Supplemental Material

Hornbachner et al.,

MSX2 safeguards syncytiotrophoblast fate of human trophoblast stem cells

List of Material provided:

Supplementary Information

Additional details on Materials & Methods

Chromatin immunoprecipitation

Cells (1-2x10^8) were fixed in 2mM Di(N-succinimidyl) glutarate (DSG) (80424, Sigma) in PBS at room temperature (RT) for 45 min. After washing with PBS, cells were fixed again in 1% formaldehyde (28908, Sigma) in TS base media at RT for 12 min. Fixation was stopped by adding glycine to a final concentration of 0.125 M. Cells were washed twice with PBS and resuspended in wash buffer 1 (10mM Hepes pH7.5, 10mM EDTA, 0.5mM EGTA and 0.75% Triton X-100) and incubated at 4C for 10min. After pelleting, cells were resuspended in wash buffer 2 (10mM Hepes pH7.5, 200mM NaCl, 1mM EDTA and 0.5mM EGTA) and incubated at 4°C for 10min. After pelleting, cells were lysed in the lysis/sonication buffer (150mM NaCl, 25mM Tris pH 7.5, 5mM EDTA, 0.1%Triton, 1% SDS and 0.5% sodium deoxycholate) with cOmplete™ EDTA-free Protease Inhibitor Cocktail (11873580001, Sigma) on ice for 30min. Chromatin was sonicated 15s on /30s off for 23-25 cycles using the UW2070 sonicator (Bandelin) to the average 300-bp fragments. Chromatin was diluted 1/10 with the dilution buffer (150mM NaCl, 25mM Tris pH 7.5, 5mM EDTA, 1% Triton X-100, 0.1% SDS and 0.5 % sodium deoxycholate) containing complete protease inhibitors. Protein G magnetic Dynabeads (10004D, Invitrogen) were blocked with 1mg/ml BSA and tRNA at 4°C for 1h and washed with buffer A (150mM NaCl, 25mM Tris pH 8.0, 1mM EDTA, 1% NP-40, 0.1 % SDS and 0.5 % sodium deoxycholate). Chromatin was pre-cleared with pre-blocked beads at 4°C for 1h. Four hundred micrograms of chromatin and seven micrograms of antibody rabbit anti-MSX2 (HPA005652, Sigma) and rabbit normal IgG (NI01, Sigma) were used per IP. IP was performed overnight at 4°C with rotation. Pre-blocked magnetic beads were added next morning for 7-8 h. Beads were washed at 4°C with buffer A three times, buffer B (50mM Tris pH 8.0, 500mM NaCl, 0.1% SDS, 0.5 % sodium deoxycholate, 1% NP-40 and 1mM EDTA), buffer C (50mM Tris pH 8.0, 250mM LiCl, 0.5% sodium deoxycholate, 1% NP-40 and 1mM EDTA) and rinsed with TE buffer. DNA was eluted from beads in the elution buffer (1% SDS,

0.1M $NaHCO_3$). Samples were treated with RNAse A and Proteinase K and reverse-crosslinked overnight by incubation for 2h at 37°C, followed by 12h at 65C. DNA was purified on the PCR purification columns (Qiagen). To generate a library, DNA from 2-4 IPs was pooled and the NEBNext Ultra II DNA library preparation master mix (New England Biolabs, E7645) was used according to the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 2500 sequencer with 50-base pair (bp) single-end protocol.

For ARID1A, SMARCA4, H3K27ac, H3K27me3 and H3K9me3 ChIP-seq on WT and MSX2 KD cells, cells were fixed in 1% FA for 10min and fixation was quenched by 0.125 M glycine at RT. Cells were scraped in PBS, centrifuged at 500g for 5min and pellet was resuspended in an ice-cold IP-buffer, a 2:1 mixture of SDS-buffer (100mM NaCl, 50mM Tris-Cl pH 8.0, 5mM EDTA pH 8.0, 0.5% SDS) and Triton-X-buffer (100mM NaCl, 100mM Tris-Cl pH 8.6, 5mM EDTA pH 8.0, 5% Triton X), containing complete Protease inhibitor tabs (11873580001, Sigma). After 30min incubation on ice chromatin was sonicated 15s on (pulsed, with 90% actual sonication time) /30s off for 20 cycles using the UW2070 sonicator (Bandelin) at ~40% power. 30µg and 10µg chromatin (DNA equivalent) were diluted to 500µl and precipitated at 4°C overnight using the antibodies against ARID1A (12354, Cell Signaling), SMARCA4 (ab110641, abcam) and H3K27ac (39685, Active Motif), H3K27me3 (07-442, Merck Millipore), H3K9me3 (07-449, Merck Millipore) respectively. For ChIPs of histone modifications we added 0.8µg Spike-In antibody (61686, Active Motif) and 25ng Spike-In chromatin per ChIP (53083, Active Motif), (1). 50µl Dynabeads Protein G (10004D, Invitrogen) equilibrated in IP-buffer, blocked with 1mg/ml BSA and tRNA at 4°C for 1h and washed with IP-buffer were added to the IP reaction and incubated for 3h at 4°C. ChIPs were washed three times with Low Salt Buffer (50mM HEPES pH 7.5, 140mM NaCl, 1% Triton X), once with High Salt Buffer (50mM HEPES pH 7.5, 500mM NaCl, 1% Triton X) and once with TE buffer before elution with 200µl Elution Buffer (1% SDS, 100mM $NaHCO_3$) per reaction. Eluate was separated from the beads and decrosslinked over night at 65°C. After RNaseA and proteinase K treatment 5ng of DNA were subjected to library prep (see above). Libraries were sequenced paired-end 50bp on an Illumina NovaSeq 6000 SP1 flowcell.

Bioinformatic analysis

RNA quantification, QuantSeq

Libraries generated with the QuantSeq FWD kit (Lexogen) were sequenced on an Illumina HiSeq2500v4 at single read 50bp. Reads were trimmed with Trim Galore! (0.6.5). Quality control was performed using fastQC (0.11.9) (2). Transcripts were mapped to the hg38 human reference genome using STAR (2.7.3a) (3). After indexing with samtools (1.10) reads in genes were counted using Rsubread (2.0.0) (4). Differential expression analysis was conducted using DESeq2 (1.24.0) (5). MA plots were generated with ggplot2 (3.3.3) and the heatmaps using pheatmap (1.0.12) (6).

