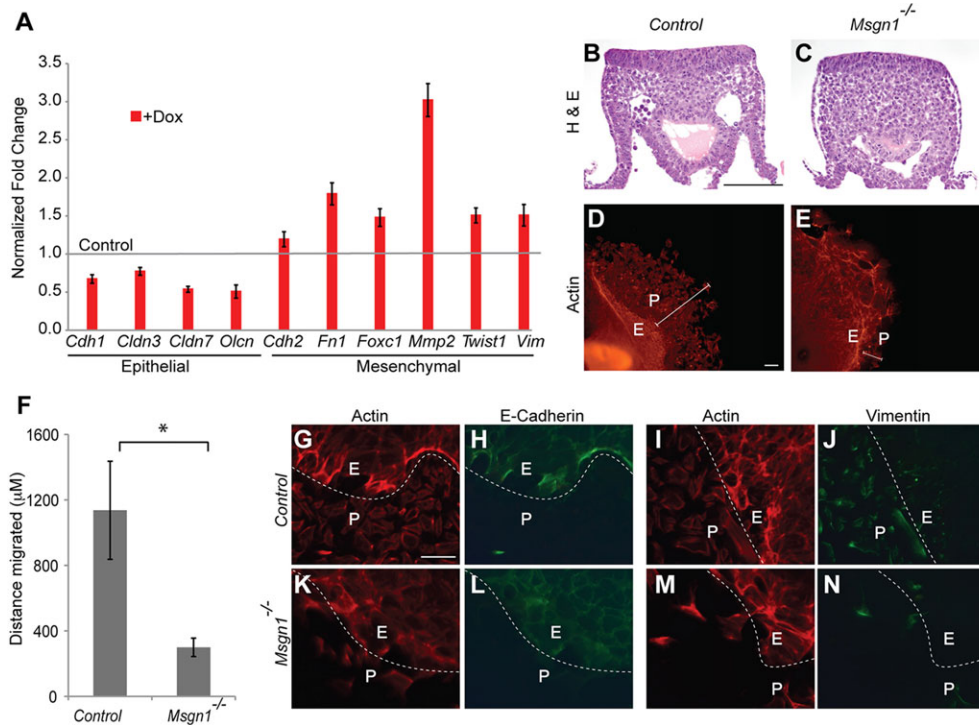
## DISCUSSION

We have shown that *Msgn1* is a master regulatory switch for paraxial PSM fate. *Msgn1* possesses the extraordinary capacity to

convert progenitor cell populations into paraxial mesoderm and is essential for paraxial mesoderm formation throughout the majority of the vertebrate trunk axis. Of the 65 total pairs of somites that form in the mouse embryo, only 8–10 develop in *Msgn1*<sup>-/-</sup> mutants, demonstrating that *Msgn1* is essential for proper trunk PSM and somite development. It is likely that the anterior somites form under the control of a separate genetic program, as suggested by the many mouse mutants that display defects in PSM formation and axis truncation but do not have defects in the anterior-most 8–10 somites (Takada et al., 1994; Chapman and Papaioannou, 1998; Galceran et al., 1999; Dunty et al., 2008, 2014; Nowotschin et al., 2012). By contrast, *Pdgfra*<sup>-/-</sup> mutants show defects in the most rostral somites (Soriano, 1997). Our demonstration that *Pdgfra* is a direct target of *Msgn1* predicts that the rostral somites should also be affected in *Msgn1*<sup>-/-</sup> embryos. The absence of a *Pdgfra*<sup>-/-</sup>-like phenotype in these mutants suggests that the rostral somites that form in *Msgn1* mutants could arise owing to redundancies between *Msgn1* and the closely related *Mesp1* and *Mesp2* bHLH factors (Yoon et al., 2000).

## Wnt3a regulates EMT and PSM cell motility through *Msgn1*

The PSM is a transitory musculoskeletal population formed from a trunk stem cell population. PSM cells are uniquely identified by their early mesodermal progenitor state and their motile mesenchymal phenotype. Our extensive gene-regulatory analysis of pathways downstream of *Msgn1* provides a unique look at how a



**Fig. 6. *Msn1* is required for EMT and motility.** (A) qPCR analysis of epithelial and mesenchymal genes expressed in Dox-treated (+Dox) iF-*Msn1* EBs after 48 h. The y-axis shows the normalized fold change; error bars are the s.d. (B,C) Histological analysis (H&E) of PS cross-sections of control ( $n=7$ ) and *Msn1*<sup>-/-</sup> ( $n=9$ ) embryos at E8.5. (D,E) Rhodamine-phalloidin staining of control (D,  $n=21$ ) and *Msn1*<sup>-/-</sup> (E,  $n=14$ ) tail bud explants dissected at the 18–22 somite stage and cultured for 48 h. The distance migrated from explant (E) to periphery (P) is indicated by white line. (F) The average distance migrated in control ( $n=13$ ) and *Msn1*<sup>-/-</sup> ( $n=7$ ) explants is shown on y-axis. \* $P \leq 9.88 \times 10^{-5}$ , two-sample unequal variance Student's *t*-test. (G–N) Immunostaining of control and *Msn1*<sup>-/-</sup> explants cultured for 48 h and stained with Rhodamine-phalloidin, and anti-E-cadherin and anti-vimentin antibodies. The dashed white line demarcates the explant (E) from periphery (P) region. Scale bars: 100 μm.

master regulator functions to activate pathways essential for the many features that define cell fate. Our data show that *Msn1* can act as a multifaceted master regulator through activation of other master regulators. Previous analyses of *Wnt3a*<sup>-/-</sup> embryos have shown that migrating mesenchymal PSM cells are missing in the mutant PS (Yoshikawa et al., 1997). *Snai1* is downregulated in *Wnt3a* and *Ctnnb1* loss-of-function mutants (this work; Dale et al., 2006) and is induced by recombinant Wnt3a treatment of ESCs (R.B.C and T.P.Y., unpublished). Our data show that *Msn1* – a Wnt3a target gene – is a major developmental regulator of *Snai1* and other EMT genes. Given that *Snai1* is a direct target of *Msn1*, the loss of *Snai1* expression in *Wnt3a* and *Ctnnb1* mutants is presumably secondary to the loss of *Msn1* expression in these mutants. Thus, *Msn1* is an excellent candidate to transduce the Wnt3a/β-catenin signal that regulates EMT and cell movement during gastrulation and axis extension. Interestingly, *Msn1* directly activates many of the EMT pathway genes that *Snai1* activates, suggesting that *Msn1* ensures the activation of the EMT and motility programs in PSM cells through a feed-forward mechanism involving *Snai1*. Notably, *Msn1*<sup>-/-</sup> mesodermal progenitors still ingress and display mesenchymal morphology in the PS, but are unable to move out of the PS. This argues that the *Msn1*-*Snai1* axis primarily functions to drive cell movement rather than EMT. The demonstration that *Msn1* regulates cell movements in zebrafish is consistent with this conclusion; however, the mechanisms appear to be substantially different because *Msn1* suppresses *snail1a* in fish (Fior et al., 2012). Alternatively, it is possible that additional Wnt3a target genes compensate for the loss of *Msn1* and initiate, but fail to complete, the EMT process.

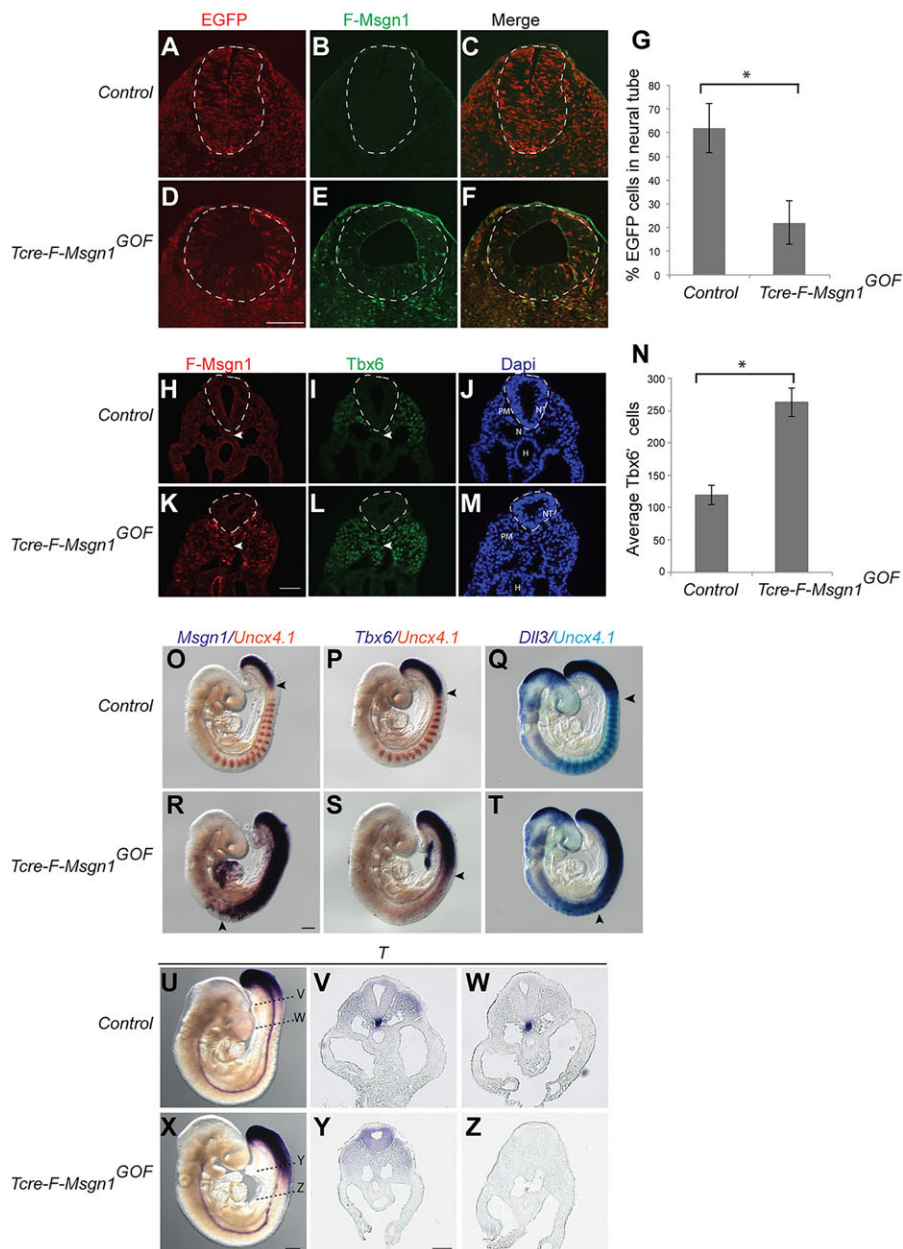
An enormous interest in EMT, and the reverse mesenchymal-epithelial transition (MET), has developed in recent years owing to the important roles that these evolutionarily conserved developmental processes play in embryogenesis, stem cell reprogramming, tumor progression and metastasis (Esteban et al., 2012; Lim and Thiery, 2012; De Craene and Berx, 2013). The only known tissue and developmental stage that *Msn1* is expressed is in the mesenchymal

cells of the PSM, making *Msn1* a remarkably stage- and tissue-specific transcription factor. Our studies demonstrate that *Msn1* regulates the EMT and PSM differentiation programs by binding to a unique E-box sequence that is distinct from the binding sites for Twist1 (Eckert et al., 2011), MyoD and NeuroD1 (Fong et al., 2012). Other non-bHLH EMT transcription factors, such as the Zn-finger homeodomain-containing factors Zeb1 and Zeb2, also function by binding to E-boxes and closely related Z-boxes (Burk et al., 2008). Interestingly, *Msn1*-binding sites can be found clustered with binding sites for these EMT transcription factors in transcriptional enhancers of EMT genes suggesting that differentiation factors and EMT transcription factors target the same enhancers. For example, *Msn1* activates *Pdgfra* expression through the same highly conserved enhancer that is targeted by Twist1 during mammary tumor metastasis (Eckert et al., 2011). In addition to binding sites for *Msn1* and Twist1, this enhancer also contains binding sites for MyoD1 and Zeb1. These results are consistent with the EMT genetic program lying embedded within broader differentiation programs, and suggest a mechanism for how transcription factors that specify unique cellular phenotypes can similarly modulate EMT. In that sense, the EMT program is highly receptive to regulation by bHLH transcription factors to promote mesenchymal cell behavior.

