## SUPPLEMENTARY MATERIAL

## SUPPLEMENTARY FIGURES

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

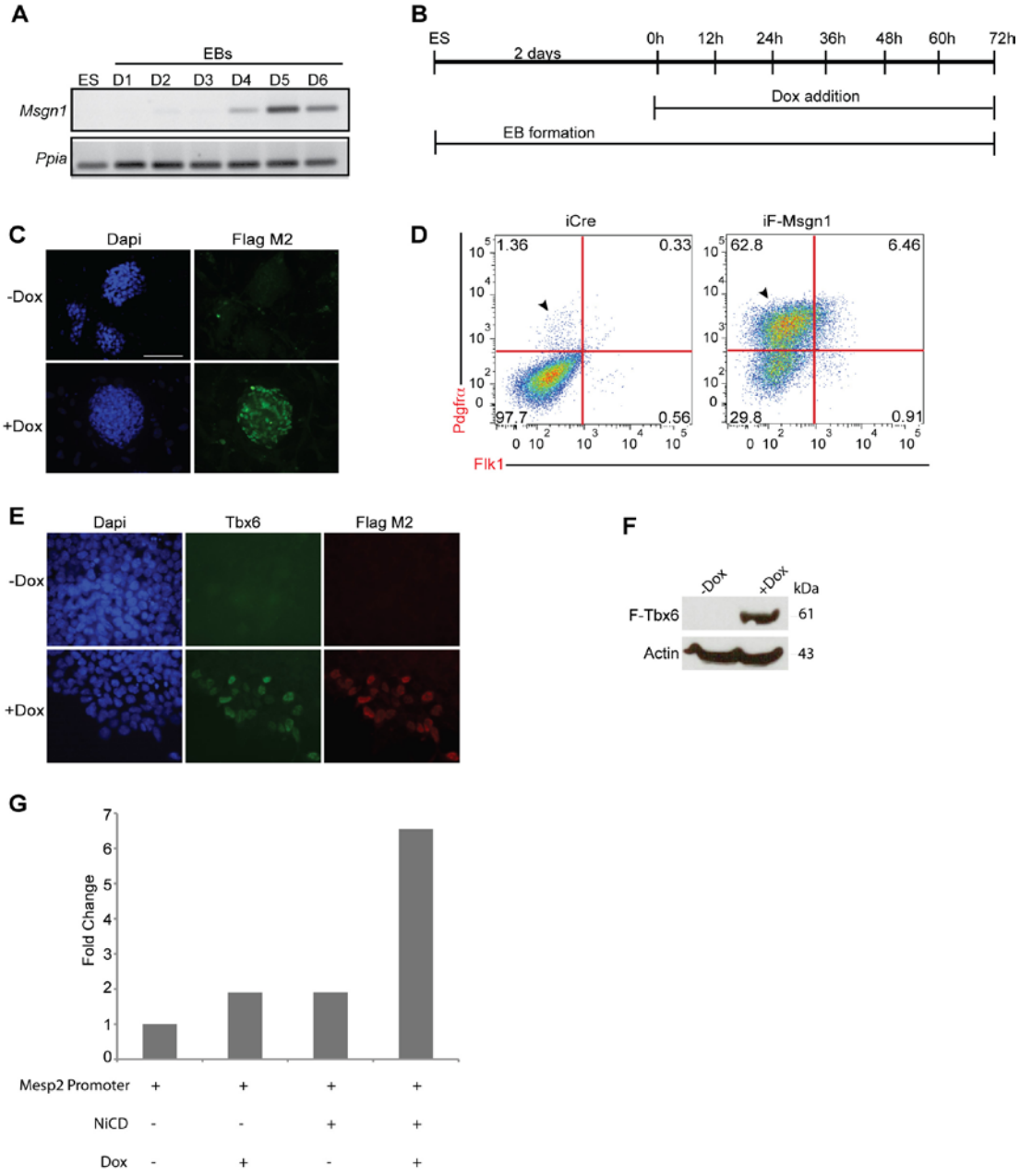


**Fig. S1. Overexpression of Msgn1 is sufficient to induce a Pdgfrα-positive population.** (A) RT-PCR analysis of endogenous *Msgn1* and *Ppia* (loading control) expression in serum free EBs from Day 0 (ES) to Day 6 (D6) of differentiation. (B) Schematic view of ESCs to EB differentiation timeline. (C)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