Chromatin Immunoprecipitation Sequencing

Libraries were sequenced on the Illumina HiSeq2500 v4 at single read 50bp or on the Illumina NovaSeq 6000 SP1 at paired-end 50bp aiming for a sequencing depth of 20 - 40 million reads/sample. Raw reads were trimmed with CUTADAPT (2.8) and aligned to the human reference genome GRCh38 with bowtie 2 (2.3.5.1) (7). Alignments were further processed using samtools (1.10). Peaks were called using MACS2 (2.2.7.1) with default parameters taking uniquely mapped reads with a merged input as the control for MSX2 ChIPseq, while separate replicated inputs were used for ARID1A, BRG1, H3K27ac, H3K27me3 and H3K9me3 ChIPseq (8). For paired-end sequencing the software was run in paired-end mode. High confidence peaks were generated by IDR (1.2), using a cut-off of FDR<0.05 across all replicates via the ChIPpeakAnno package (3.26.0) (9, 10). Overlaps between ChIPseq peaks were computed by findOverlapsOfPeaks and makeVennDiagram. For visualization we generated RPKM normalized coverage files for MSX2, ARID1A and SMARCA4 with deeptools (3.5.0) and bamCoverage (11), while ChIPseq data for histone modifications were additionally normalized by drosophila spike-in reads (1). ChIPseq tracks were visualized using Galaxy and UCSC genome browser custom tracks (11, 12).

Integration of ChIPseq and Quantseq data

Feature distribution was determined, and peaks were annotated with ChIPseeker (1.26.0) with following priorities: Promoter, 5'UTR, 3'UTR, Exon, Intron, Downstream, Intergenic (9). Gene symbols were retrieved using biomaRt (13) and gene overlaps between ChIPseq and/or Quantseq data were plotted using Eulerr (6.1.0) or UpSetR (1.4.0) (14, 15). After extraction of specific gene sets functional enrichment analysis was conducted as described below.

Differential ChIP enrichment analysis

Differential ChIP enrichment analysis was done using DiffBind (3.2.3) (16). All MSX2 peaks found in at least one replicate were considered, i.e. reads for each ChIP-seq (ARID1A, SMARCA4 and H3K27ac) were counted within a 400 bp window around MSX2 peaks. Correlation heatmaps of reads in ChIP-seq peaks were generated using DiffBind considering all peaks found in at least one sample. MA plots showing differential enrichment of ChIP-seq signals were constructed using the internal DiffBind function.

Functional Enrichment Analysis

GREAT analysis was conducted via the GREAT webpage (17) and visualized using ggplot2 (18). EnrichR analysis (19–21) was run on genes annotated by ChIPseeker or gene sets from Quantseq overlaps and visualized using custom R scripts. If not stated otherwise, P-values of enrichment analyses were BH adjusted and only terms with an adjusted P-value < 0.05 were considered.

Genetic modifications

To deplete MSX2, we generated pLKO.1-based constructs containing either shRNA against MSX2 (TRCN0000234848 and TRCN0000234849) or a shRNA against GFP as a control by cloning annealed oligos into the AgeI/EcoRI sites of pLKO.1-neo (13425, Addgene). The following plasmids were simultaneously transfected into HEK293T cells: pLKO.1 (individual constructs containing different shRNA), psPAX2 (encodes Gag and Pol sequences to package the lentivirus), and pMD2.G (encodes the G protein envelope protein) using lipofectamine 3000 (Thermo Fisher Scientific). Supernatant was collected after 48h and used to transduce hTSC for a minimum of 16h. Cells were selected using 300 µg/ml G418 (A1720, Sigma).

Cell culture conditions

Briefly, the hTSCs were cultured in the basal media (DMEM/F12 (11320074, Thermo Fisher), 1x ITS-X (51500056, Thermo Fisher), 0.2% FBS (PAA), 1.5 µg/mL L-ascorbic acid (A4403, Sigma)) supplemented with 2 µM CHIR99021 (2520691, Peprotech), 50ng/mL EGF (AF-100-15, Peprotech), 0.5 µM A83-01 (9094360, Peprotech), and 5µM Y27632 (1293823, Peprotech), on dishes coated with 10µg/ml Fibronectin (FC0101, Sigma). Cells were passaged every 3-4 days using TrypLE (12605036, Thermo Scientific). ST differentiation was induced in basal medium supplemented with 2 µM Forskolin (6652995, Peprotech), 5 µM Y27632 and 4% KnockOut Serum Replacement (10828010, Thermo Scientific).

Mass Spectrometry

Liquid Chromatography mass spectrometry (LC-MS) was performed on an UltiMate 3000 RSLC nano system (Thermo Scientific) coupled to a Q Exactive HF-X mass spectrometer (Thermo Scientific) and equipped with a Nanospray Flex Ion Source (Thermo Scientific). For peptide identification, the RAW-files were processed using Proteome Discoverer (2.3.0.523) (Thermo Scientific). The MS/MS spectra were searched against the Uniprot human proteome database 2020-10-12 (20,536 sequences; 11,395,384 residues) and the default list of common contaminants using MSAmanda (v2.0.0.12368) (12). Beta-methylthiolation on cysteine was set as a fixed modification. Localization of post-translational modification sites within the peptides was performed using the function ptmRS of phospho-RS (13). The maximal number of missed cleavages was set to 2, using tryptic enzymatic specificity. The results were filtered to 1 % FDR on the protein level using the Percolator algorithm integrated in Proteome Discoverer. Peptide areas were quantified using apQuant (14).