### ***Msn1* is a potent activator of PSM differentiation**

Careful assessment of *Msn1*<sup>-/-</sup> embryos has revealed that primitive mesodermal progenitors form but are unable to differentiate into PSM or move away from the PS. This explains the grossly enlarged tailbud observed in *Msn1*<sup>-/-</sup> mutants as being the result of the combined inability of immature mesodermal progenitors to undergo both differentiation and migration. *Tbx6* mutants also display a similarly enlarged tailbud (Chapman and Papaioannou, 1998). These findings in mice contrast with zebrafish studies where an enlarged tailbud is only seen in the absence of both *msn1* and *spadetail* suggesting perhaps that the basic regulatory networks that define the PSM phenotype has diverged to some extent between mammals and fish (Fior et al., 2012; Yabe and Takada, 2012). In mice, *Wnt3a*<sup>-/-</sup>



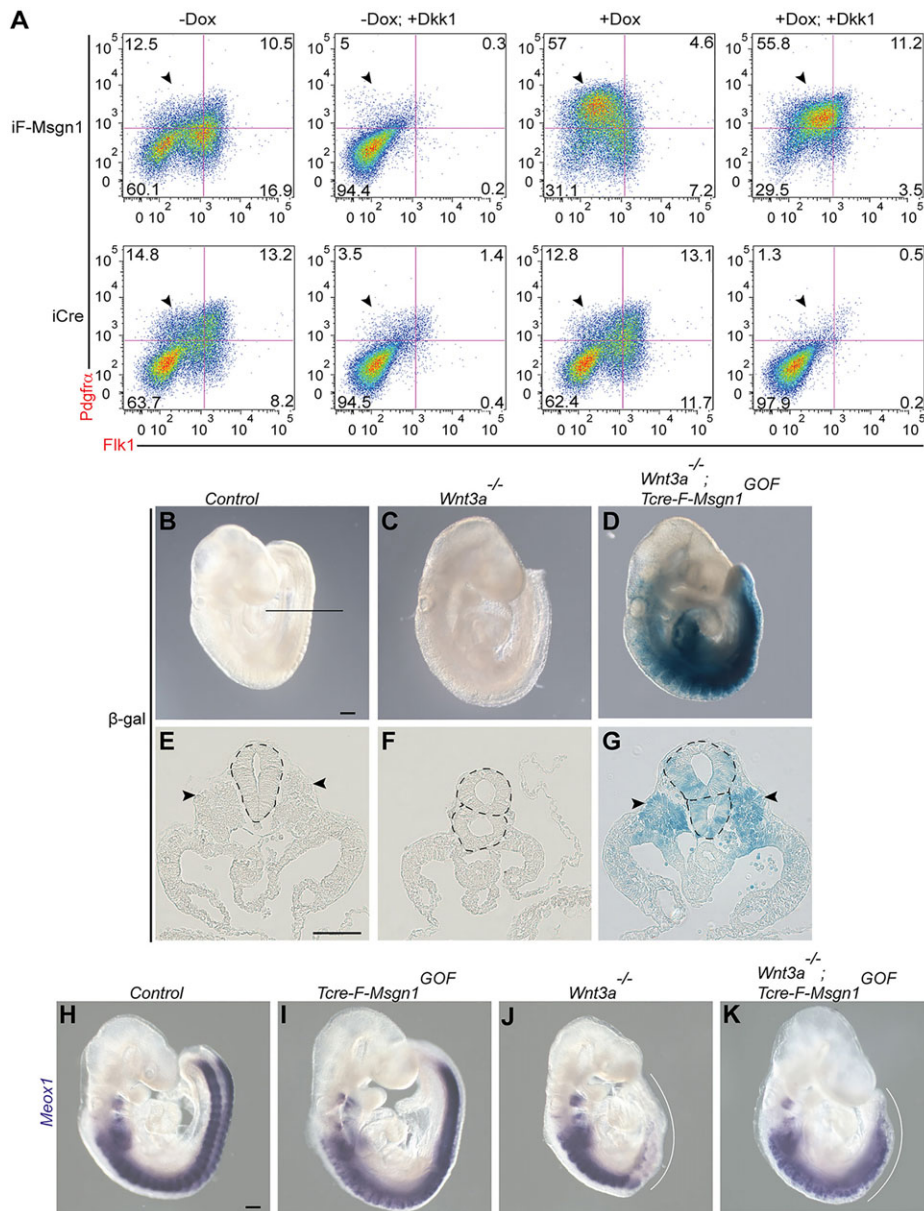


**Fig. 7. Ectopic expression of F-Msgn1 in NM stem cells excludes cells from the neural tube and specifies PSM fates.** (A–F) Immunodetection of EGFP (A, D) and F-Msgn1 (B, E) in cross-sections through the trunk neural tube (dashed lines) of E10.5 control (*T-Cre<sup>tg/+</sup>; R26-flox-rtTA-Ires-EGFP<sup>+</sup>*) (A–C) and *Tcre-F-Msgn1<sup>GOF</sup>* mutant embryos (*T-Cre<sup>tg/+</sup>; R26-flox-rtTA-Ires-EGFP<sup>+</sup>; lacZ-TRE-F-Msgn1<sup>tg/+</sup>*) (D–F). (G) Quantitative lineage tracing analysis of *T-Cre* labeled cells in neural tubes of control (A,  $n=5$ ) and *Tcre-F-Msgn1<sup>GOF</sup>* mutant (D,  $n=5$ ) are shown.  $*P \leq 2.18 \times 10^{-4}$ , two-sample unequal variance Student's *t*-test. (H–M) Immunodetection of F-Msgn1 and Tbx6 protein in cross-sections through the PSM of E9.5 control (H–J) and *Tcre-F-Msgn1<sup>GOF</sup>* mutant (K–M) embryos. NT, neural tube; PM, paraxial mesoderm; N, notochord; H, hindgut. Arrowheads denote the notochord. (N) Quantitative analysis of Tbx6<sup>+</sup> cells in control (I,  $n=5$ ) and *Tcre-F-Msgn1<sup>GOF</sup>* mutants (L,  $n=5$ ).  $*P \leq 7.38 \times 10^{-6}$ . (O–T) WISH analysis of *Msgn1* (O, R), *Tbx6* (P, S) and *Dll3* (Q, T), with the posterior somite marker *Uncx4.1* (orange in O, P, R and S; blue in Q and T) in E9.5 controls and *Tcre-F-Msgn1<sup>GOF</sup>* mutants. (U–Z) WISH of *T* mRNA expression in control (U–W) and *Tcre-F-Msgn1<sup>GOF</sup>* mutants (X–Z). V, W, cross-sections in U and Y; Z, cross-sections in X at the indicated embryonic locations. Scale bars: 100  $\mu$ m (D, Y), 50  $\mu$ m (K), 200  $\mu$ m (R, X).

and *T<sup>-/-</sup>* embryos do not form a tailbud and are completely devoid of NM stem cells and posterior mesoderm by E9.5 (Wilkinson et al., 1990; Takada et al., 1994; Yamaguchi et al., 1999; Yoon and Wold, 2000). Thus, *Wnt3a* and *T* define a class of genes necessary for mesodermal progenitor formation and maintenance, whereas *Msgn1* and *Tbx6* belong to a separate class that is necessary for mesodermal differentiation (Fig. 9). Consistent with *Msgn1* being the ultimate effector of PSM differentiation, *Msgn1* is not expressed in *Wnt3a<sup>-/-</sup>*, *T<sup>-/-</sup>* or *Tbx6<sup>-/-</sup>* embryos (Wittler et al., 2007; Chalamalasetty et al., 2011; Nowotschin et al., 2012), whereas *Wnt3a*, *T* and, at least in mesodermal progenitors, *Tbx6*, continue to be expressed in *Msgn1<sup>-/-</sup>* mouse embryos (this work; Yoon and Wold, 2000; but see Nowotschin et al., 2012). The continued expression of *Wnt3a*, *T* and *Tbx6* in *Msgn1<sup>-/-</sup>* mutants demonstrates that their expression is insufficient for PSM differentiation in the absence of *Msgn1*.

*Msgn1* promotes PSM differentiation through the direct activation of a PSM genetic program. *Msgn1* activates *Tbx6*, *Snai1*, *Foxc1*,

*Pdgfra* and *Dact1*, which are all essential for mesodermal development. Through the activation of multiple mediators of PSM cell fate, *Msgn1* broadly instills PSM characteristics on mesodermal progenitors. A key feature of this activation of PSM fate is the parallel inhibition of competing cell fates. NM stem cells have the capacity to differentiate into both neural and mesodermal cells; however, in addition to activating the PSM program, *Msgn1* also activates a program that inhibits entry into the neural cell fate. Specifically, *Msgn1* directly activates the *Tbx6* transcription factor, which is a known inhibitor of the neural determinant *Sox2*, in PSM tissue. Indeed, misexpression of *Sox2* in paraxial mesoderm is sufficient to generate ectopic neural tube formation (Takemoto et al., 2011). *Tbx6* suppresses *Sox2* and neural fates through a neural-specific N1 enhancer in the *Sox2* locus (Takemoto et al., 2011). Thus, the reprogramming of neurons to a PSM fate likely requires the expression of at least two genes, *Msgn1* to specify PSM fates and *Tbx6* to suppress neural fates. Although we have shown that *Msgn1* is



**Fig. 8. Msn1 rescues paraxial mesoderm in *Wnt3a*<sup>-/-</sup> mutants.** (A) Flow cytometry analysis of iF-Msgn1 or control iCre EBs treated with (+) or without (-) Dkk1 and/or Dox for 48 h for the expression of Pdgfra and Flk1. The arrowheads indicate Pdgfra<sup>+</sup> cells. (B-G)  $\beta$ -Gal staining of E9.5 control (B), *Wnt3a*<sup>-/-</sup> (C), and *Wnt3a*<sup>-/-</sup>; *Tcre-F-Msgn1*<sup>GOF</sup> (D) mutants. (E-G) These panels represent corresponding cross sections of embryos depicted in B-D, taken at the indicated location (black line, B). Arrowheads indicate paraxial mesoderm. (H-K) WISH analysis of *Meox1* in control (H), *Tcre-F-Msgn1*<sup>GOF</sup> (I), *Wnt3a*<sup>-/-</sup> (J), and *Wnt3a*<sup>-/-</sup>; *Tcre-F-Msgn1*<sup>GOF</sup> (K) mutants. Curved white line in J, K highlights the rescued posterior paraxial mesoderm domain. Scale bars: 200  $\mu$ m (B,H), 100  $\mu$ m (E).

sufficient to activate *Tbx6* transcription in luciferase reporter assays in heterologous cell lines, it is apparently insufficient in neurons to reprogram neural fates or requires higher levels of expression than we were able to achieve experimentally. Nevertheless, given the powerful ability of Msn1 to function as a transcriptional regulator, it will be interesting to test whether Msn1 is capable of directly reprogramming human somatic cells into musculoskeletal stem cells (i.e. PSM), given that the PSM gives rise to several cell types, including skeletal muscle, dermis, bone and cartilage, that could be used in regenerative therapies for aging or musculoskeletal disease.

## MATERIALS AND METHODS

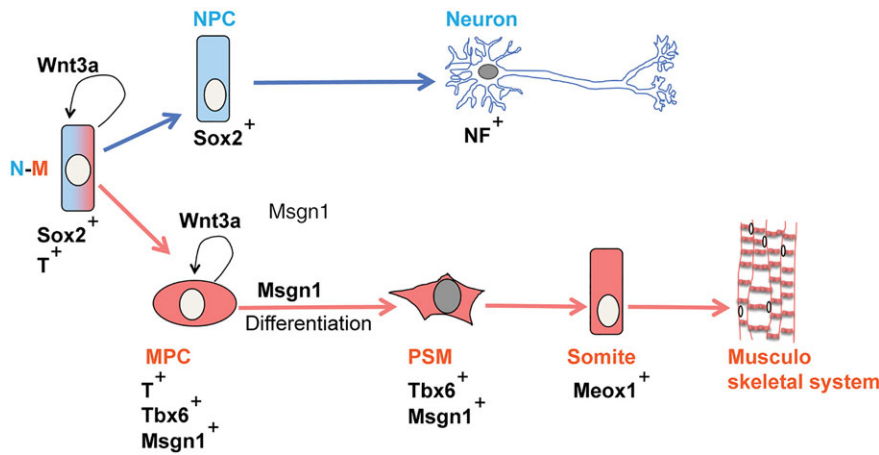
### Mice and embryos

The *Wnt3a*<sup>tm1Amc</sup> (Takada et al., 1994), *Msn1*<sup>tm1Wb</sup> (Yoon and Wold, 2000), *Ctnnb1*<sup>tm2Kem</sup> (Brault et al., 2001), *Ctnnb1*<sup>tm1Mmt</sup> (Harada et al., 1999), *Gt(Rosa)26Sor*<sup>tm1Sor</sup> (Soriano, 1999), *B6.Cg-Gt(ROSA)26Sortm1<sup>trTA,EGFP</sup>Nagy/J* (Belteki et al., 2005) and *Tg(T-cre)1Lwd* (Perantoni et al., 2005) mice were used to generate the allelic combinations featured in this study. For conditional expression of Flag-tagged Msn1 in mice, *LacZ-TRE-F-Msgn1*<sup>Tg/+</sup> transgenic

mice were generated by standard pronuclear injection methods. Detailed experimental details on the generation and characterization of *LacZ-TRE-F-Msgn1*<sup>Tg/+</sup> mice can be found in the methods in the supplementary materials. Embryos were dissected at E8.5 (4–8 somites) and E9.5 (18–22 somites). All animal experiments were performed in accordance with the guidelines established by the NCI-Frederick Animal Care and Use Committee.