μm. (D) Flow cytometry analysis for the expression of cell surface markers Pdgfrα and Flk1 in control (iCre EBs) and iF-Msgn1 EBs induced for 12 h. Arrowheads indicate the Pdgfrα 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)

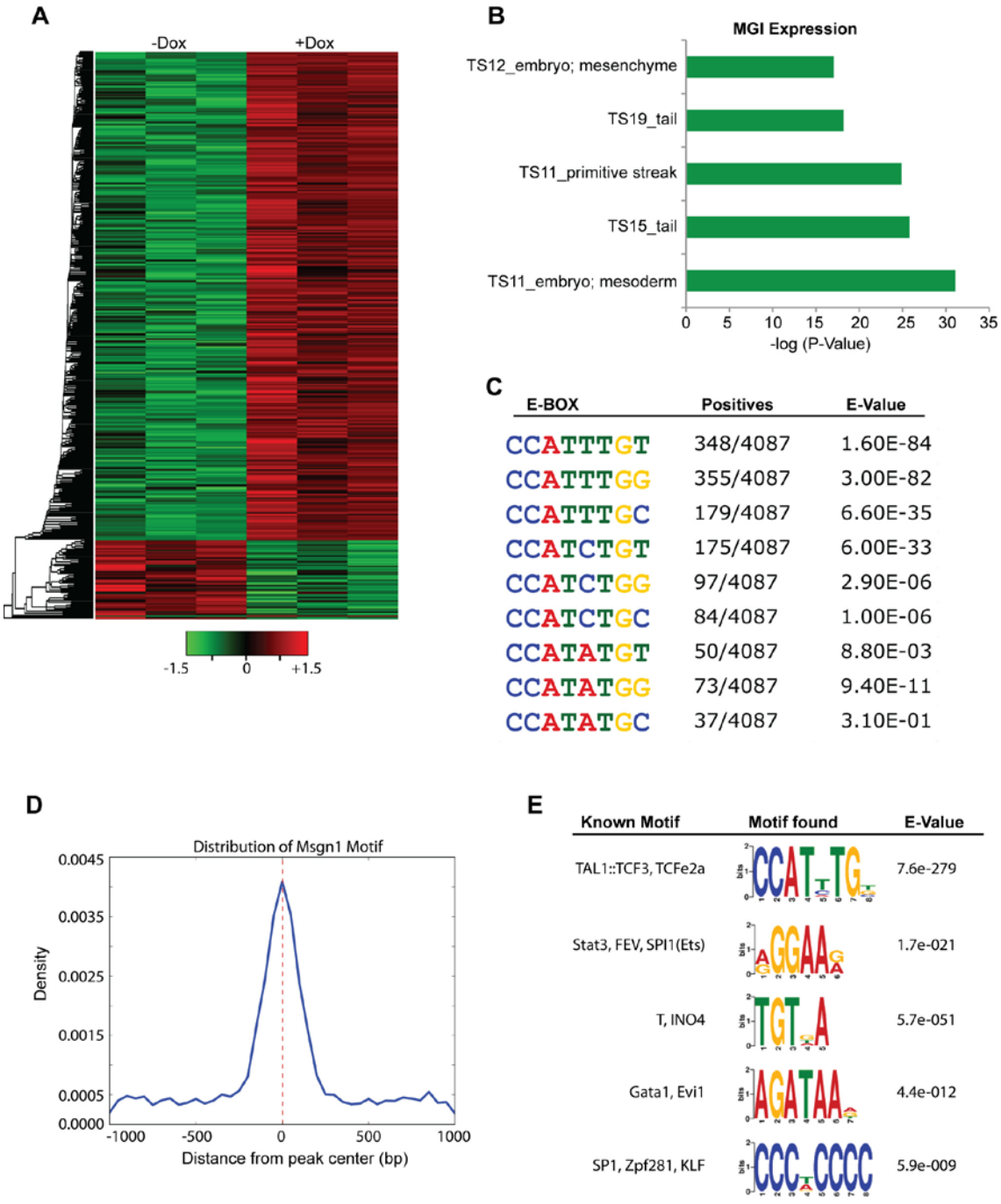


**Fig. S2. Microarray and ChIP-seq analysis of Msgn1-binding sites.** (A) Hierarchical clustering of 443 differentially expressed Affymetrix IDs or 332 differentially expressed unique genes at $P \leq 0.05$, Fold change$\geq$1.5 upon induction of F-Msgn1 with (+) or without (–) Dox for 24 h. Expression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