Primer sequences

RTQPCR

hPBGD-1F GGAGCCATGTCTGGTAACGG

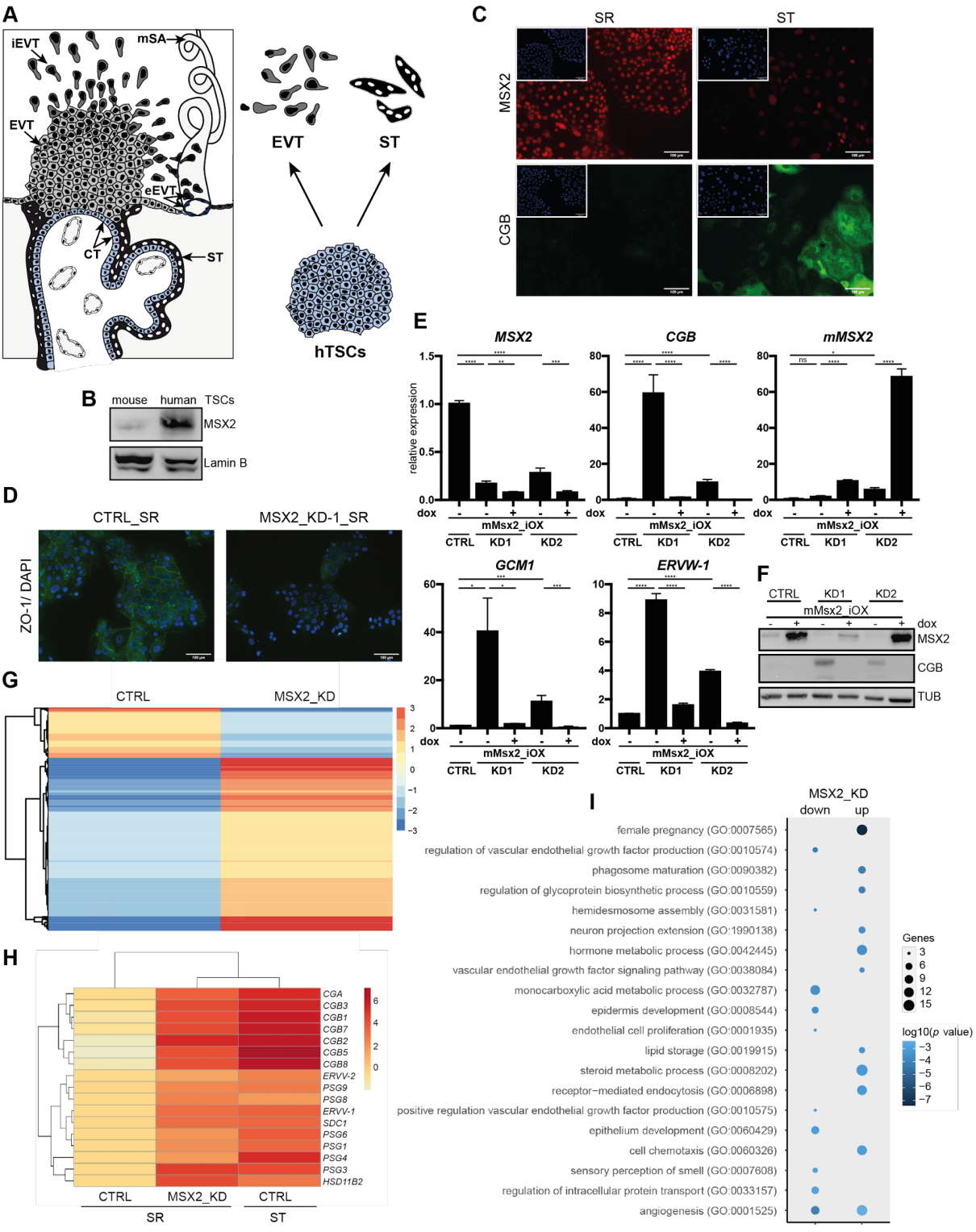
hPBGD-1R CCACGCGAATCACTCTCATCT

hMSX2-1F GCAGGAACCCGGCCGATATT

hMSX2-1R	*CTGACGGAACTTGCGCTCCA*
hCGB-F	*CAGCATCCTATCACCTCCTGGT*
hCGB-R	*CTGGAACATCTCCATCCTTGGT*
hGCM1-1F	*GCTGGGACTTGAACCAGCAGT*
hGCM1-1R	*CTGGATCGGCCCACTCAAGC*
hSDC1_F	*CTATTCCCACGTCTCCAGAACC*
hSDC1_R	*GGACTACAGCCTCTCCCTCCTT*
ERVW-1_1F	*CTACCCCAACTGCGGTTAAA*
ERVW-1_1R	*GGTTCCTTTGGCAGTATCCA*
hTEAD4_F	*GCCAGTCCAGCCCAAGCTAC*
hTEAD4_R	*CATTGGAGGGTCCCCGTTCG*

shRNAs

CTRL	*GCAAGCTGACCCTGAAGTTCAT*
MSX2_KD1	*AGCGCAAGTTCCGTCAGAAAC*
MSX2_KD2	*TGCAGGCAGCGTCCATATATG*


A
iEVT
mSA
EVT
eEVT
CT
ST
EVT
ST
hTSCs
B
mouse
human
TSCs
MSX2
Lamin B
C
SR
ST
MSX2
CGB
D
CTRL_SR
MSX2_KD-1_SR
ZO-1/ DAPI
E
MSX2
CGB
mMSX2
GCM1
ERVW-1
relative expression
dox
mMsx2_iOX
CTRL
KD1
KD2
F
CTRL
KD1
KD2
mMsx2_iOX
dox
MSX2
CGB
TUB
G
CTRL
MSX2_KD
H
CGA
CGB3
CGB1
CGB7
CGB2
CGB5
CGB8
ERVV-2
PSG9
PSG8
ERVV-1
SDC1
PSG6
PSG1
PSG4
PSG3
HSD11B2
CTRL
MSX2_KD
CTRL
SR
ST
I
MSX2_KD
down
up
female pregnancy (GO:0007565)
regulation of vascular endothelial growth factor production (GO:0010574)
phagosome maturation (GO:0090382)
regulation of glycoprotein biosynthetic process (GO:0010559)
hemidesmosome assembly (GO:0031581)
neuron projection extension (GO:1990138)
hormone metabolic process (GO:0042445)
vascular endothelial growth factor signaling pathway (GO:0038084)
monocarboxylic acid metabolic process (GO:0032787)
epidermis development (GO:0008544)
endothelial cell proliferation (GO:0001935)
lipid storage (GO:0019915)
steroid metabolic process (GO:0008202)
receptor−mediated endocytosis (GO:0006898)
positive regulation vascular endothelial growth factor production (GO:0010575)
epithelium development (GO:0060429)
cell chemotaxis (GO:0060326)
sensory perception of smell (GO:0007608)
regulation of intracellular protein transport (GO:0033157)
angiogenesis (GO:0001525)
Genes
log10(p value)

Supplemental Figure S1

Depletion of MSX2 results in loss of silencing of syncytiotrophoblast genes.

(A) Schematic representation of a fragment of a villous tree in human placenta with main trophoblast cell types marked and their *in vitro* counterpart derivatives. EVT: extravillous trophoblast, iEVT: interstitial EVT, eEVT: endovascular EVT, ST: syncytiotrophoblast, CT: cytotrophoblast, mSA: maternal spiral artery, hTSCs: human trophoblast stem cells.

(*B*) Western blot analysis for MSX2 in mouse and human trophoblast stem cells. Lamin B serves as loading control.