### Whole-mount *in situ* hybridization, $\beta$ -gal staining and immunohistochemistry

Whole-mount *in situ* hybridization (WISH) experiments and probe synthesis (*Msn1*, *Tbx6*, *Snai1*, *Meox1*, *T*, *Pdgfra*, *Uncx4.1* and *Dll3*) were carried out as described previously (Biris and Yamaguchi, 2014). All observed phenotypes were reproduced in at least three separate embryos unless otherwise specified.  $\beta$ -gal staining of cells and embryos was performed according to standard protocols (Whiting et al., 1991). For immunodetection on frozen sections, mouse embryos were fixed in 4% paraformaldehyde in PBS for 30 min at room temperature, incubated in 20% sucrose overnight at 4°C, embedded in OCT, and sectioned to a thickness of 5  $\mu$ m. Immunodetection was performed as described previously (Dunty et al., 2014). Detailed experimental details can be found in the methods in the supplementary materials.



**Fig. 9. *Msgn1* is a master regulator of PSM differentiation.**

Bipotential NM stem cells, defined by Sox2 and T expression, generate the progenitors of the trunk spinal cord (NPC, neural progenitor cells) and musculoskeletal system (MPC, mesodermal progenitor cells). Wnt3a maintains both the NM stem cell and the MPC. In our model, the Wnt3a/ $\beta$ -catenin pathway activates the target genes *T* and *Msgn1* in NM stem cells in the PS to specify paraxial mesoderm fate (Yamaguchi et al., 1999; Dunty et al., 2008; Chalamalasetty et al., 2011). *Msgn1* initiates the genetic programs that define PSM identity and cell behavior by activating downstream master regulators of EMT and motility (*Snai1*) and neural suppression (*Tbx6*) among others. *Msgn1* and *Tbx6* expression must be extinguished for PSM cells to undergo MET and segmentation to form somites and, ultimately, to differentiate into skeletal muscle, bone, cartilage, tendons, endothelial cells and dermis.

### Tailbud explant migration assay

E9.5 18–22 ss embryos were dissected at room temperature in PBS. Excised tailbuds extended anteriorly from the tail tip up to, but not including, the hindgut. Isolated explants were transferred by pipette onto fibronectin-coated eight-well  $\mu$ -slide dishes (Ibidi) containing 300  $\mu$ l of Dulbecco's modified Eagle's medium (DMEM) and 10% fetal bovine serum (FBS) (Gibco), and were cultured for 48 h at 37°C in 5% CO<sub>2</sub>. For immunocytochemistry analysis, explants were fixed in 4% PFA in PBS for 15 min at room temperature, washed with PBS and processed for antibody staining. Pilot experiments revealed that after the initial attachment and spreading of the explant on the dish, the tight cell-cell contacts visualized by intense phalloidin and E-cadherin staining remained intact in the ectoderm over the 48 h culture and readily defined the explant boundaries. Mesenchymal cells migrated radially away from the explant as individual cells and were defined by a low-level of phalloidin staining, and being negative for E-cadherin and positive for vimentin staining.

### Cloning, ESC culture and differentiation

1 $\times$ Flag-*Msgn1*, 3 $\times$ Flag-*Msgn1*-Ires2-AcGFP and 3 $\times$ Flag-*Tbx6*-Ires2-AcGFP were cloned into p2lox vector to generate iF-*Msgn1* or iF-*Tbx6* ESCs. Site-specific recombination in A2lox Cre ESCs was achieved as described previously (Iacovino et al., 2011). ESCs were cultured on mitomycin C-treated mouse embryo fibroblasts (MEFs) in DMEM, 15% FBS, 1 $\times$  MEM non-essential amino acids, penicillin-streptomycin (all Life Technologies), 10<sup>3</sup> units/ml ESGRO (Chemicon) and 100  $\mu$ M  $\beta$ -mercaptoethanol. For differentiation, ESCs were separated from MEFs and grown in suspension cultures for 2 days to form EBs. Dox was added on day 2 to induce *Msgn1* and samples were harvested from 6–72 h post-induction, depending on the experiment. Detailed experimental details can be found in the methods in the supplementary materials.

### EMSA assays

0.8 ng of DIG-labeled oligonucleotides (DIG Gel Shift Assay Kit, Roche, catalog number 03353591910) were incubated with *Msgn1* and DNA-protein complexes were migrated on 6% DNA retardation gels (Invitrogen) and analyzed on Hybond-N+ membrane (Amersham, catalog number RPN303B) according to manufacturer's instructions. Detailed experimental details can be found in the methods in the supplementary materials.

### Luciferase assays

200 ng of luciferase reporter and 5 ng of *Renilla* constructs were co-transfected with 50 ng of expression construct per well in 24-well plates. Luciferase assays were performed according to Dual Luciferase Assay Kit (Promega) instructions. Detailed experimental details can be found in the methods in the supplementary materials.

### Reverse transcription and qPCR

RNA isolation, reverse transcription and qPCR were set up as described previously (Chalamalasetty et al., 2011). Total RNA was isolated using RNeasy mini kit (Qiagen), reverse transcription and qPCR were set up using iScript cDNA synthesis kit and SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), respectively. Oligos used in this study are described in supplementary material Table S5. Detailed experimental details can be found in the methods in the supplementary materials.

### Flow cytometry analysis

EBs were washed in PBS and briefly trypsinized with 0.25% Trypsin-EDTA for 1 min to dissociate into single cells. After neutralizing in DMEM, 10% FBS medium, cells were washed in ice-cold SORT buffer (PBS, penicillin-streptomycin, 0.2% Fraction-V BSA) and stained with antibodies for 20 min on ice. Cells were washed twice with SORT buffer and fixed in 1% PFA. For analysis, cells were passed through cell strainer and analyzed using a BD FACS machine. Raw data was imported to FlowJo software for extensive analysis. Antibodies used in flow cytometry are described in Table S4 in the supplementary materials.

### Gene expression profiling, ChIP-qPCR and ChIP-Seq studies

For microarray analysis, day 2 EBs formed from iF-*Msgn1* ESC were treated with 1  $\mu$ g/ml of Dox and samples were subsequently collected at 12 h, 24 h, and 48 h time points. Microarray and ChIP-seq analyses were performed as described previously (Chalamalasetty et al., 2011). Briefly, ChIP-seq was performed on day 2 EBs treated with Dox for an additional 36 h, fixed in 1% formaldehyde solution for 15 min, quenched with 0.125 M glycine and flash-frozen. Nuclear extracts were incubated with anti-Flag M2 antibody (Sigma), DNA fragments were purified and amplified using Illumina ChIP-seq DNA sample prep kit (Illumina), and DNA libraries were sequenced on a Genome Analyzer II. 35-nt sequence reads were mapped to the mouse genome (build mm9) using the Eland algorithm. Validation of ChIP-seq peaks was done by ChIP-qPCR on iF-*Msgn1* EBs and E9.5 PSM extracts using control (Clone: NS-1; Sigma) and anti-*Msgn1* ascites (Chalamalasetty et al., 2011) as described previously. Detailed experimental details can be found in the methods in the supplementary materials.

### Accession numbers

Microarray and ChIP-seq data reported in this paper are deposited in NCBI GEO database with accession numbers GSE29848 and GSE55263, respectively.

### Acknowledgements

We are grateful to J. Yoon and M. Lewandoski for *Msgn1* and T-Cre mice, respectively, and to M. Kyba for A2lox ESCs reagents. We thank L. Feigenbaum for

generating transgenic animals, Ruth Wolfe for assistance with animal husbandry, and the Flow Cytometry Core at NCI-Frederick for assistance with FACS experiments. We are grateful to A. Perantoni and M. Lewandoski for critiquing the manuscript.

#### Competing interests

The authors declare no competing financial interests.

#### Author contributions

R.B.C. and T.P.Y. conceived the project. R.B.C. and T.P.Y. designed and analyzed the experiments. R.J.G. performed the immunohistochemistry experiments in supplementary material Fig. S4 and *lacZ* sections in Fig. 8. W.C.D., P.J. and H.S. assisted R.B.C. with genomics and analysis. M.W.K. performed some of the luciferase assays. T.P.Y., R.B.C. and R.J.G. wrote the manuscript.

#### Funding

This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. Deposited in PMC for release after 12 months.