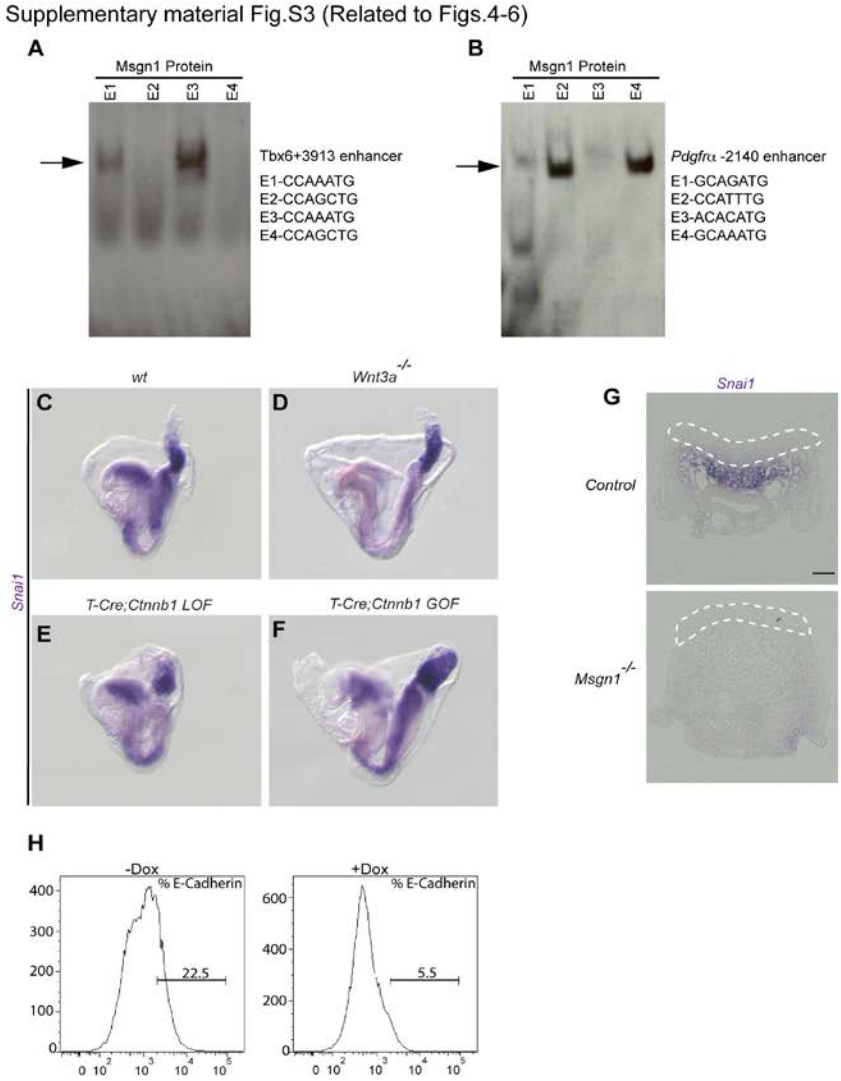


**Fig. S3. Biochemical and genetic analyses of *Tbx6, Pdgfrα* and *Snai1* regulation by Msgn1.** (A) Electrophoretic Mobility Gel Shift Analysis (EMSA) of Msgn1 binding to DIG-labeled oligos encompassing E boxes E1-E4 of *Tbx6+3913* enhancer. Msgn1 only binds to specific E-box sequence 'CATTTG'. Arrow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

E8.5 *WT* (C), *Wnt3a*$^{-/-}$ (D), *T-Cre; Ctnnb1 LOF* (β-catenin loss-of-function) (E) and *T-Cre; Ctnnb1 GOF* embryos (β-catenin gain-of-function) (F). (G) PS cross-sections of *Snai1* WISH in *WT* and *Msgn1*$^{-/-}$ embryos from Fig. 5E-H are shown. Dotted white line indicates the epithelial PS. *Snai1* expression is down regulated in the mesodermal domain of the PS of *Msgn1*$^{-/-}$ mutants compared to control embryos. Scale bar: 100 μm. (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).