(C) Immunostaining of self-renewing (SR) and 3 day syncytiotrophoblast differentiated (ST) hTSCs for MSX2 and CGB. DAPI (insets) indicates nuclei.

(*D*) Immunostaining for ZO-1 of MSX2-depleted (MSX2_KD-1_SR) and control (CTRL) cells cultured in self-renewal (SR) conditions. DAPI indicates nuclei.

(E) RTQPCR analysis of control (CTRL) and MSX2-depleted (KD1 and KD2) hTSCs carrying a doxycycline (dox) inducible mouse Msx2 transgene (mMsx2_iOX) and cultured in self-renewal conditions in the presence (+dox) or absence (-dox) of dox. The bars represent a mean of three biological replicates (n=3) with S.E.M., expression in CTRL was set to 1. ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05, ns: not significant. MSX2 indicates expression of the human, endogenous MSX2, while mMsx2 of the mouse Msx2 transgene.

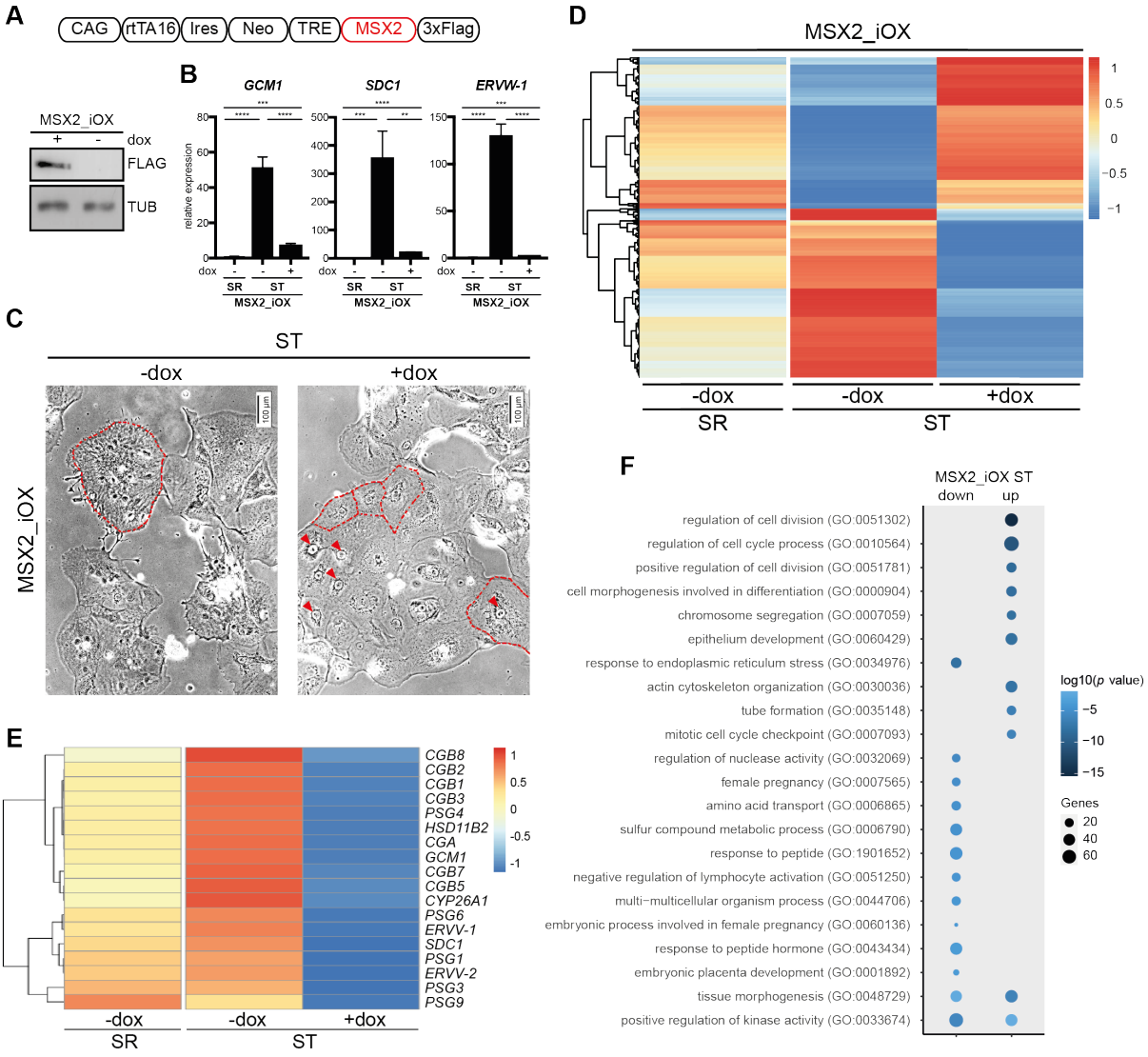
(F) Western blot analysis for CGB and MSX2 of control (CTRL) and MSX2-depleted (KD1 and KD2) hTSCs carrying a doxycycline (dox) inducible mouse Msx2 transgene (mMsx2_iOX), cultured in self-renewal conditions in the presence (+dox) or absence (-dox) of dox. Tubulin (TUB) serves as a loading control.

(G) Hierarchical clustering of genome-wide RNA profiles of MSX2-depleted (MSX2_KD: MSX2_KD1 (n=3) and MSX2_KD2 (n=1)) and control (CTRL (n=3) hTSCs cultured in self-renewal conditions. Analysis (|log2FC|>1, *p* adj<0.05) revealed 522 up-regulated and 152 down-regulated genes in MSX2_KD lines compared to control.

(H) Clustered heat map depicting expression of selected syncytiotrophoblast markers in control (CTRL) and MSX2-depleted (MSX2_KD) hTSC lines cultured in self-renewal

conditions (SR) (details as in A.) and, as comparison, in cells differentiated for 6 days to syncytiotrophoblast (ST).

(l) Gene Ontology (GO) term enrichment for biological processes of de-regulated genes in MSX2_KD compared to control (CTRL), based on RNA-Seq analysis as in Supplemental Fig.1G.