#### Supplementary material

Supplementary material available online at

<http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.110908/-/DC1>

#### References

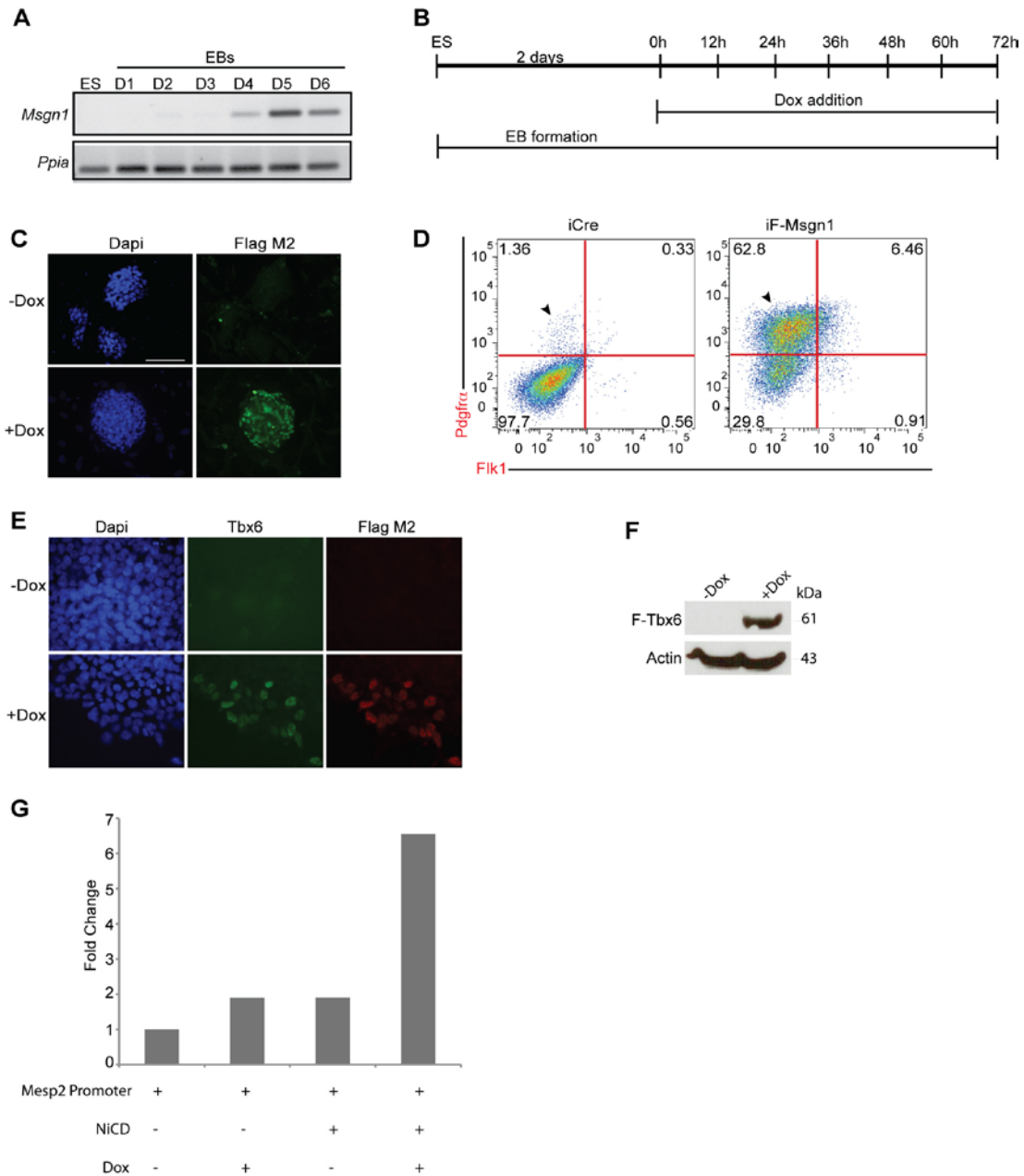
- Barrallo-Gimeno, A. and Nieto, M. A. (2005). The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* **132**, 3151–3161.
- Battle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J. and García De Herreros, A. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat. Cell Biol.* **2**, 84–89.
- Belteki, G., Haigh, J., Kabacs, N., Haigh, K., Sison, K., Costantini, F., Whitsett, J., Quaggin, S. E. and Nagy, A. (2005). Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. *Nucleic Acids Res.* **33**, e51.
- Biris, K. K. and Yamaguchi, T. P. (2014). Two-color in situ hybridization of whole-mount mouse embryos. *Methods Mol. Biol.* **1092**, 17–30.
- Braut, V., Moore, R., Kutsch, S., Ishibashi, M., Rowitch, D. H., McMahon, A. P., Sommer, L., Boussadia, O. and Kemler, R. (2001). Inactivation of the beta-catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development* **128**, 1253–1264.
- Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S. and Brabletz, T. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep.* **9**, 582–589.
- Cambay, N. and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. *Development* **134**, 2829–2840.
- Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F. and Nieto, M. A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat. Cell Biol.* **2**, 76–83.
- Chalamalasetty, R. B., Dunty, W. C., Jr, Biris, K. K., Ajima, R., Iacovino, M., Beisaw, A., Feigenbaum, L., Chapman, D. L., Yoon, J. K., Kyba, M. et al. (2011). The Wnt3a/beta-catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. *Nat. Commun.* **2**, 390.
- Chapman, D. L. and Papaioannou, V. E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695–697.
- Dale, J. K., Malapert, P., Chal, J., Vilhais-Neto, G., Maroto, M., Johnson, T., Jayasinghe, S., Trainor, P., Herrmann, B. and Pourquié, O. (2006). Oscillations of the snail genes in the presomitic mesoderm coordinate segmental patterning and morphogenesis in vertebrate somitogenesis. *Dev. Cell* **10**, 355–366.
- De Craene, B. and Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. *Nat. Rev. Cancer* **13**, 97–110.
- Dequéant, M. L. and Pourquié, O. (2008). Segmental patterning of the vertebrate embryonic axis. *Nat. Rev. Genet.* **9**, 370–382.
- Dodd, J., Jessell, T. M. and Placzek, M. (1998). The when and where of floor plate induction. *Science* **282**, 1654–1657.
- Dunty, W. C., Jr, Biris, K. K., Chalamalasetty, R. B., Taketo, M. M., Lewandoski, M. and Yamaguchi, T. P. (2008). Wnt3a/beta-catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation. *Development* **135**, 85–94.
- Dunty, W. C., Jr, Kennedy, M. W. L., Chalamalasetty, R. B., Campbell, K. and Yamaguchi, T. P. (2014). Transcriptional profiling of Wnt3a mutants identifies Sp transcription factors as essential effectors of the Wnt/beta-catenin pathway in neuromesodermal stem cells. *PLoS ONE* **9**, e87018.
- Eckert, M. A., Lwin, T. M., Chang, A. T., Kim, J., Danis, E., Ohno-Machado, L. and Yang, J. (2011). Twist1-induced invadopodia formation promotes tumor metastasis. *Cancer Cell* **19**, 372–386.
- Ema, M., Takahashi, S. and Rossant, J. (2006). Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. *Blood* **107**, 111–117.
- Esteban, M. A., Bao, X., Zhuang, Q., Zhou, T., Qin, B. and Pei, D. (2012). The mesenchymal-to-epithelial transition in somatic cell reprogramming. *Curr. Opin. Genet. Dev.* **22**, 423–428.
- Fior, R., Maxwell, A. A., Ma, T. P., Vezzano, A., Moens, C. B., Amacher, S. L., Lewis, J. and Saude, L. (2012). The differentiation and movement of presomitic mesoderm progenitor cells are controlled by Mesogenin 1. *Development* **139**, 4656–4665.
- Fong, A. P., Yao, Z., Zhong, J. W., Cao, Y., Ruzzo, W. L., Gentleman, R. C. and Tapscott, S. J. (2012). Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev. Cell* **22**, 721–735.
- Fujimoto, H. and Yanagisawa, K. O. (1979). Effect of the T-mutation on histogenesis of the mouse embryo under the testis capsule. *J. Embryol. Exp. Morphol.* **50**, 21–30.
- Galceran, J., Farinas, I., Depew, M. J., Clevers, H. and Grosschedl, R. (1999). Wnt3a-like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. *Genes Dev.* **13**, 709–717.
- Guaita, S., Puig, I., Franci, C., Garrido, M., Domínguez, D., Battle, E., Sancho, E., Dedhar, S., de Herreros, A. G. and Baulida, J. (2002). Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression. *J. Biol. Chem.* **277**, 39209–39216.
- Harada, N., Tamai, Y., Ishikawa, T., Sauer, B., Takaku, K., Oshima, M. and Taketo, M. M. (1999). Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. *EMBO J.* **18**, 5931–5942.
- Iacovino, M., Bosnakovski, D., Fey, H., Rux, D., Bajwa, G., Mahen, E., Mitanoska, A., Xu, Z. and Kyba, M. (2011). Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* **29**, 1580–1588.
- Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biol.* **5**, 226.
- Lim, J. and Thiery, J. P. (2012). Epithelial-mesenchymal transitions: insights from development. *Development* **139**, 3471–3486.
- Lindsley, R. C., Gill, J. G., Murphy, T. L., Langer, E. M., Cai, M., Mashayekhi, M., Wang, W., Niwa, N., Nerbonne, J. M., Kyba, M. et al. (2008). Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. *Cell Stem Cell* **3**, 55–68.
- McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501.
- Mugford, J. W., Sipilä, P., McMahon, J. A. and McMahon, A. P. (2008). Osr1 expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an Osr1-dependent nephron progenitor compartment within the mammalian kidney. *Dev. Biol.* **324**, 88–98.
- Nicolas, J. F., Mathis, L., Bonnerot, C. and Saurin, W. (1996). Evidence in the mouse for self-renewing stem cells in the formation of a segmented longitudinal structure, the myotome. *Development* **122**, 2933–2946.
- Nowotzschin, S., Ferrer-Vaquer, A., Concepcion, D., Papaioannou, V. E. and Hadjantonakis, A.-K. (2012). Interaction of Wnt3a, Msn1 and Tbx6 in neural versus paraxial mesoderm lineage commitment and paraxial mesoderm differentiation in the mouse embryo. *Dev. Biol.* **367**, 1–14.
- Palmer, M. B., Majumder, P., Green, M. R., Wade, P. A. and Boss, J. M. (2007). A 3' enhancer controls snail expression in melanoma cells. *Cancer Res.* **67**, 6113–6120.
- Perantoni, A. O., Timofeeva, O., Naillat, F., Richman, C., Pajni-Underwood, S., Wilson, C., Vainio, S., Dove, L. F. and Lewandoski, M. (2005). Inactivation of FGF8 in early mesoderm reveals an essential role in kidney development. *Development* **132**, 3859–3871.
- Sakurai, H., Era, T., Jakt, L. M., Okada, M., Nakai, S., Nishikawa, S. and Nishikawa, S.-I. (2006). In vitro modeling of paraxial and lateral mesoderm differentiation reveals early reversibility. *Stem Cells* **24**, 575–586.
- Soriano, P. (1997). The PDGF alpha receptor is required for neural crest cell development and for normal patterning of the somites. *Development* **124**, 2691–2700.
- Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21**, 70–71.
- Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., McMahon, J. A. and McMahon, A. P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev.* **8**, 174–189.
- Takebe, A., Era, T., Okada, M., Jakt, L. M., Kuroda, Y. and Nishikawa, S.-I. (2006). Microarray analysis of PDGFR alpha+ populations in ES cell differentiation culture identifies genes involved in differentiation of mesoderm and mesenchyme including ARID3b that is essential for development of embryonic mesenchymal cells. *Dev. Biol.* **293**, 25–37.
- Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E. and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394–398.
- Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. and Nicolas, J.-F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* **17**, 365–376.

- Wehn, A. K. and Chapman, D. L.** (2010). Tbx18 and Tbx15 null-like phenotypes in mouse embryos expressing Tbx6 in somitic and lateral plate mesoderm. *Dev. Biol.* **347**, 404–413.
- Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W., Stott, D. and Allemann, R. K.** (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of Hox-2.6 gene expression. *Genes Dev.* **5**, 2048–2059.
- Wilkinson, D. G., Bhatt, S. and Herrmann, B. G.** (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* **343**, 657–659.
- Wilson, V., Olivera-Martinez, I. and Storey, K. G.** (2009). Stem cells, signals and vertebrate body axis extension. *Development* **136**, 1591–1604.
- Wittler, L., Shin, E.-h., Grote, P., Kispert, A., Beckers, A., Gossler, A., Werber, M. and Herrmann, B. G.** (2007). Expression of Msn1 in the presomitic mesoderm is controlled by synergism of WNT signalling and Tbx6. *EMBO Rep.* **8**, 784–789.
- Yabe, T. and Takada, S.** (2012). Mesogenin causes embryonic mesoderm progenitors to differentiate during development of zebrafish tail somites. *Dev. Biol.* **370**, 213–222.
- Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. and McMahon, A. P.** (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev.* **13**, 3185–3190.
- Yoon, J. K. and Wold, B.** (2000). The bHLH regulator pMesogenin1 is required for maturation and segmentation of paraxial mesoderm. *Genes Dev.* **14**, 3204–3214.
- Yoon, J. K., Moon, R. T. and Wold, B.** (2000). The bHLH class protein pMesogenin1 can specify paraxial mesoderm phenotypes. *Dev. Biol.* **222**, 376–391.
- Yoshikawa, Y., Fujimori, T., McMahon, A. P. and Takada, S.** (1997). Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. *Dev. Biol.* **183**, 234–242.

**SUPPLEMENTARY MATERIAL**

**SUPPLEMENTARY FIGURES**

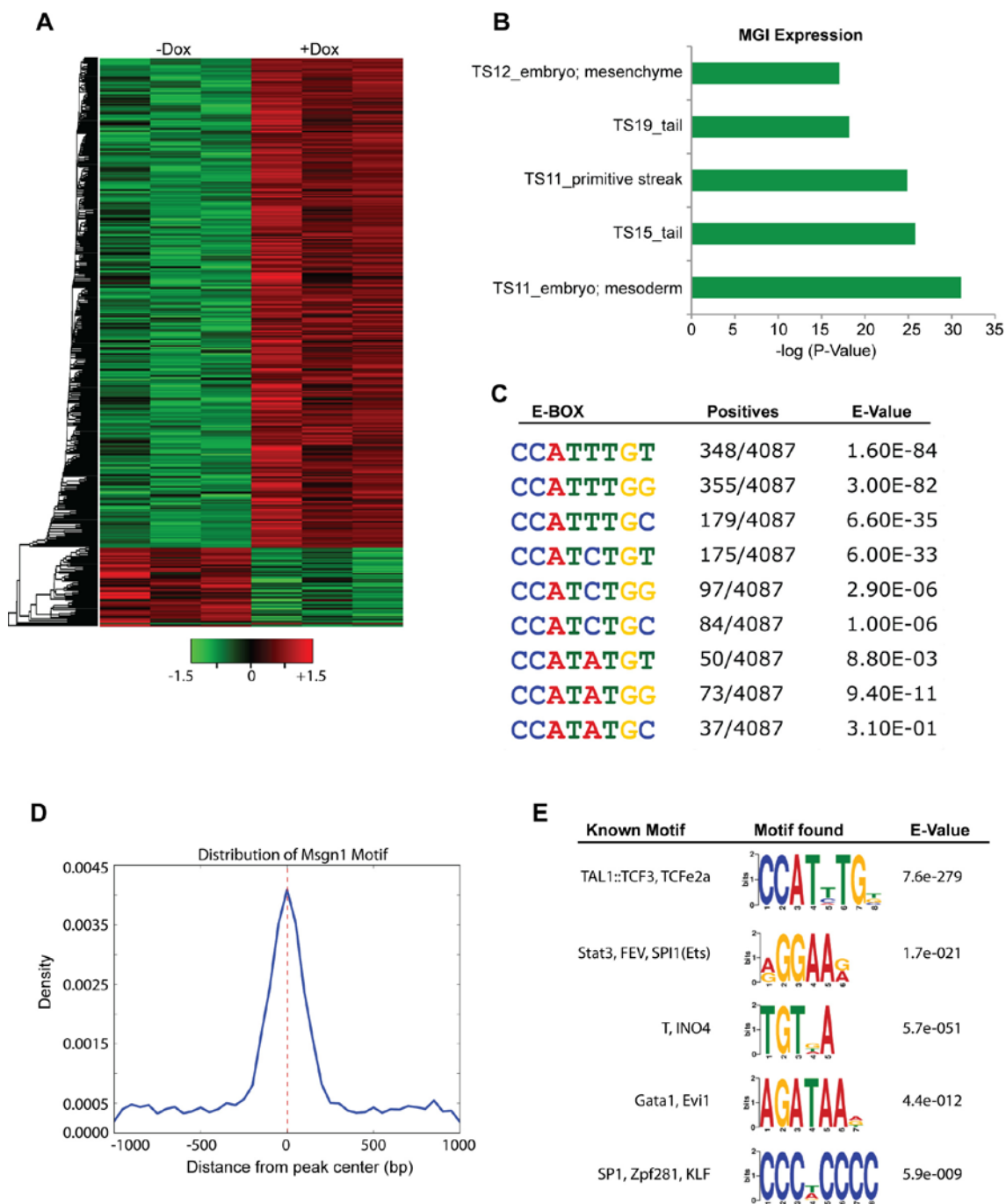
Supplementary material Fig.S1 (Related to Fig.1)



**Fig. S1. Overexpression of *Msgn1* is sufficient to induce a *Pdgfra*-positive population.** (A) RT-PCR analysis of endogenous *Msgn1* and *Ppia* (loading control) expression in serum free EBs from Day 0 (ES) to Day 6 (D6) of differentiation. (B) Schematic view of ESCs to EB differentiation timeline. (C) iF-*Msgn1* EBs treated with (+) or without (-) Dox were stained with DAPI and anti-Flag M2 antibody. Scale bars: 50

$\mu\text{m}$ . (D) Flow cytometry analysis for the expression of cell surface markers *Pdgfra* and *Flk1* in control (iCre EBs) and iF-*Msgn1* EBs induced for 12 h. Arrowheads indicate the *Pdgfra* single positive population in that quadrant. (E) iF-*Tbx6* ES cells grown on MEFs treated with (+) or without (–) Dox were stained with DAPI and anti-Flag M2 or anti-*Tbx6* antibodies. (F) Western blot analysis of F-*Tbx6* protein expression in dox treated iF-*Tbx6* EBs for 48 h. Blots were probed with anti-Flag M2 and anti-Actin antibodies (loading control). (G) A functional luciferase assay to test the activity of F-*Tbx6* protein in differentiating iF-*Tbx6* ESCs. Differentiating iF-*Tbx6* ESCs were co-transfected with *Mesp2*-promoter luciferase reporter (Yasuhiko et al., 2008), NICD (Notch Intracellular Domain) expression constructs and simultaneously treated with Dox to induce F-*Tbx6* expression. F-*Tbx6* and NiCD showed synergistic activation of the *Mesp2* reporter. Y-axis is normalized fold change.