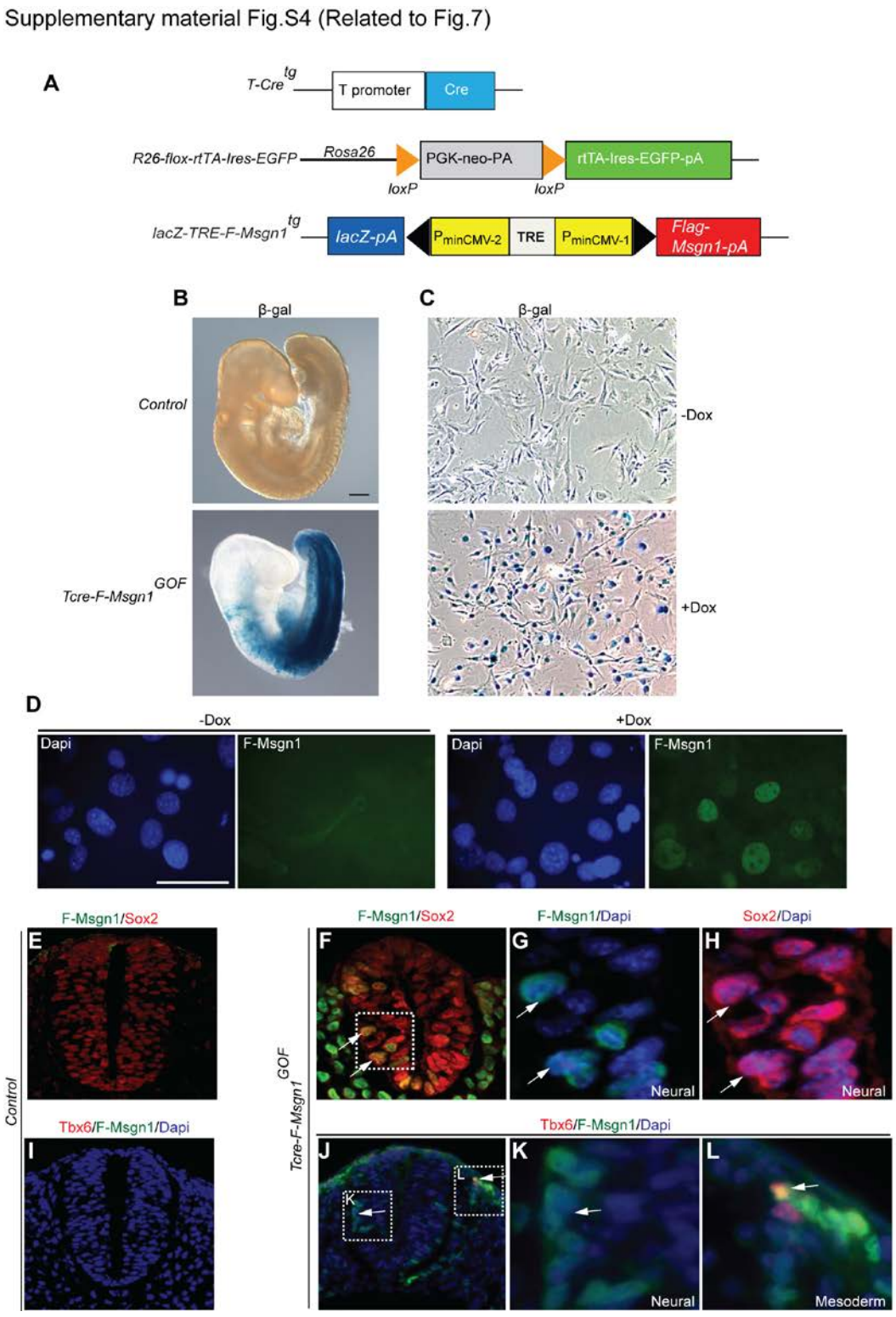


**Fig. S4. Triple transgenic mouse strategy for inducible expression of Msgn1 *in vivo*.**

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

Msgn1. Transgenic mouse lines *T-cre*$^{tg}$, *R26-flox-rtTA-Ires-EGFP* and *LacZ-TRE-F-Msgn1*$^{tg}$ 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* $^{tg}$ 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*$^{GOF}$ mutants treated with Dox were processed for β-gal staining. Scale bars: 200 μm. (C, D) Fibroblasts derived from E13.5 constitutively expressing *F-Msgn1*$^{GOF}$ embryos treated with (+) or without (–) Dox for 24 h were β-gal stained (C) or labeled with dapi and anti-Flag antibodies (D). Scale bars: 50 μm. (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*$^{GOF}$ 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)