A
CAG
rtTA16
Ires
Neo
TRE
MSX2
3xFlag
MSX2_iOX
dox
FLAG
TUB
B
GCM1
SDC1
ERVW-1
relative expression
dox
SR
ST
MSX2_iOX
C
ST
-dox
+dox
MSX2_iOX
100 µm
D
MSX2_iOX
-dox
-dox
+dox
SR
ST
E
CGB8
CGB2
CGB1
CGB3
PSG4
HSD11B2
CGA
GCM1
CGB7
CGB5
CYP26A1
PSG6
ERVV-1
SDC1
PSG1
ERVV-2
PSG3
PSG9
-dox
-dox
+dox
SR
ST
F
MSX2_iOX ST
down
up
regulation of cell division (GO:0051302)
regulation of cell cycle process (GO:0010564)
positive regulation of cell division (GO:0051781)
cell morphogenesis involved in differentiation (GO:0000904)
chromosome segregation (GO:0007059)
epithelium development (GO:0060429)
response to endoplasmic reticulum stress (GO:0034976)
actin cytoskeleton organization (GO:0030036)
tube formation (GO:0035148)
mitotic cell cycle checkpoint (GO:0007093)
regulation of nuclease activity (GO:0032069)
female pregnancy (GO:0007565)
amino acid transport (GO:0006865)
sulfur compound metabolic process (GO:0006790)
response to peptide (GO:1901652)
negative regulation of lymphocyte activation (GO:0051250)
multi−multicellular organism process (GO:0044706)
embryonic process involved in female pregnancy (GO:0060136)
response to peptide hormone (GO:0043434)
embryonic placenta development (GO:0001892)
tissue morphogenesis (GO:0048729)
positive regulation of kinase activity (GO:0033674)
log10(p value)
Genes

Supplemental Figure S2

Ectopic expression of MSX2 blocks syncytiotrophoblast cell fate.

(A) Schematic depiction of a doxycycline (dox) inducible construct carrying a 3xFlag-tagged coding sequence of human MSX2 and a Western blot showing expression of this transgene upon 48h of dox treatment.

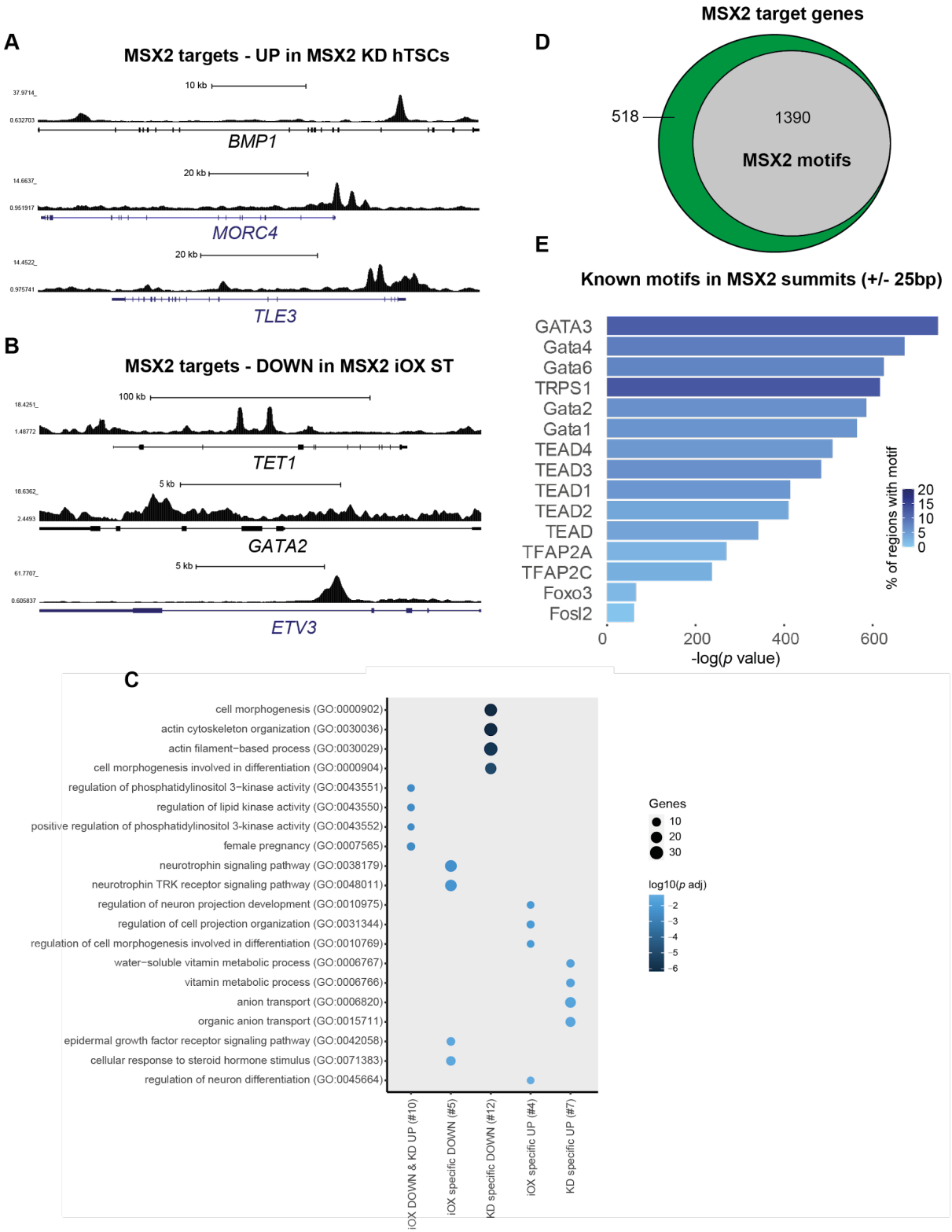
(B) RTQPCR analysis for syncytiotrophoblast markers *GCM1*, *SDC1* and *ERVW-1* in hTSC line carrying a dox inducible MSX2 transgene and differentiated to syncytiotrophoblast (ST) in the presence (+dox) or absence (-dox) of dox. The bars represent an average of three independent biological replicates (n=3) with S.E.M.; expression in self-renewal conditions (SR) was set to 1. ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05, ns: not significant.

(C) Bright field images of hTSCs carrying a doxycycline (dox) inducible MSX2 transgene (MSX2_iOX) differentiated for 6 days into syncytiotrophoblast (ST) in the presence (+dox) or absence (-dox) of dox. The red, dashed lines indicate syncytium (left panel) and single cells (right panel); arrowheads indicate enlarged nuclei (right panel).

(D) Hierarchical clustering of genome-wide expression profiles of hTSC carrying a doxycycline (dox) inducible MSX2 transgene (MSX2_iOX) differentiated for 6 days into syncytiotrophoblast (ST) in the presence (+dox) or absence (-dox) of doxycycline. Analysis (|log2FC|>1, *p* adj<0.05) revealed 1285 up-regulated and 1433 down-regulated genes in MSX2_iOX (+dox) compared to (-dox) control. For comparison, levels of these differentially expressed genes in -dox, self-renewal conditions (SR) are displayed.