Supplementary material Fig.S2 (Related to Fig.2 and Fig.3)

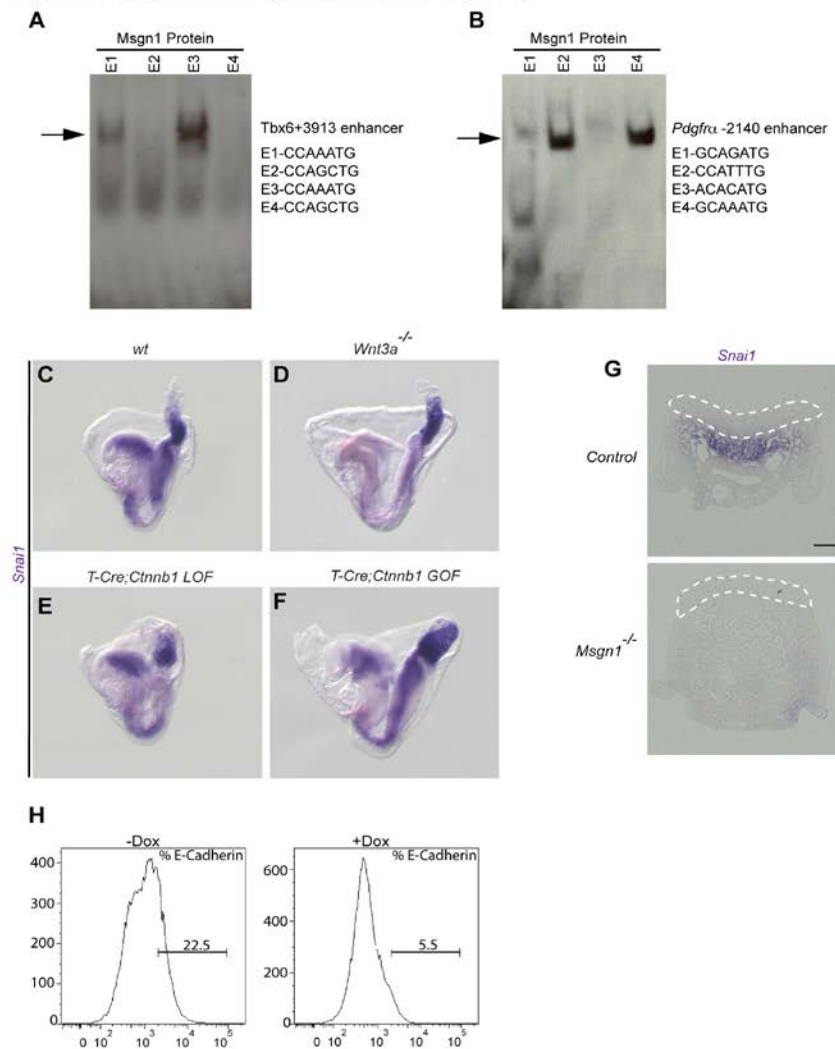


**Fig. S2. Microarray and ChIP-seq analysis of Msgn1-binding sites.** (A) Hierarchical clustering of 443 differentially expressed Affymetrix IDs or 332 differentially expressed unique genes at  $P \leq 0.05$ , Fold change  $\geq 1.5$  upon induction of F-Msgn1 with (+) or without (-) Dox for 24 h. Expression of upregulated genes (red) and downregulated genes (green)



were shown. About 85% of the genes were upregulated upon induction of *Msgn1* suggesting that *Msgn1* is a transcriptional activator. Color bar represents  $\log_2$  intensity values. (B) Representative annotations of Mouse Genome Informatics (MGI) expression analysis of F-*Msgn1* peaks discovered using GREAT algorithm. X-axis shows the *P*-value ( $-\log_{10}$ ). (C) Enriched matching words for the top motif 'CCATHTGB' identified by DREME analysis. Positives indicate the number of positive sequences with matches to the word and E-value indicates the *P*-value times the number of candidates tested. (D) Density graph of *Msgn1* binding motif CCATHTGB in relation to the distance from peak center. Here, *Msgn1* motif from Fig. 3E was mapped to F-*Msgn1* peaks, and the motif locations from peak center were plotted against the motif occurrence density. (E) The top 5 motifs identified by DREME analysis of *Msgn1* binding peaks include a motif that is similar to TAL1/TCF3 bHLH transcription factors. The other motifs identified were similar to Stat3/FEV/Ets, Brachyury (T)/INO4, Gata1/Evi1, and Sp1/Zpf281/KLF factors.

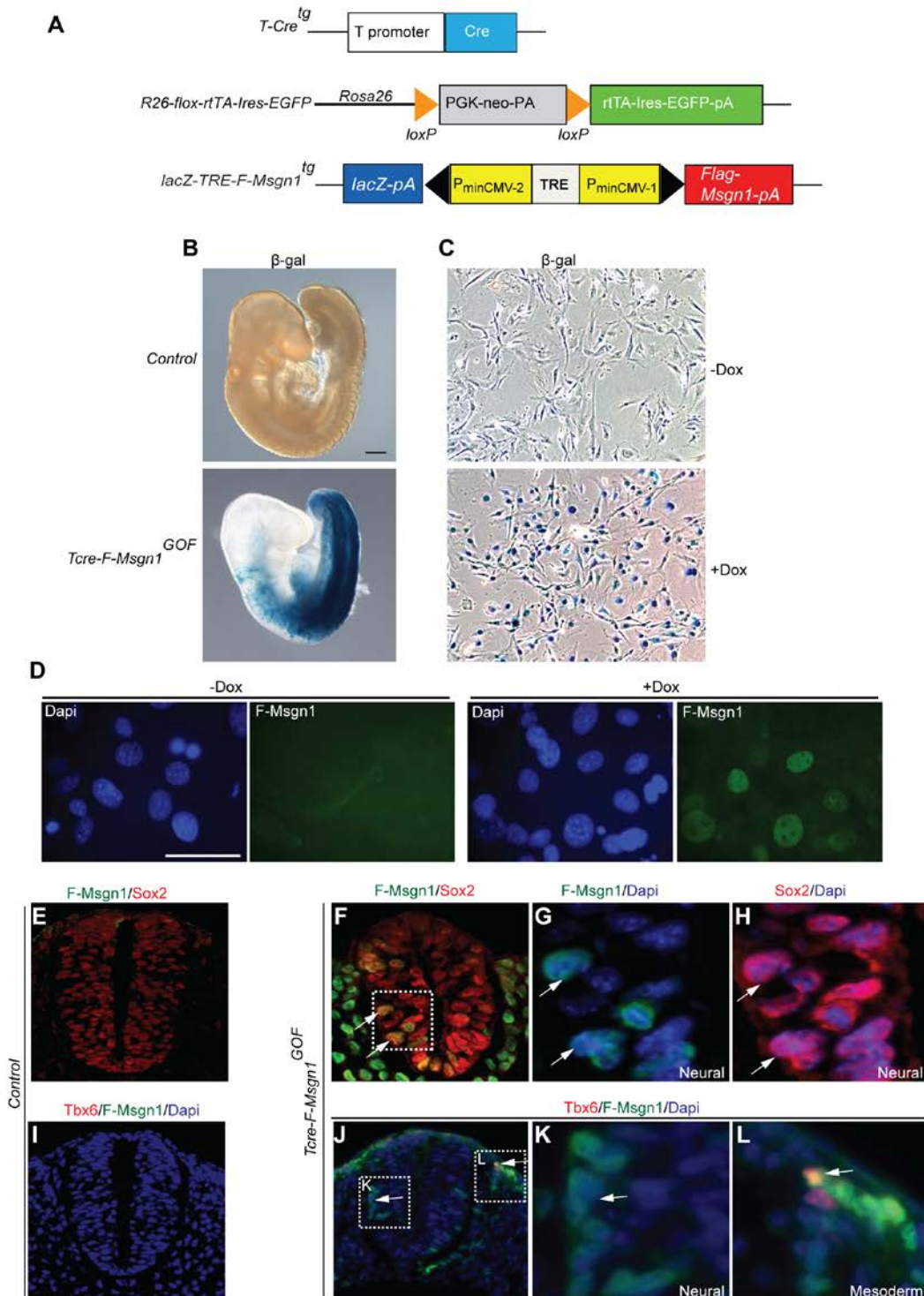
Supplementary material Fig.S3 (Related to Figs.4-6)



**Fig. S3. Biochemical and genetic analyses of *Tbx6*, *Pdgfra* and *Snai1* regulation by *Msgn1*.** (A) Electrophoretic Mobility Gel Shift Analysis (EMSA) of *Msgn1* binding to DIG-labeled oligos encompassing E boxes E1-E4 of *Tbx6*+3913 enhancer. *Msgn1* only binds to specific E-box sequence ‘CATTG’. Arrow indicates the gel shift of DNA-protein complex. (B) Binding of *Msgn1* protein to E-box sequences E1-E4 of *Pdgfra*-2140 enhancer. Arrow shows protein-DNA complexes. (C-F) WISH analysis of *Snai1* in

E8.5 *WT* (C), *Wnt3a*<sup>-/-</sup> (D), *T-Cre; Ctnnb1 LOF* ( $\beta$ -catenin loss-of-function) (E) and *T-Cre; Ctnnb1 GOF* embryos ( $\beta$ -catenin gain-of-function) (F). (G) PS cross-sections of *Snail* WISH in *WT* and *Msgn1*<sup>-/-</sup> embryos from Fig. 5E-H are shown. Dotted white line indicates the epithelial PS. *Snail* expression is down regulated in the mesodermal domain of the PS of *Msgn1*<sup>-/-</sup> mutants compared to control embryos. Scale bar: 100  $\mu$ m. (H) iF-*Msgn1* EBs were differentiated with (+) or without (-) Dox and flow cytometry analysis was performed for E-cadherin expressing cells. E-cadherin positive cells decreased from 22.5% to 5.5% in F-*Msgn1* expressing EBs (+Dox) compared to untreated (-Dox).

Supplementary material Fig.S4 (Related to Fig.7)

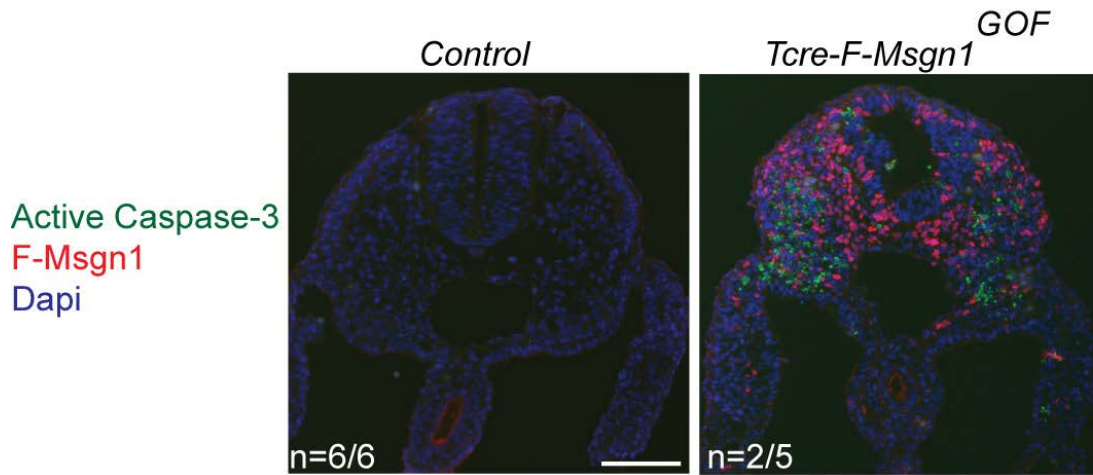


**Fig. S4. Triple transgenic mouse strategy for inducible expression of *Msgn1* *in vivo*.**

(A) Schematic view of triple transgenic mouse strategy used to ectopically express F-

Msgn1. Transgenic mouse lines *T-cre<sup>tg</sup>*, *R26-flox-rtTA-Ires-EGFP* and *LacZ-TRE-F-Msgn1<sup>tg</sup>* were used for ectopic expression of F-Msgn1 in the *T-cre* domain. *R26-rtTA-Ires-EGFP* mice were crossed to transgenic line *LacZ-TRE-F-Msgn1<sup>tg</sup>* to generate *R26-flox-rtTA-Ires-EGFP; LacZ-TRE-F-Msgn1* which were then crossed to *T-cre* mice. To induce expression of F-Msgn1 in the *T* domain, pregnant females were fed with Dox water from E6.5. (B) E9.5 littermate control and *Tcre-F-Msgn1<sup>GOF</sup>* mutants treated with Dox were processed for  $\beta$ -gal staining. Scale bars: 200  $\mu$ m. (C, D) Fibroblasts derived from E13.5 constitutively expressing *F-Msgn1<sup>GOF</sup>* embryos treated with (+) or without (-) Dox for 24 h were  $\beta$ -gal stained (C) or labeled with dapi and anti-Flag antibodies (D). Scale bars: 50  $\mu$ m. (E-H) F-Msgn1 and Sox2 protein expression was assessed by immunofluorescence performed on sections of the neural tube of control and *Tcre-F-Msgn1<sup>GOF</sup>* mutants taken from the trunk region. Arrows indicate F-Msgn1 expressing cells in the neural tube that co-express Sox2. Dotted squares in F represent magnified views depicted in G and H. (I-L) Detection of F-Msgn1 and Tbx6 in F-Msgn1 expressing embryos showing absence of ectopic Tbx6 expression in neural cells expressing F-Msgn1 (J, K, arrows). Ectopic Tbx6 expression was detected in F-Msgn1 expressing cells in trunk somitic mesoderm (J, L, arrows). Left dotted square in J depicts the magnified view shown in K, right dotted square in J is represented in L.