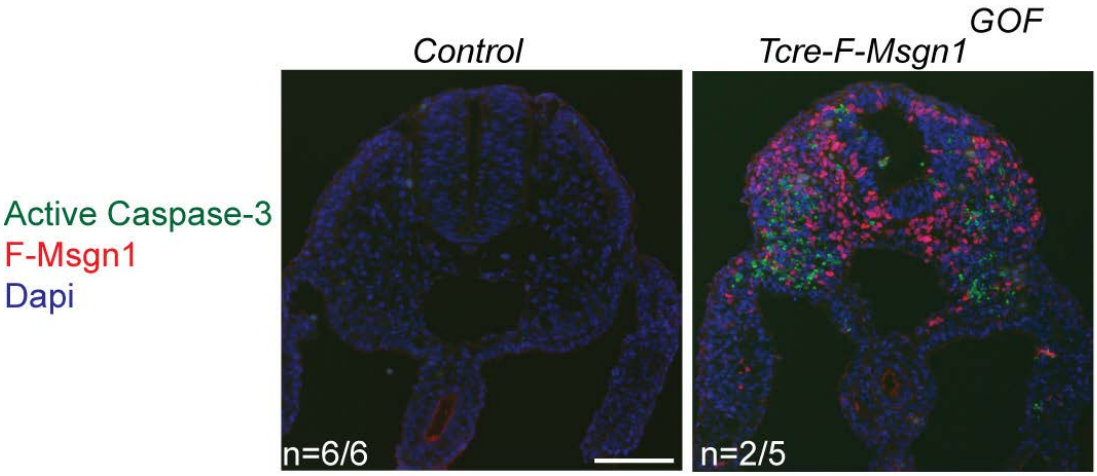


**Fig. S5. Cell death analysis in *Tcre-F-Msgn1*$^{GOF}$ mutants.**

Immunostaining of active Caspase-3 and F-Msgn1 in cross-sections of control and *Tcre-F-Msgn1*$^{GOF}$ 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 μm.

## SUPPLEMENTARY TABLES

**Table S1. Differentially expressed genes ($P \leq 0.05$, fold change ≥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≥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*$^{Tg/+}$ 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 β-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*$^{Tg/Tg}$. 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 μg/ml Dox for 24 h and stained for β-gal activity and F-Msgn1 protein expression. 1 out of 10 transgenic founder males were positive for β-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*$^{Tg/Tg}$*; LacZ-TRE-F-Msgn1*$^{Tg/+}$ line was crossed to *Tcre*$^{Tg/Tg}$. *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 β-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 $4^0$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 $4^{\circ}$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 haemotoxylin 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 *Xho*I and *Hin*dIII to generate p2lox 3x-Flag-GFP. Ires2-AcGFP from pIRES2-AcGFP1 (Clontech) replaced GFP via *Eco*R1 and *Not*1 sites in p2lox 3x-Flag-GFP to generate p2lox-3x-Flag-Ires2-AcGFP. Msgn1 cDNA was cloned into the *Mlu*1 and *Eco*R1 sites to generate p2lox 3x-Flag Msgn1-Ires2-AcGFP. Similarly, Tbx6 cDNA was cloned into the *Mlu*1 and *Sal*1 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; $10^3$ 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 ΔΔC(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 log2-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×log10 *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 Pdgfrα -2140* ChIP-seq peak sequences containing the Msgn1 response elements were PCR-amplified and inserted into the *Nhe*I and *Eco*RV 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 Kpn1 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
CTTCTTCCGTTCTCATACCTGTCTCCTGTGACTAGGACCCCCAGCTTCGCTGA
GGCTCCAGACCCTGGGCGCCCAGCCCCCTACTCAGCTGCATTTCTGGACCTAC
AGCCTGGACCAGGGGGCTCTGCCTATCAGGCAGCTCCATCTGTACCATCCTTT
GCCCCACACTTCATCCAAGGGGGTCCCTTCCCTCTACCGTACCCAGGACCTGG
AGGTTATCTGGACATGGGATCCAAGCCAATGTACTGAGCCTCGGTGTAGTAA
CCCTATGCCATCTTCCCTTGATCCTCCAGCTCCCTTCCCCCAGCCTGGTAGCAT
CCGCATTGAAGTGGTATCCCCCTCCCCCCCACCAAATGGCTGGCTTGGGCCTC
TCTTCCACCCTTTAGTTCACACCTTGATTTCACTCCACCCCCTCTGGCTTCAAA
GCTCAGGCAAGGCAGCTCTGAGTCCAGCTGGGGCCTGCTTCCCCTCAGCTCTC
ACCTTAGTGCGAATGGAGGGAGCCTCTGCCTGGCCAAATGGGGCTCCCAGCC
CAGTACCCCCACCTCCTGGGGGCCTTGTTACTAGTGCAAGCCATGCCAATTTG
TTCTCAGGATCAGAGTATTTTTGTTAATAAAACTCCAAAGACATTAGTGCTCA
GGAAACTTGAGCTAATGGTGTTGGCCCTGGGTGGGGCTGGGGGAGCCACAGC
TTTAAGGTCCAGCTGCTTGCCCTGCCCTTGTGGCTAGGAAAGGGCCTGCTCTG
GGCTGTCCCCTTCAAGATGGAAAAGCTTCCTCACAATGGTTGCTTTAAGAATC
AAGGCTGTCCATGAACCTAGATTTCATGAAGCACTAAGGAAGCCTTTGGGAA
AGATTAGAAGATTAGCAGATGGAGCCTCAAGGCTCCGGGCCCAACGGACAA
GCAGAGCAGGTCTCCATGTCCACCCCAATCAAACCTGACAGAACCTAAGTCT