(E) Clustered heat map depicting expression of selected ST markers as in Supplemental Fig. S2D.

(F) Gene Ontology (GO) term enrichment for biological processes of de-regulated genes in MSX2_iOX ST (+dox) in comparison to MSX2_iOX ST (-dox), based on RNA-Seq analysis as in Supplemental Fig. S2D.


A
MSX2 targets - UP in MSX2 KD hTSCs
10 kb
BMP1
20 kb
MORC4
20 kb
TLE3
B
MSX2 targets - DOWN in MSX2 iOX ST
100 kb
TET1
5 kb
GATA2
5 kb
ETV3
D
MSX2 target genes
518
1390
MSX2 motifs
E
Known motifs in MSX2 summits (+/- 25bp)
GATA3
Gata4
Gata6
TRPS1
Gata2
Gata1
TEAD4
TEAD3
TEAD1
TEAD2
TEAD
TFAP2A
TFAP2C
Foxo3
Fosl2
0
200
400
600
-log(p value)
% of regions with motif
20
15
10
5
0
C
cell morphogenesis (GO:0000902)
actin cytoskeleton organization (GO:0030036)
actin filament−based process (GO:0030029)
cell morphogenesis involved in differentiation (GO:0000904)
regulation of phosphatidylinositol 3−kinase activity (GO:0043551)
regulation of lipid kinase activity (GO:0043550)
positive regulation of phosphatidylinositol 3-kinase activity (GO:0043552)
female pregnancy (GO:0007565)
neurotrophin signaling pathway (GO:0038179)
neurotrophin TRK receptor signaling pathway (GO:0048011)
regulation of neuron projection development (GO:0010975)
regulation of cell projection organization (GO:0031344)
regulation of cell morphogenesis involved in differentiation (GO:0010769)
water−soluble vitamin metabolic process (GO:0006767)
vitamin metabolic process (GO:0006766)
anion transport (GO:0006820)
organic anion transport (GO:0015711)
epidermal growth factor receptor signaling pathway (GO:0042058)
cellular response to steroid hormone stimulus (GO:0071383)
regulation of neuron differentiation (GO:0045664)
iOX DOWN & KD UP (#10)
iOX specific DOWN (#5)
KD specific DOWN (#12)
iOX specific UP (#4)
KD specific UP (#7)
Genes
10
20
30
log10(p adj)
-2
-3
-4
-5
-6

Supplemental Figure S3

MSX2 binds the syncytiotrophoblast genes in hTSCs.

(*A*) Genome browser snapshots for MSX2 targets identified among the upregulated genes in hTSCs upon MSX2_KD.

(B) Genome browser snapshots for MSX2 targets identified among the downregulated genes in ST upon MSX2 overexpression (MSX2_iOX ST).

(C) Gene Ontology (GO) term enrichment of specific groups of MSX2 target genes derived from the UpSet plot in Fig. 3C. #10, genes that were conversely misregulated in MSX2_KD hTSCs and MSX2_iOX STs; #5, genes that were only downregulated in MSX2_iOX STs but not significantly regulated in MSX2_KD hTSCs; #12, genes that were only downregulated in MSX2_KD hTSCs but not significantly regulated in MSX2_iOX STs; #4, genes that were only upregulated in MSX2_iOX STs but not significantly regulated in MSX2_KD hTSCs and #7, genes that were only upregulated in MSX2 KD hTSCs but not significantly regulated in MSX2 iOX STs.

(D) Number of MSX2 target genes with the MSX2 motif, identified by HOMER *de novo* motif detection. Total targets, blue; MSX2 motif containing targets, grey.

(E) Bar graph showing the identified known TF motifs around the MSX2 peak summits. The bar length represents -log(*p* value) of enrichment and the colour scale indicates percentage of regions containing the respective TF motif.

A

				MSX2						IgG					
Gene	Alias	log2FC	p-value	# unique peptides			% coverage			# unique peptides			% coverage		
MSX2	MSX2	8.95	6.33E-07	3	3	2	22.5	22.1	16.9	0	0	0	0.0	0.0	0.0
SMARCE1	BAF57	7.36	3.28E-06	10	10	11	28.7	26.5	32.9	0	0	0	0.0	0.0	0.0
SMARCC2	BAF170	7.33	3.97E-06	21	23	20	25.6	28.6	24.1	3	4	0	2.6	3.5	0.0
SMARCC1	BAF155	7.18	2.59E-05	23	31	31	34.3	37.5	35.5	0	2	0	1.7	4.9	0.0
ARID1A	BAF250A	7.10	7.10E-06	31	31	42	20.1	20.0	24.6	0	1	0	0.0	0.0	0.0
SMARCD2	BAF60B	6.95	4.45E-06	18	13	15	43.9	34.3	38.4	0	0	0	0.0	0.0	0.0
SMARCA4	BRG1	6.85	7.48E-06	12	17	20	16.8	20.3	19.7	0	0	0	0.0	0.0	0.0
SMARCB1	BAF47	6.78	5.76E-06	11	12	7	33.5	42.1	27.5	0	0	0	0.0	0.0	0.0
ACTL6A	BAF53A	6.47	8.01E-06	11	11	9	32.4	37.5	25.4	0	0	0	0.0	0.0	0.0
DPF2	BAF45D	6.27	1.79E-05	8	8	11	23.0	31.0	32.5	0	0	0	0.0	0.0	0.0
SMARCA2	BRM	4.44	1.66E-04	9	10	9	16.0	15.6	13.3	0	0	0	0.0	0.0	0.0
SMARCD1	BAF60A	4.74	9.02E-05	9	10	9	23.9	28.2	24.3	0	0	0	0.0	0.0	0.0