## Supplementary material Fig.S5 (Related to Fig.7)



**Fig. S5. Cell death analysis in *Tcre-F-Msgn1<sup>GOF</sup>* mutants.**

Immunostaining of active Caspase-3 and F-Msgn1 in cross-sections of control and *Tcre-F-Msgn1<sup>GOF</sup>* mutants along the trunk region of E 9.5 embryos. In 3/5 mutants, Caspase-3 expression was not observed. In 2/5 mutants analyzed, active Caspase-3 is predominantly localized to non- F-Msgn1 expressing mesoderm cells. Scale bar: 100  $\mu$ m.

## SUPPLEMENTARY TABLES

**Table S1. Differentially expressed genes ( $P \leq 0.05$ , fold change  $\geq 1.5$ ) upon induction of *Msgn1* for 24 h and pathway analysis annotations for GO biological processes, signal transduction pathways and overrepresented tissues (related to Fig. 2).**

This Excel file contains four work sheets. The first worksheet lists differentially expressed genes upon induction of *Msgn1* at  $P \leq 0.05$ , fold change  $\geq 1.5$ . The differentially expressed gene list was analyzed in Genomatix pathway system. The next three worksheets show the most significant Gene Ontology biological process, signal transduction pathways and over represented tissues.

**Table S2. *Msgn1* binding peaks, peak distributions, annotation of peaks to genes, GO Biological processes and MGI expression annotations analyzed in GREAT (related to Fig. 3).**

First worksheet in this Excel file contains *Msgn1* ChIP-seq peaks analyzed by MACS. Columns A-F indicate MACS peaks, chromosome start to end, orientation of the read in the alignment and peak score. Column G-H indicates annotation of peaks with respect to genes. Column-I denotes distance to TSS (Transcription Start Site) and columns J-Q indicate nearest annotated genes. Second worksheet contains *Msgn1* and genome background peak distributions in intergenic, intragenic, promoter (2 kb upstream of TSS), first intron and 10 kb downstream of TES (Transcription End Site). Third worksheet contain list of association of *Msgn1* peaks to genes. *Msgn1* peaks that are present within 10 kb upstream to TSS and 5 kb downstream to TES are only considered for this analysis. Fourth worksheet contains GREAT version 2.0.2 analysis of *Msgn1* peaks. GO Biological process and MGI expression analyses are shown.

**Table S3. Integration of genes from *Msgn1* ChIP-seq and Microarray (related to Fig. 3).**

This worksheet lists the direct *Msgn1* target genes. *Msgn1* peaks associated with -10 kb upstream of the TSS and +5 kb downstream of the TES were annotated to 1860 unique genes. When this dataset was integrated with microarray data, 121 out of 332

differentially expressed genes from microarray analysis are present in both datasets and shown in column E (highlighted in yellow color), see also Fig. 3C.

**Table S1**

[Click here to Download Table S1](#)

**Table S2**

[Click here to Download Table S2](#)

**Table S3**

[Click here to Download Table S3](#)



## SUPPLEMENTARY MATERIALS AND METHODS

### Generation and characterization of *LacZ-TRE-F-Msgn1*<sup>Tg/+</sup> mice

For conditional and inducible expression of F-Msgn1 in mice, the F-Msgn1 ORF was amplified using primers Msgn1-Pst1-Flag.FW and Msgn1-Sal1.RV (see sequences below) and cloned in a bi-directional promoter containing vector pBI-G Tet (Clontech) into Sal1/Pst1 sites to generate an 8.4 kb  $\beta$ -globin polyA-*lacZ*-TRE-F-Msgn1-SV40 polyA transgenic construct. This vector was linearized with Ase1 to generate 7.2 kb fragment for pronuclear injection following standard procedures. Transgenic mice were generated in the Transgenic Core Facility, NCI-Frederick. 10 transgenic founder males were crossed to constitutively expressing *ROSA26-rtTA-IRES-EGFP*<sup>Tg/Tg</sup>. Mouse embryonic fibroblasts were generated from E13.5 day embryos. Embryos were dissected in sterile PBS, organs were removed, and tissue was dissociated with 0.25% Trypsin/EDTA. Cells from each embryo were plated on 10 cm tissue culture dishes, cultured in DMEM, 10%FBS, Pen/Strep, induced with 1  $\mu$ g/ml Dox for 24 h and stained for  $\beta$ -gal activity and F-Msgn1 protein expression. 1 out of 10 transgenic founder males were positive for  $\beta$ -gal and F-Msgn1 expression and the colony was established by crossing to C57BL/6 females. Genotyping was achieved using 3xFlag.FW and Msgn1.RV primers (for sequences see below). For generating conditional and inducible expression of F-Msgn1 in mice, *ROSA26-loxP-PGK-neo-pA-loxP-rtTA-IRES-EGFP*<sup>Tg/Tg</sup>; *LacZ-TRE-F-Msgn1*<sup>Tg/+</sup> line was crossed to *Tcre*<sup>Tg/Tg</sup>. *PGK-neo-pA* stop cassette in the mesodermal progenitor domain was excised by T-cre thereby activating the expression of rtTA (and EGFP) in NM stem cells (supplementary material Fig. S4A). For ectopic expression of F-Msgn1 (and  $\beta$ -gal) in NM stem cells, pregnant moms were fed with 1.6 mg/ml Dox and 5% Sucrose in drinking water starting from embryonic day E6.5 and harvested at E9.5 stage. Analysis of EGFP expression at E9.5 stage indicates that rtTA is expressed in both the neural tube and paraxial mesoderm as expected (Fig. 7A-F).

### Immunohistochemistry and Histology

For immunohistochemistry on frozen sections, embryos were dissected at E9.5 stage and fixed in 4% PFA for 30 min. After 3x PBS washes, embryos were suspended in 20%

sucrose at 4<sup>0</sup>C overnight followed by 5 µm sections in O.C.T. block using standard protocols. For staining, cross sections from appropriate embryonic locations were selected and permeabilized in 0.1% Triton X-100 in PBS for 15 min at room temperature, washed 3 times in PBS containing 0.1% Tween-20 (PBT), blocked with 0.2% BMBR in PBS for 30 min. Sections were incubated in primary antibodies at 4<sup>0</sup>C overnight, washed 5 times in PBT, incubated with secondary antibodies (1: 500) and DAPI (Molecular Probes, 1:5000), washed 5 times in PBT and mounted using Aqua-Poly/Mount (Polysciences). For paraffin sections, embryos were dehydrated in ethanol, cleared in Citrisolv (Fisher) twice for 20 min, imbedded in three 20 min changes of molten paraffin (Fisher) and sectioned to 10 µm. For histological analysis, embryos at E8.5 were fixed in Bouin's solution overnight at room temperature, washed in PBS until washes were clear. Embryos were then dehydrated in a graded ethanol series and embedded in paraffin. Cross sections were stained with haematoxylin and eosin according to standard procedures. Images were taken on a Zeiss Axioplan2 microscope. For immunocytochemistry, cells were grown on glass coverslips and fixed in 4% PFA for 10 min at room temperature. Cells were permeabilized with ice-cold 0.1% Triton X-100 for 5 min and blocked in 1% BSA in PBS for 30 min, followed by antibody staining. For immunocytochemistry of EBs, Day 2 EBs were plated on Matrigel (BD Biosciences) coated coverslips and grown for one more day before fixing and stained as above.

### **Cloning of 1x or 3X-Flag Msgn1 and 3x-Flag Tbx6 in p2lox vector**

P2lox-1x-Flag Msgn1 construct was generated by cloning 1x-Flag Msgn1 into p2lox vector using EcoR1 and Not1 restriction enzymes. For generating 3x-Flag Msgn1 and 3x-Flag Tbx6 constructs in p2lox vector, 105 bp oligos (Xho1-3x-Flag-Mlu1-Sac1-Afl2-Hind3) containing 3x Flag tag were cloned into *XhoI* and *HindIII* to generate p2lox 3x-Flag-GFP. Ires2-AcGFP from pIRES2-AcGFP1 (Clontech) replaced GFP via *EcoR1* and *Not1* sites in p2lox 3x-Flag-GFP to generate p2lox-3x-Flag-Ires2-AcGFP. Msgn1 cDNA was cloned into the *Mlu1* and *EcoR1* sites to generate p2lox 3x-Flag Msgn1-Ires2-AcGFP. Similarly, Tbx6 cDNA was cloned into the *Mlu1* and *Sal1* sites to generate p2lox 3x-Flag Tbx6-Ires2-AcGFP (see Table S5 for primer sequences used for cloning). All constructs were confirmed by DNA sequencing.

### **Embryonic stem cells, cell culture and differentiation**

Generation of inducible 1x or 3X Flag-Msgn1 ESCs was described previously (Chalamalasetty et al., 2011). Site-specific recombination was achieved by electroporation into A2lox Cre ESCs (Iacovino et al., 2011) induced with doxycycline for 24 h. Positive colonies were selected using 300 µg/ml G418 antibiotic and further characterized for the expression of F-Msgn1 and F-Tbx6. A2lox ESC were grown on mitomycin C (Sigma) treated mouse embryonic fibroblasts in DMEM, 15% FBS, Pen/Strep supplemented with ESGRO (Chemicon; 10<sup>3</sup> Units/ml) and differentiated as described (Chalamalasetty et al., 2011). ESCs grown on MEFs were trypsinized using 0.25% Trypsin-EDTA and neutralized with FBS containing DMEM medium. ESCs were separated from MEFs and plated on 10 cm sterile petri dishes to form EBs at a concentration of 2.5 million cells in 15 ml of DMEM, 10% FBS, Pen/Strep, non-essential amino acids, 2-Mercaptoethanol supplemented with 50 µg/ml Ascorbic acid. EBs were formed for 2 days and gene expression was induced on day 2 with 1 µg/ml doxycycline. GFP-Bry ESC were cultured and differentiated in serum independent conditions as described (Gadue et al., 2006). Mouse embryonic fibroblasts, NIH3T3 and P19 cells were cultured in DMEM, 10%FBS, and Pen/Strep medium.

### **EMSA assays**

3'end labeling with digoxigenin-11-ddUTP was accomplished on annealed double stranded oligos using terminal transferase according to manufacturer's instructions (DIG Gel Shift Assay Kit, Roche, Cat# 03353591910). Msgn1 protein was made *in vitro* from pCS2-Msgn1 expression construct using TNT Reticulocyte Lysate System (Promega, Cat# L5020). For binding reactions, 5 µl of Msgn1 protein was mixed with 0.8 ng of DIG labeled-oligonucleotides and incubated at room temperature for 20 min. Native protein-DNA complexes were migrated on 6% DNA retardation gels (Invitrogen, Cat# EC63652BOX) and transferred to Hybond-N+ (Amersham, Cat# RPN303B) in 0.5xTBE buffer. DNA-protein complexes were cross-linked using a UV Stratalinker (Stratagene). Membrane was washed, blocked, and incubated with anti-DIG AP for 30 min at room

temperature. Antibody was washed and membrane was detected using CDP-Star (Roche, Cat# 11685627001).

### **Reverse Transcription and Q-PCR**

Total RNA was isolated using RNeasy mini prep kit (Qiagen). First strand cDNA synthesis was performed using iScript cDNA synthesis kit (Bio-Rad, Cat# 170-8891) according to manufacturer's recommendations. For qPCR analysis, SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Cat# 172-5272) was used with CFX96 real-Time PCR Detection System (Bio-Rad). For normalization purposes, Gapdh expression values were used. Fold change expression was calculated using  $\Delta\Delta C(t)$  method.

### **Microarray Analysis**

Statistical analysis was performed on probe-intensity level data (CEL files) using Partek Genomics Suite and Genespring GX softwares. Background adjustment, quantile normalization and summarization of probe set intensities were carried out using RMA algorithm (Bolstad et al., 2003; Irizarry et al., 2003). Statistical analysis was done using one-way ANOVA ( $P \leq 0.05$ , Fold Change  $\geq 1.5$ ). Hierarchical clustering of the differentially expressed genes was done using gene expression values based on single linkage method and Euclidean distance measure. In order to find over-representation of functional categories in the list of differentially expressed genes at 24 h time point, comprehensive analysis was done in Genomatix Pathway System (GePS). GePS uses information extracted from public and proprietary databases to display canonical pathways or to extend networks based on literature data. For each annotation the  $P$ -value, the number of observed genes and the number of total genes of the annotation is noted. Differentially expressed genes in mesoderm (Genomatix literature mining) were extracted from GePS and hierarchical clustering and heatmap plotting of these genes were performed using gplots package from CRAN R package repository. Expression values were log<sub>2</sub>-transformed and expression values per gene across all samples were scaled for the comparison.