ATTTTTCTTCTTTTATTAGAATTTTTTTCTTTTTTTCTCAAAATGTTTTATCTAA
AAACAAAAACAAAAAGAAAAAAGAAAGGAAAAAAAAGGAAAAAAGAAAAA
AA

**Snai1 +5148 peak (Chr2: 167,368,488-167,369,262)**

GGGCTCTCTTGGGAAGGTGCTTTGTGGGGCTTAGCCACAGCCTGGCCAGGAG
CTCTGGCTCCACCATAGCAGGTCTTTGGGGCAGAGCTAGCTGCCCCTCCCAGC
TGAAGGGAGGATGTCTTTCCTGTTCCCAAGCGCCTCCCTGTCAGGCAGGCCC
ACACCTCTGTCTGGCTGCTGGGTACCAAATGGTCTTTGATGCAAGGAAATGG
GCTGGCTGGAGCCAGGAGGGTGAGAAGCCAGGCCAAAGGGTAGGGGTCATA
CAGCTTCCTGCTTGGAGGACCAGGGCTGGAGCAATTTACCTCAGGCTAGTTG
GTGCTGGTGACCTCAACTCTGAGGGCCTCCTGGGGCTCGAGAGAAAATGAGT
GTCCTTTCTAAGCACAAAAGGTGATGAAGGTGCCAGTTTCCTGTGGGGAGCT
GAGCAGGCGTCGGCTGGGAGCTTCAGCCAGCCCAGCTCCAACCCGGCTGGGG
GGTGGGGAGCTTCCAGCTCCAAAGGAGAGATGAGTCTTCTAAACAGGAAATG
GTGACCCGCAGGCGGGAACAGCCTAATGACTTACAGCTTTCCTCCCAAAAGG
AAAAAAAAAAGTCTCTAACAATCGCCAAATTGTGGCAGGCCTCCCTGAAGGG
GAGGTTGGTGAGGTGGTATTGGGGGCTGGGCATCAACGAGCTGAAGGCCTAC
TGTGTCCCCAGGAGGTTCCTCCACATAGGTTGAACAGTGAGTGACTTCACCCC
GGGCCTGGAGGGATGGGATGG GCCTCCAGGACCAGACTCAGCGCTG

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

GGCTGGGTTCTTGACACACTAACCCAAGCAGAAAGCGATTCCGTTAGGCTCA
CCAAGTGACCTGTTCCACCCCCCCCCCCCCCCGACATCAGGGACAGAAGTTC
AGAAGGGGAGAAAAGACCTTGGCTAGGCACTGGCACTTGCCTTCTTCGGCCC
CGAGACCACCCAGGACCCGGGGTGTTGAGGGGGCAATCTGGAAGGATTTTAG
AACAGGGCAAGAACTGGGCTGGCCCCCTCCCCCTGACACCCTGGGTTGAGTG
ACGGGCACAGACCGGACCTTTCCTTCAGCCCCCAGAGAGTGGAGGGCGGAA
AATGCGATTCAAACCCGCCTATGAGCAGTGGGAGCAGTGGGCGGGGGCAGG
ACCTGGTTTCAGAAACCAACCGCTTGAGCTCCCAAGCCTCCTGCCCTGAGGT
GGCTCCAGACCACTGCGTTTATTTTGCTTTTGCAGGCAAGGCAGATGCTTTGC