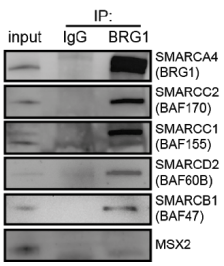
B

				ARID1A						IgG					
Gene	Alias	log2FC	p value	# unique peptides			% coverage			# unique peptides			% coverage		
ARID1A	BAF250A	10.71	1.72E-04	48	99	114	28.1	34.1	36.1	0	0	0	0.0	0.0	0.0
SMARCC1	BAF155	9.56	2.47E-05	25	41	35	32.3	40.0	34.9	8	3	9	12.6	5.7	15.6
DPF2	BAF45D	7.75	1.08E-04	7	13	11	23.0	36.8	34.8	0	0	0	0.0	0.0	0.0
SMARCE1	BAF57	7.66	1.72E-04	9	13	20	28.0	38.2	45.5	1	0	0	2.4	0.0	0.0
SMARCD2	BAF60B	7.40	1.37E-04	9	12	14	22.8	21.1	25.2	1	1	1	1.9	1.9	1.9
SMARCC2	BAF170	7.38	4.17E-04	13	18	29	19.4	23.0	27.7	4	1	4	5.5	2.7	6.3
SMARCB1	BAF47	6.54	1.07E-04	7	6	9	17.9	20.8	19.0	2	0	1	5.7	0.0	3.4
SMARCA2	BRM	5.74	1.42E-03	11	17	18	11.1	15.2	17.2	5	3	1	6.7	3.6	2.2
ACTL6A	BAF53A	3.79	3.96E-02	5	14	14	14.7	29.4	26.8	4	1	1	14.7	2.3	3.0

C

				BRG1						IgG					
Gene	Alias	log2FC	p value	# unique peptides			% coverage			# unique peptides			% coverage		
SMARCC1	BAF155	11.6	5.6671E-06	37	47	27	45.4	45.2	27.2	8	3	9	12.6	5.7	15.6
SMARCA4	BRG1	11.5	2.0959E-05	30	48	47	32.0	32.2	26.7	0	0	1	2.6	0.7	2.2
SS18		9.9	0.00010526	3	7	3	7.2	12.4	4.8	1	0	0	4.3	0.0	0.0
DPF2	BAF45D	9.8	8.4602E-06	10	15	9	35.3	41.2	18.7	0	0	0	0.0	0.0	0.0
SMARCC2	BAF170	9.7	3.0104E-05	23	47	23	26.1	41.5	20.3	4	1	4	5.5	2.7	6.3
SMARCE1	BAF57	9.6	2.9191E-05	17	21	14	41.6	45.3	31.6	1	0	0	2.4	0.0	0.0
ACTL6A	BAF53A	9.4	3.5801E-05	15	18	14	40.8	36.8	25.2	4	1	1	14.7	2.3	3.0
SMARCD2	BAF60B	9.1	2.1762E-05	17	18	10	42.7	42.4	22.0	1	1	1	1.9	1.9	1.9
SMARCB1	BAF47	9.1	2.746E-05	9	9	8	22.9	27.8	18.4	2	0	1	5.7	0.0	3.4
ARID2	BAF200	7.7	0.00048737	9	17	10	7.0	13.5	5.6	1	1	1	0.5	0.5	0.5
BCL7B		7.1	0.00077619	0	5	0	5.0	27.7	5.0	0	0	0	0.0	0.0	0.0
SMARCD1	BAF60A	6.9	0.0007143	8	10	8	27.8	27.8	20.8	0	0	0	0.0	0.0	0.0
PBRM1	BAF180	6.6	0.00188292	9	27	15	6.5	19.3	9.5	0	0	0	0.0	0.0	0.0
BCL7C		5.9	0.00122949	2	1	2	18.4	12.9	18.4	0	0	0	0.0	0.0	0.0
BRD7		5.8	0.00114908	4	9	6	10.1	20.7	12.3	0	0	0	0.0	0.0	0.0
PHF10	BAF45A	5.4	0.00475248	3	7	3	6.4	14.1	7.0	0	0	0	0.0	0.0	0.0
BRD9		5.4	0.00703802	2	4	2	5.5	7.9	4.7	0	0	0	0.0	0.0	0.0
ARID1A	BAF250A	5.0	0.0001725	31	52	38	20.1	30.9	18.9	0	0	0	0.0	0.0	0.0
ARID1B	BAF250B	5.0	0.00035688	2	9	4	2.1	6.0	2.2	0	0	0	0.0	0.0	0.0
DPF3	BAF45C	4.0	0.03327768	3	3	1	9.0	8.2	2.4	0	0	0	0.0	0.0	0.0

D



Supplemental Figure S4 MSX2 interacts with components of the cBAF complex.

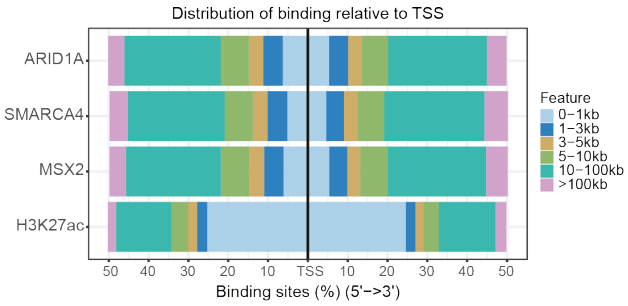
(A) Details of anti-MSX2 rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) from Fig.4A.

(*B*) Details of anti-ARID1A RIME from Fig.4D.

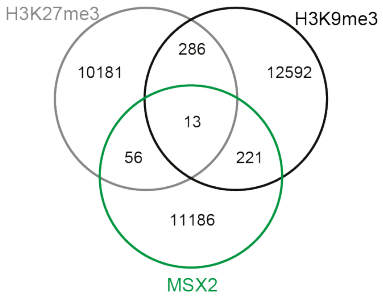
(*C*) Details of anti-SMARCA4 RIME from Fig.4E.

(*D*) SMARCA4 immunoprecipitates analyzed by Western blot probed with anti-SMARCA4 (BRG1), anti-SMARCC2 (BAF170), anti-SMARCC1 (BAF155), anti-SMARCD2 (BAF60B), anti-SMARCB1 (BAF47), and anti-MSX2.

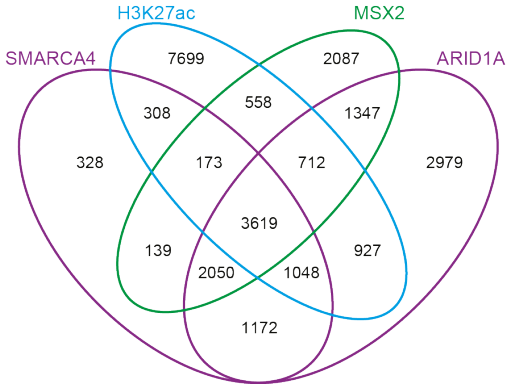
A


Distribution of binding relative to TSS
ARID1A
SMARCA4
MSX2
H3K27ac
50 40 30 20 10 TSS 10 20 30 40 50
Binding sites (%) (5'->3')
Feature
0-1kb
1-3kb
3-5kb
5-10kb
10-100kb
>100kb