## ChIP-seq analysis

FASTQC software was used for quality check on raw sequence data. For peak calling MACS Version 1.4.1 (Model-based analysis of ChIP-seq) was used to identify the peak enrichments (Zhang et al., 2008). Significant enriched ChIP regions or peaks were called with  $P$ -value cutoff  $1.00e-05$  between F-Msgn1 and the control sample. The default bandwidth of 300 bp was used to slide 2 bandwidth windows across the genome to find regions with MFOLD = 10,30 to build the paired-peaks model. MACS generated BED format files that contained the peak locations, with the fifth column representing the  $-10 \times \log_{10} P$ -value of the peak region. This BED file was used to carry out Genomic Regions Enrichment of Annotations Tool (GREAT) analysis. For visualization of the peak regions for both Msgn1 and control samples, the peak files in BED format were converted to wiggle format (WIG) files which were then converted to a binary read count density file (TDF format) using Integrative Genomics Viewer (IGV) tools package (Thorvaldsdottir et al., 2013). The TDF files were loaded into the IGV for extensive analysis.

## ChIP-seq data statistics

Statistics of ChIP-seq data for Flag-Msgn1 and Input control as follows:

Sample	Total Reads	Mapped Unique	%	Mapped >1	%	Unmapped	%
Flag-Msgn1	13,539,303	10,028,765	74.1%	3,009,110	22.2%	501,428	3.70%
Input Control	10,181,527	7,413,755	72.8%	2,342,725	23.0%	425,047	4.17%

## Assigning Peaks to Genes

For assigning 4087 Msgn1 peaks to genes, peaks present within  $-10$  kb to TSS, gene body and  $+5$  kb to TTS, were extracted for association with respective genes. Peak distributions at intergenic, intragenic, promoter region, TES region were calculated using CisGenome (Ji et al., 2008). Peaks located in first intron region were annotated using PeakAnalyzer 1.4 (Salmon-Divon et al., 2010). Genome background distribution was calculated using in-house developed python scripts with gene annotation from mm9 ref Flat file downloaded from UCSC genome database.

### **Genomic Regions Enrichment of Annotations Tool (GREAT) analysis**

Bed files comprising the genomic coordinates of Msgn1 peaks were uploaded to GREAT, version 2.0.2 (McLean et al., 2010) and run against the whole genome background using NCBI build 37 or UCSC mm9. Association of genomic regions was defined as 5 kb upstream and 1 kb downstream to TSS up to 1000 kb max extension. GO Biological process and MGI expressions were analyzed for each peak category.

### **De novo Motif Analysis**

As the average length of Msgn1 sequences is 705.4 residues and total length of 2,883,053 residues, for de novo motif analysis we exploited the use of MEME-ChIP (Bailey et al., 2009) as it was designed specifically for analyzing motifs in large datasets. Fasta files of Msgn1 ChIP sequences were uploaded to MEME-ChIP website using default JASPAR and UniPROBE database sets. Multiple Em for Motif Elicitation (MEME) (Bailey and Elkan, 1994), Discriminative DNA Motif Discovery (DREME) (Bailey, 2011) and TOMTOM (Gupta et al., 2007) results from MEME-ChIP were analyzed with default parameters for de novo motif analysis.

### **Luciferase Assays**

Enhancer regions of *Tbx6* +3913, *Snai1*+5148, and *Pdgfra* -2140 ChIP-seq peak sequences containing the Msgn1 response elements were PCR-amplified and inserted into the *NheI* and *EcoRV* restriction sites of the pGL4.23 luciferase (minp2-luc) reporter vector (Promega). A 300 bp Mesp2 promoter (Yasuhiko et al., 2008) was amplified from C57BL/6 genomic DNA and cloned into *KpnI* and *XhoI* sites of pGL4.10 vector (see oligo sequences below). For luciferase assays, a total of 500 ng of DNA containing 200 ng of reporter constructs and 50-100 ng of pCS2-MT-Msgn1 or pCS2-MT empty vector were transiently transfected with lipofectamine 2000 (Invitrogen) or Fugene6 (Roche) into NIH3T3 or P19 cells for 40 h. About 5 ng of pGL4.74 (hRluc/TK) vector (Promega) was used as internal control in all the transfections and equal amounts of DNA was accomplished with empty pCS2 MT vector. Cell lysis and luciferase activity was measured using Dual Luciferase Assay Kit (Promega, Cat# E1960) following

manufacturer's recommendations. Fold change analysis for reporter activity was described previously (Chalamalasetty et al., 2011). For luciferase assays of differentiating iF-Tbx6 ESC, 20000 cells/well were plated on matrigel-coated 24 well plates for 2 days without ESGRO. On day 2, differentiated cells were transfected with *Mesp2* promoter luciferase reporter, Notch Intracellular Domain (NICD) expression construct and treated with (+) or without (-) Dox to induce F-Tbx6 and cultured for another 2 days before harvesting for luciferase assay.

**ChIP-seq peak sequences: potential E-box consensus sequences are highlighted in yellow.**

**Tbx6 +3913 peak (chr7: 133,928,330-133,929,489)**

TGAGGCCTACCTTCTACACCCTGCCGCTTTTCATGGCGCCCCCAGTCACCTAC  
CAGCCAGGTAAGGTGAAGGGTAGAGGGCTGGGTTGTCCAGGAGGGGGTGGC  
CTTCTTCCGTTCTCATACTGTCTCCTGTGACTAGGACCCCCAGCTTCGCTGA  
GGCTCCAGACCCTGGGCGCCCAGCCCCCTACTCAGCTGCATTTCTGGACCTAC  
AGCTGGACCAGGGGGCTCTGCCTATCAGGCAGCTCCATCTGTACCATCCTTT  
GCCCCACACTTCATCCAAGGGGGTCCCTTCCCTTACCGTACCCAGGACCTGG  
AGGTTATCTGGACATGGGATCCAAGCCAATGTACTGAGCCTCGGTGTAGTAA  
CCCTATGCCATCTTCCCTTGATCCTCCAGCTCCCTTCCCCCAGCCTGGTAGCAT  
CCGCATTGAAGTGGTATCCCCCTCCCCCCA **CCAAATG** GCTGGCTTGGGCCTC  
TCTTCCACCCTTTAGTTCACACCTTGATTTCACTCCACCCCCTCTGGCTTCAA  
GCTCAGGCAAGGCAGCTCTGAGT **CCAGCTG** GGGCCTGCTTCCCCTCAGCTCTC  
ACCTTAGTGCGAATGGAGGGAGCCTCTGCCTGG **CCAAATG** GGGCTCCCAGCC  
CAGTACCCCCACCTCCTGGGGGCCTTGTTACTAGTGCAAGCCATGCCAATTTG  
TTCTCAGGATCAGAGTATTTTTGTTAATAAACTCCAAAGACATTAGTGCTCA  
GGAACTTGAGCTAATGGTGTGGCCCTGGGTGGGGCTGGGGGAGCCACAGC  
TTAAGGT **CCAGCTG** CTTGCCCTGCCCTTGTGGCTAGGAAAGGGCCTGCTCTG  
GGCTGTCCCCTTCAAGATGGAAAAGCTTCCCTCACAATGGTTGCTTTAAGAATC  
AAGGCTGTCCATGAACCTAGATTTTCATGAAGCACTAAGGAAGCCTTTGGGAA  
AGATTAGAAGATTAGCAGATGGAGCCTCAAGGCTCCGGGCCCAACGGACAA  
GCAGAGCAGGTCTCCATGTCCACCCCAATCAAACCTGACAGAACCTAAGTCT

ATTTTCTTCTTTTATTAGAATTTTTTCTTTTTTCTCAAATGTTTTATCTAA  
AAACAAAACAAAAGAAAAAGAAAGGAAAAAAAAGGAAAAAAGAAAA  
AA

**Snail +5148 peak (Chr2: 167,368,488-167,369,262)**

GGGCTCTCTTGGGAAGGTGCTTTGTGGGGCTTAGCCACAGCCTGGCCAGGAG  
CTCTGGCTCCACCATAGCAGGTCTTTGGGGCAGAGCTAGCTGCCCTC**CCAGC**  
**TG**AAGGGAGGATGTCTTTCCTGTTCCCAAGCGCCTCCCTGTCAGGCAGGCC  
ACACCTCTGTCTGGCTGCTGGGTA**CCAAATG**GTCTTTGATGCAAGGAAATGG  
GCTGGCTGGAGCCAGGAGGGTGAGAAGCCAGGCCAAAGGGTAGGGGTCATA  
CAGCTTCCTGCTTGGAGGACCAGGGCTGGAGCAATTTACCTCAGGCTAGTTG  
GTGCTGGTGACCTCAACTCTGAGGGCCTCCTGGGGCTCGAGAGAAAATGAGT  
GTCCTTTCTAAGCACAAAAGGTGATGAAGGTGCCAGTTTCCTGTGGGGAGCT  
GAGCAGGCGTCGGCTGGGAGCTTCAGCCAGCCCAGCTCCAACCCGGCTGGGG  
GGTGGGGAGCTTCCAGCTCCAAAGGAGAGATGAGTCTTCTAAACAGGAAATG  
GTGACCCGCAGGCGGGAACAGCCTAATGACTTACAGCTTTCCTCCAAAAGG  
AAAAAAAAAAGTCTCTAACAATCGCCAAATTGTGGCAGGCCTCCCTGAAGGG  
GAGGTTGGTGAGGTGGTATTGGGGGCTGGGCATCAACGAGCTGAAGGCCTAC  
TGTGTCCCAGGAGGTTCTCCACATAGGTTGAACAGTGAGTGACTTCACCCC  
GGGCCTGGAGGGATGGGATGG GCCTCCAGGACCAGACTCAGCGCTG

**Pdgfra -2140 peak (chr5: 75,545,700-75,546,652)**

GGCTGGGTTCTTGACACACTAACCCAAGCAGAAAGCGATTCCGTTAGGCTCA  
CCAAGTGACCTGTTCCACCCCCCCCCCCCCCGACATCAGGGACAGAAGTTC  
AGAAGGGGAGAAAAGACCTTGGCTAGGCACTGGCACTTGCCTTCTTCGGCCC  
CGAGACCACCAGGACCCGGGGTGTGAGGGGGCAATCTGGAAGGATTTTAG  
AACAGGGCAAGAACTGGGCTGGCCCCCTCCCCCTGACACCCTGGGTTGAGTG  
ACGGGCACAGACCGGACCTTTCCTTCAGCCCCAGAGAGTGGAGGGCGGAA  
AATGCGATTCAAACCCGCCTATGAGCAGTGGGAGCAGTGGGCGGGGGCAGG  
ACCTGGTTTTAGAAACCAACCGCTTGGAGCTCCAAGCCTCCTGCCCTGAGGT  
GGCTCCAGACCACTGCGTTTATTTTGCTTTTGCAGGCAAG**GCAGATG**CTTTGC



TGGGTCCCGACTCAGCGAGATCTTTGCCCAA**CCATTTG**CTTGCCTGCTCCACC  
CCCTACTGGGGCACTATCCCGGGAGTCAGCTCTCCAGGCCTTTAAACCTTAGG  
GAATGTC**ACACATG**GAAACCTTTA**GCAAATG**TTTGTTAATGATCATAACAAA  
GGCATCATTTAAATTAGGCAGGTAATTACTACCAGAAGA**GCAACTG**GGTGCT  
CGATACTCATCCTTTTCTCCCCTTTAAACTCCCGAGGAATTACTGTCTCAGGA  
TACCCGAGCCGGCCCGCTTTTCGGACTCCTTTAAAACGGCTCTGAAATGAACT  
GTGAAGTTA**CCATTTG**TGGGAGGAGAACGCAGGGGCCCTGCTTGAGGCAGGT  
AGTGGGATTAGAGCTGGTAGGGTGTTCCTTGCAAAGCAGGCTCTCCTCTGCA  
GACGCTCAGCCCCAAACTCACCCCCTCCTTGGGTACCTGCTACAAAAACCTA  
AATCTTGACTTGCT

**Pdgfr $\alpha$ -2140 ChIP-sequence with point mutations in the E-boxes that were used in the luciferase assay. Here red colored bases are substituted in the consensus E-box sequence**