TGGGTCCCGACTCAGCGAGATCTTTGCCCAACCATTTGCTTGCCTGCTCCACC
CCCTACTGGGGCACTATCCCGGGAGTCAGCTCTCCAGGCCTTTAAACCTTAGG
GAATGTCACACATGGAAACCTTTAGCAAATGTTTGTTAATGATCATAACAAA
GGCATCATTTAAATTAGGCAGGTAATTACTACCAGAAGAGCAACTGGGTGCT
CGATACTCATCCTTTTCTCCCCTTTAAACTCCCGAGGAATTACTGTCTCAGGA
TACCCGAGCCGGCCCGCTTTTCGGACTCCTTTAAAACGGCTCTGAAATGAACT
GTGAAGTTACCATTTGTGGGAGGAGAACGCAGGGGCCCTGCTTGAGGCAGGT
AGTGGGATTAGAGCTGGTAGGGTGTTCCTTGCAAAGCAGGCTCTCCTCTGCA
GACGCTCAGCCCCAAACTCACCCCCTCCTTGGGTACCTGCTACAAAAACCTA
AATCTTGACTTGCT

**Pdgfrα-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
CCAAGTGACCTGTTCCACCCCCCCCCCCCCCCGACATCAGGGACAGAAGTTC
AGAAGGGGAGAAAAGACCTTGGCTAGGCACTGGCACTTGCCTTCTTCGGCCC
CGAGACCACCCAGGACCCGGGGTGTTGAGGGGGCAATCTGGAAGGATTTTAG
AACAGGGCAAGAACTGGGCTGGCCCCCTCCCCCTGACACCCTGGGTTGAGTG
ACGGGCACAGACCGGACCTTTCCTTCAGCCCCCAGAGAGTGGAGGGCGGAA
AATGCGATTCAAACCCGCCTATGAGCAGTGGGAGCAGTGGGCGGGGGCAGG
ACCTGGTTTCAGAAACCAACCGCTTGAGCTCCCAAGCCTCCTGCCCTGAGGT
GGCTCCAGACCACTGCGTTTATTTTGCTTTTGCAGGCAAGGCAGGCACTTTGC
TGGGTCCCGACTCAGCGAGATCTTTGCCCAACCATCCACTTGCCTGCTCCACC
CCCTACTGGGGCACTATCCCGGGAGTCAGCTCTCCAGGCCTTTAAACCTTAGG
GAATGTCACACATGGAAACCTTTAGCAAGCATTTGTTAATGATCATAACAAA
GGCATCATTTAAATTAGGCAGGTAATTACTACCAGAAGAGCAACTGGGTGCT
CGATACTCATCCTTTTCTCCCCTTTAAACTCCCGAGGAATTACTGTCTCAGGA
TACCCGAGCCGGCCCGCTTTTCGGACTCCTTTAAAACGGCTCTGAAATGAACT
GTGAAGTTACCATTTGTGGGAGGAGAACGCAGGGGCCCTGCTTGAGGCAGGT
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 (Pdgfrα), 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-β-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 |
| Western Blotting | |
| Mouse Anti-Flag M2 | Sigma, Cat# F3165, 1:500 |
| Mouse anti-Actin | Chemicon, Cat# MAB1501, 1:1000 |
| ChIP-seq | |
| Mouse Anti-Flag M2 | Sigma, Cat# F3165 |
| ChIP-qPCR | |
| Mouse anti-Msgn1 Ascites (Chalamalasetty et al., 2011) raised against a cocktail of three peptides in Msgn1 protein:<br>CWKSRARPLELVQESP (16aa)<br>CDLLNSSGREPRPQSV (16aa)<br>CSHEAAGLVELDYS (14aa) | 2µl per ChIP assay |
| Mouse control Ascites,<br>Clone: NS-1 | Sigma, Cat# M8273, 2µl per ChIP assay |

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

| | |
|---|---|
| RT-PCR | |
| 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 | GCAGTGACCACCATTCCTG |
| 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 Msgn1.RV | GACCTCAGAGTAGGACTCCA |
| RT-Tbx6.FW | CCTTCCGATTTCCTGAGACCACAT |
| RT-Tbx6.RV | AGGTCCAGAAATGCAGCTGAGTAG |
| RT-Pax3.FW | TTTCACCTCAGGTAATGGGACT |
| RT-Pax3.RV | GAACGTCCAAGGCTTACTTTGT |
| RT-OSR1.FW | CCGCCACTTCACTAAGTCCTAT |