GGCTGGGTTCTTGACACACTAACCCAAGCAGAAAGCGATTCCGTTAGGCTCA  
CCAAGTGACCTGTTCCACCCCCCCCCCCCCCGACATCAGGGACAGAAGTTC  
AGAAGGGGAGAAAAGACCTTGGCTAGGCACTGGCACTTGCCTTCTTCGGCCC  
CGAGACCACCAGGACCCGGGGTGTGAGGGGGCAATCTGGAAGGATTTTAG  
AACAGGGCAAGAACTGGGCTGGCCCCCTCCCCCTGACACCCTGGGTTGAGTG  
ACGGGCACAGACCGGACCTTTCCTTCAGCCCCAGAGAGTGGAGGGGCGGAA  
AATGCGATTCAAACCCGCCTATGAGCAGTGGGAGCAGTGGGCGGGGGCAGG  
ACCTGGTTTCAGAAACCAACCGCTTGAGCTCCAAGCCTCCTGCCCTGAGGT  
GGCTCCAGACCACTGCGTTTATTTTGCTTTTGCAGGCAAG**GCAGGCA**CTTTGC  
TGGGTCCCGACTCAGCGAGATCTTTGCCCAA**CCATCCA**CTTGCCTGCTCCACC  
CCCTACTGGGGCACTATCCCGGGAGTCAGCTCTCCAGGCCTTTAAACCTTAGG  
GAATGTC**ACACATG**GAAACCTTTA**GCAAGCA**TTTGTTAATGATCATAACAAA  
GGCATCATTTAAATTAGGCAGGTAATTACTACCAGAAGA**GCAACTG**GGTGCT  
CGATACTCATCCTTTTCTCCCCTTTAAACTCCCGAGGAATTACTGTCTCAGGA  
TACCCGAGCCGGCCCGCTTTTCGGACTCCTTTAAAACGGCTCTGAAATGAACT  
GTGAAGTTA**CCATTTG**TGGGAGGAGAACGCAGGGGCCCTGCTTGAGGCAGGT  
AGTGGGATTAGAGCTGGTAGGGTGTTCCTTGCAAAGCAGGCTCTCCTCTGCA

GACGCTCAGCCCCAAACTCACCCCCTCCTTGGGTACCTGCTACAAAAACCTA  
AATCTTGACTTGCT

**Table S4. Antibodies used in this study:**

Flow Cytometry	Company, Cat#, Dilution
Rat anti-mouse CD309 (Flk1), Clone: Avas12a1	eBioscience, Cat# 12-5821-81, 1:100
Rat anti-Mouse CD140a (Pdgfr $\alpha$ ), Clone: APA5	eBioscience, Cat# 17-1401-81, 1:100
Rat PE-conjugated IgG2a isotype control, Clone: eBR2a	eBioscience, Cat# 12-4321-81, 1:100
Rat APC-conjugated IgG2a control, Clone: eBR2a	eBioscience, Cat# 17-4321-81, 1:100
Rat Biotin-anti-CD324 (E- Cadherin), Clone: DECMA-1	eBioscience, Cat# 13-3249-80, 1:400
Biotin IgG isotype control, Clone: HTK888	Biolegend, Cat# 400903, 1:400
PE Streptavidin	eBioscience, Cat# 558774, 1:400
Immunohistochemistry	
Mouse anti-Flag M2	Sigma, Cat# F3165, 1:500
Goat anti-Tbx6	R&D Systems, Cat# AF4744, 1:100
Rabbit anti- $\beta$ -galactosidase	MP Biomedicals, Cat# 559761, 1:5000
Rabbit anti-Sox2	Millipore, Cat# AB5603, 1:250
Rabbit anti-GFP	eBioscience, Cat# 14-6774-63, 1:100
Rabbit anti-E-Cadherin	Cell Signaling, Cat# 24E10, 1:100
Rabbit anti-Vimentin	Cell Signaling, Cat# D21H3, 1:100
Rhodamine Phalloidin	Molecular Probes, 1:40
Rabbit Active Caspase-3	Millipore, Cat#AB3623, 1:250
Alexa Fluor 488 and Alexa Fluor	Invitrogen, 1:500

598 secondary antibodies	
DAPI	Molecular Probes, 1:5000
<b>Western Blotting</b>	
Mouse Anti-Flag M2	Sigma, Cat# F3165, 1:500
Mouse anti-Actin	Chemicon, Cat# MAB1501, 1:1000
<b>ChIP-seq</b>	
Mouse Anti-Flag M2	Sigma, Cat# F3165
<b>ChIP-qPCR</b>	
Mouse anti-Msgn1 Ascites (Chalamalasetty et al., 2011) raised against a cocktail of three peptides in Msgn1 protein: CWKSRARPLELVQESP (16aa) CDLLNSSGREPRPQSV (16aa) CSHEAAGLVELDYS (14aa)	2 $\mu$ l per ChIP assay
Mouse control Ascites, Clone: NS-1	Sigma, Cat# M8273, 2 $\mu$ l per ChIP assay

**Table S5. Oligos used in this study:**

<b>RT-PCR</b>	
RT-Snai1.FW	CACACGCTGCCTTGTGTCT
RT-Snai1.RV	GGTCAGCAAAAGCACGGTT
RT-CDH1.FW	CAGGTCTCCTCATGGCTTTGC
RT-CDH1.RV	TTCCGAAAAGAAGGCTGTCC
RT-Cldn3.FW	ACCAACTGCGTACAAGACGAG
RT-Cldn3.RV	CAGAGCCGCCAACAGGAAA
RT-Cldn7.FW	GGCCTGATAGCGAGCACTG
RT-Cldn7.RV	GTGACGCACTCCATCCAGA
RT-Olcn.FW	TTGAAAGTCCACCTCCTTACAGA

RT-Olcn.RV	CCGGATAAAAAGAGTACGCTGG
RT-MMP2.FW	CAAGTTCCCCGGCGATGTC
RT-MMP2.RV	TTCTGGTCAAGGTCACCTGTC
RT-Vimentin.FW	CGTCCACACGCACCTACAG
RT-Vimentin.RV	GGGGATGAGGAATAGAGGCT
RT-Fn1.FW	GCAGTGACCACCATTTCCTG
RT-Fn1.RV	GGTAGCCAGTGAGCTGAACAC
RT-Twist1.FW	GGACAAGCTGAGCAAGATTCA
RT-Twist1.RV	CGGAGAAGGCGTAGCTGAG
RT-CDH2.FW	AGCGCAGTCTTACCGAAGG
RT-CDH2.RV	TCGCTGCTTTCATACTGAACTTT
RT-Gapdh.FW	AATGTGTCCGTCGTGGATCTG
RT-Gapdh.RV	CTGCTTCACCACCTTCTTGATGT
RT-Msgn1.FW	CTTCTGACACCGCTGGTCTG
RT-Msgn1.RV	GTGACTGCCGTAGCCATCG
RT-Flag.FW	ATGGACTACAAGGACGACGAT
RT-Flag Msn1.RV	GACCTCAGAGTAGGACTCCA
RT-Tbx6.FW	CCTTCCGATTTCTGAGACCACAT
RT-Tbx6.RV	AGGTCCAGAAATGCAGCTGAGTAG
RT-Pax3.FW	TTTCACCTCAGGTAATGGGACT
RT-Pax3.RV	GAACGTCCAAGGCTTACTTTGT
RT-OSR1.FW	CCGCCACTTCACTAAGTCCTAT

RT-OSR1.RV	GAGTGTAGCGTCTTGTGGACAG
RT-Pdgfra.FW	TCCATGCTAGACTCAGAAGTCA
RT-Pdgfra.RV	TCCCGGTGGACACAATTTTTC
RT-Foxc1.FW	CCCGTTCTTTCGACATAGGA
RT-Foxc1.RV	CAGAGACTCGCTTTCCTGCT
RT-Ephrin B1.FW	TGTGGCTATGGTCGTGCTG
RT-Ephrin B1.RV	CCAAGCCCTTCCCCTTAGG
RT-Dact1.FW	AAGAGATGCCGGTTTGTTGAA
RT-Dact1.RV	GCCCGAAGCTCCATCACTC
RT-Epha4.FW	CCAGACAGAGTAGCTGTGCG
RT-Epha4.RV	CGAGGAGCAGAATGGTGAAT
RT-Gata4.FW	AAGGCAGAGAGTGTGTCAATTGTGG
RT-Gata4.RV	TGGTAGTCTGGCAGTTGGCACAG
<b>ChIP-qPCR</b>	
ChIP-Snai1-3'enh.FW	AAGGGAGGATGTCTTTCCTGTT
ChIP-Snai1-3'enh.RV	ATTCCTTGCATCAAAGACCATT
ChIP-Tbx6-3'enh.FW	CCCCCACCAAATGGCTGGCTT
ChIP-Tbx6-3'enh.RV	GCAGGCCCCAGCTGGACTCAGA
<b>EMSA</b>	
DIG-Pdgfra-Ebox1	CTTTTGCAGGCAAGGCAGATGCTTTGCTGG GTC
DIG-Pdgfra-Ebox2	TGCCCAACCATTTGCTTGCCTGCTCCA

DIG-Pdgfra-Ebox3	AGGGAATGTCACACATGGAAACCTTTAG
DIG-Pdgfra-Ebox4	GAAACCTTTAGCAAATGTTTGTTAATGA
DIG-Tbx6-Ebox1	CCCCCACCAAATGGCTGGCTT
DIG-Tbx6-Ebox2	TCTGAGTCCAGCTGGGGCCTGC
DIG-Tbx6-Ebox3	TGCTGGCCAAATGGGGCTCCC
DIG-Tbx6-Ebox4	TTAAGGTCCAGCTGCTTGCCCT
<b>Cloning</b>	
Msgn1-Pst1-Flag.FW	CCCTGCAGACCGCCGCCACCATGGACTACA AAG
Msgn1-Sal1.RV	CCGTCGACTCACACTCTGTGGCCTGGGCT CTCTC
Xho1-3x-Flag-Mlu1-Sac1-Afl2- Hind3	5'TCGAGACCGCCGCCACCATGGACTACAAA GACCATGACGGTGATTATAAAGATCATGAT ATCGATTACAAGGATGACGATGACAAGACG CGTGAGCTCCTTAAGA 3'
Msgn1-Mlu1.FW	CCACGCGTGACAACCTGGGTGAGACCTTCC TCAGC
Msgn1-EcoR1.RV	CCGAATTCTCACACTCTGTGGCCTGGGCT CTCTC
Tbx6-Mlu1.FW	CCACGCGTTACCATCCACGAGAGTTGTACC CCTCC
Tbx6-Sal1.RV	CCGTCGACTCAGTACATTGGCTTGGATCCCA TGCCAG
Snai1 +5148 -Nhe1-Pst1.FW	CGTAGCTGCAGCAAGCTTCCTGAACTCTTA GTTCTTC
Snai1 +5148-EcoRV.RV	CCGATATCCAGAACCAGGCTGACCAAGTCC TA
Tbx6 +3913-Nhe1-Pst1.FW	CGTAGCTGCAGGGACCTGGAGGTTATCTG GACAT
Tbx6 +3913-EcoRV.RV	CCGATATCCTTGATTCTTAAAGCAACCATTG TG

Pdgfra -2140-Nhe1-Pst1.FW	GCGCTAGCCTGCAGCTCCAGACCACTGCG
Pdgfra -2140-EcoRV.RV	GCGATATCGTATCCTGAGACAG
Mesp2 -300bp-Kpn1.FW	CCGGTACCGTAGCAGGGCGGAGTGAAGGTG GGGGCTCG
Mesp2 -300-Xho1.RV	CCCTCGAGGGCTGGTGTGACTCTGGGAAGC TGGACCGG
<b>Genotyping</b>	
3xFlag.FW	CGCCACCATGGACTACAAAG
Msgn1.RV	GACCTCAGAGTAGGACTCCA

### Supplementary References:

- Bailey, T. L.** (2011). DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653-1659.
- 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.
- Bailey, T. L. and Elkan, C.** (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* **2**, 28-36.
- Bolstad, B. M., Irizarry, R. A., Astrand, M. and Speed, T. P.** (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193.
- Chalamalasetty, R. B., Dunty, W. C., Jr., Biris, K. K., Ajima, R., Iacovino, M., Beisaw, A., Feigenbaum, L., Chapman, D. L., Yoon, J. K., Kyba, M., et al.** (2011). The Wnt3a/beta-catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. *Nature communications* **2**, 390.
- Gadue, P., Huber, T. L., Paddison, P. J. and Keller, G. M.** (2006). Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 16806-16811.
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. and Noble, W. S.** (2007). Quantifying similarity between motifs. *Genome biology* **8**, R24.
- Iacovino, M., Bosnakovski, D., Fey, H., Rux, D., Bajwa, G., Mahen, E., Mitanoska, A., Xu, Z. and Kyba, M.** (2011). Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem cells* **29**, 1580-1588.
- Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B. and Speed, T. P.** (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research* **31**, e15.
- Ji, H., Jiang, H., Ma, W., Johnson, D. S., Myers, R. M. and Wong, W. H.** (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nature biotechnology* **26**, 1293-1300.
- McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. and Bejerano, G.** (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495-501.
- Salmon-Divon, M., Dvinge, H., Tammoja, K. and Bertone, P.** (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC bioinformatics* **11**, 415.
- Thorvaldsdottir, H., Robinson, J. T. and Mesirov, J. P.** (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-192.
- Yasuhiko, Y., Kitajima, S., Takahashi, Y., Oginuma, M., Kagiwada, H., Kanno, J. and Saga, Y.** (2008). Functional importance of evolutionally conserved Tbx6



binding sites in the presomitic mesoderm-specific enhancer of *Mesp2*.  
*Development* **135**, 3511-3519.

**Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al.** (2008). Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137.