| | |
|---|---|
| RT-OSR1.RV | GAGTGTAGCGTCTTGTGGACAG |
| RT-Pdgfrα.FW | TCCATGCTAGACTCAGAAGTCA |
| RT-Pdgfrα.RV | TCCCGGTGGACACAATTTTTC |
| RT-Foxc1.FW | CCCGTTCTTTCGACATAGGA |
| RT-Foxc1.RV | CAGAGACTCGCTTTCCTGCT |
| RT-Ephrin B1.FW | TGTGGCTATGGTCGTGCTG |
| RT-Ephrin B1.RV | CCAAGCCCTTCCCACTTAGG |
| RT-Dact1.FW | AAGAGATGCCGGTTTGTTGAA |
| RT-Dact1.RV | GCCCGAAGCTCCATCACTC |
| RT-Epha4.FW | CCAGACAGAGTAGCTGTGCG |
| RT-Epha4.RV | CGAGGAGCAGAATGGTGAAT |
| RT-Gata4.FW | AAGGCAGAGAGTGTGTCAATTGTGG |
| RT-Gata4.RV | TGGTAGTCTGGCAGTTGGCACAG |
| ChIP-qPCR | |
| ChIP-Snai1-3'enh.FW | AAGGGAGGATGTCTTTCCTGTT |
| ChIP-Snai1-3'enh.RV | ATTTCCTTGCATCAAAGACCATT |
| ChIP-Tbx6-3'enh.FW | CCCCCCACCAAATGGCTGGCTT |
| ChIP-Tbx6-3'enh.RV | GCAGGCCCCAGCTGGACTCAGA |
| EMSA | |
| DIG-Pdgfrα-Ebox1 | CTTTTGCAGGCAAGGCAGATGCTTTGCTGG GTC |
| DIG-Pdgfrα-Ebox2 | TGCCCAACCATTTGCTTGCCTGCTCCA |

| | |
|---|---|
| DIG-Pdgfrα-Ebox3 | AGGGAATGTCACACATGGAAACCTTTAG |
| DIG-Pdgfrα-Ebox4 | GAAACCTTTAGCAAATGTTTGTTAATGA |
| DIG-Tbx6-Ebox1 | CCCCCCACCAAATGGCTGGCTT |
| DIG-Tbx6-Ebox2 | TCTGAGTCCAGCTGGGGCCTGC |
| DIG-Tbx6-Ebox3 | TGCCTGGCCAAATGGGGCTCCC |
| DIG-Tbx6-Ebox4 | TTAAGGTCCAGCTGCTTGCCCT |
| Cloning | |
| Msgn1-Pst1-Flag.FW | CCCTGCAGACCGCCGCCACCATGGACTACAAAG |
| Msgn1-Sal1.RV | CCGTCGACTCACACACTCTGTGGCCTGGGCTCTCTC |
| Xho1-3x-Flag-Mlu1-Sac1-Afl2-Hind3 | 5'TCGAGACCGCCGCCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGACGCGTGAGCTCCTTAAGA 3' |
| Msgn1-Mlu1.FW | CCACGCGTGACAACCTGGGTGAGACCTTCCTCAGC |
| Msgn1-EcoR1.RV | CCGAATTCTCACACACTCTGTGGCCTGGGCTCTCTC |
| Tbx6-Mlu1.FW | CCACGCGTTACCATCCACGAGAGTTGTACCCCTCC |
| Tbx6-Sal1.RV | CCGTCGACTCAGTACATTGGCTTGGATCCCATGTCCAG |
| Snai1 +5148 -Nhe1-Pst1.FW | CGCTAGCTGCAGCAAGCTTCCTGAACTCTTAGTTCTTC |
| Snai1 +5148-EcoRV.RV | CCGATATCCAGAACCAGGCTGACCAAGTCCTA |
| Tbx6 +3913-Nhe1-Pst1.FW | CGCTAGCTGCAGGGACCTGGAGGTTATCTGGACAT |
| Tbx6 +3913-EcoRV.RV | CCGATATCCTTGATTCTTAAAGCAACCATTGTG |

| | |
|---|---|
| Pdgfrα -2140-Nhe1-Pst1.FW | GCGCTAGCCTGCAGCTCCAGACCACTGCG |
| Pdgfrα -2140-EcoRV.RV | GCGATATCGTATCCTGAGACAG |
| Mesp2 -300bp-Kpn1.FW | CCGGTACCGTAGCAGGGCGGAGTGAAGGTG GGGGCTCG |
| Mesp2 -300-Xho1.RV | CCCTCGAGGGCTGGTGTGACTCTGGGAAGC TGGACCGG |
| Genotyping | |
| 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.