B


H3K27me3
H3K9me3
MSX2
10181
286
12592
13
56
221
11186

C


H3K27ac
MSX2
SMARCA4
ARID1A
7699
2087
308
558
1347
328
173
712
2979
3619
139
927
2050
1048
1172

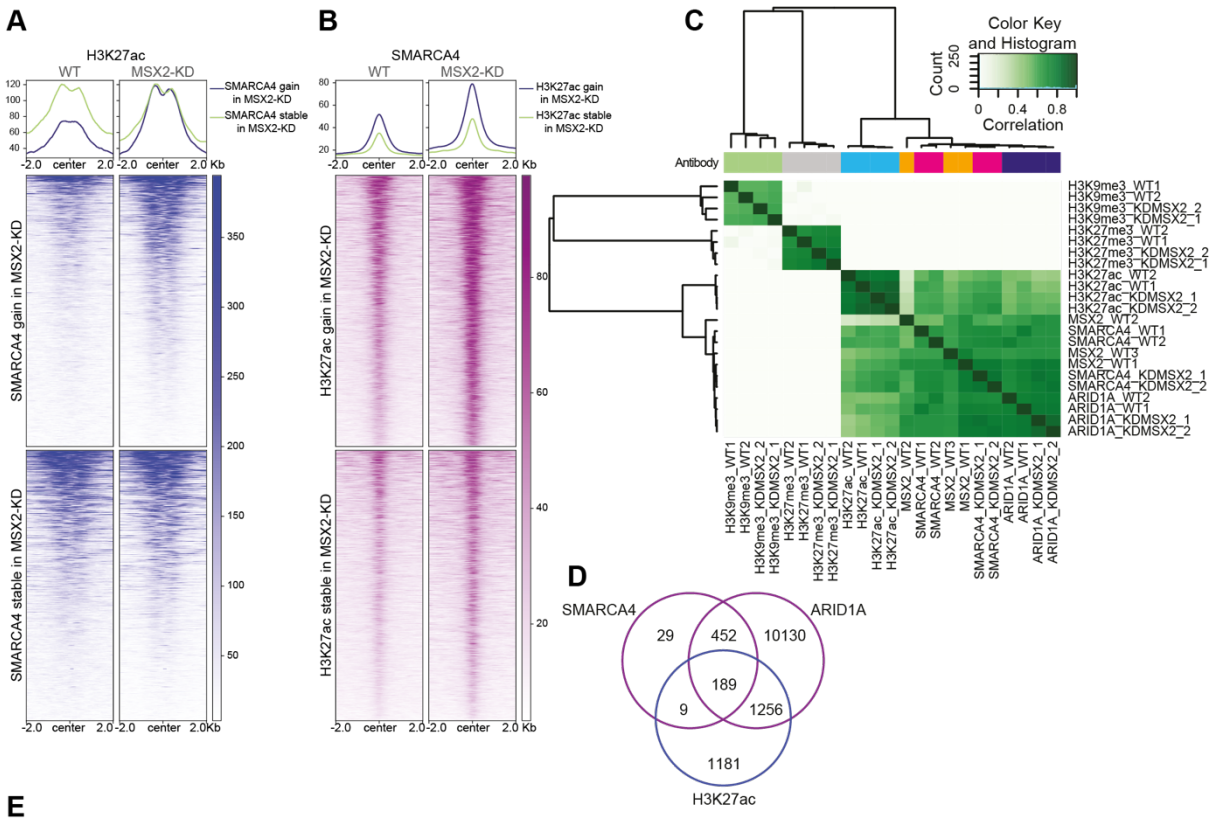
Supplemental Figure S5

MSX2 and cBAF complex co-bind trophoblast genes marked by H3K27ac.

(*A*) Distribution of MSX2, ARID1A, SMARCA4 binding sites and regions marked by H3K27ac relative to transcription start sites (TSS).

(*B*) Venn diagram depicting overlap between regions bound by MSX2 and regions marked by H3K27me3 and H3K9me3.

(*C*) Venn diagram showing overlap between regions bound by MSX2, ARID1A, SMARCA4 and marked by H3K27ac.


A
H3K27ac
WT
MSX2-KD
SMARCA4 gain in MSX2-KD
SMARCA4 stable in MSX2-KD
SMARCA4 gain in MSX2-KD
SMARCA4 stable in MSX2-KD
B
SMARCA4
WT
MSX2-KD
H3K27ac gain in MSX2-KD
H3K27ac stable in MSX2-KD
H3K27ac gain in MSX2-KD
H3K27ac stable in MSX2-KD
C
Color Key and Histogram
Count
Correlation
Antibody
H3K9me3_WT1
H3K9me3_WT2
H3K9me3_KDMSX2_2
H3K9me3_KDMSX2_1
H3K27me3_WT2
H3K27me3_WT1
H3K27me3_KDMSX2_2
H3K27me3_KDMSX2_1
H3K27ac_WT2
H3K27ac_WT1
H3K27ac_KDMSX2_1
H3K27ac_KDMSX2_2
MSX2_WT2
SMARCA4_WT1
SMARCA4_WT2
MSX2_WT3
MSX2_WT1
SMARCA4_KDMSX2_1
SMARCA4_KDMSX2_2
ARID1A_WT2
ARID1A_WT1
ARID1A_KDMSX2_1
ARID1A_KDMSX2_2
D
SMARCA4
ARID1A
H3K27ac
29
452
10130
189
9
1256
1181
E
Scale
200 kb
hg38
MSX2 WT
ARID1A WT
ARID1A MSX2-KD
SMARCA4 WT
SMARCA4 MSX2-KD
H3K27ac WT
H3K27ac MSX2-KD
LIPE-AS1
PSG3
PSG8
PSG8-AS1
PSG1
PSG6
PSG7
AC0047841
PSG11
PSG2
AC0053921
PSG5
PSG4
PSG9
AC0053922

Supplemental Figure S6

MSX2 depletion leads to increases in both cBAF occupation and H3K27ac.

(*A*) Heatmap and profile plot of H3K27ac ChIP-seq signal at MSX2 bound regions differentially enriched for SMARCA4 in MSX2-KD vs WT hTSCs and compared to an equally sized set of random MSX2 bound regions.

(*B*) Heatmap and profile plot of SMARCA4 ChIP-seq signal at MSX2 bound regions differentially enriched for H3K27ac in MSX2-KD vs WT hTSCs as determined in lower MA-plot in Fig 6A (top heatmap) and an equally sized set of random MSX2 bound regions (bottom heatmap).

(*C*) Correlation between MSX2, ARID1A, SMARCA4, H3K27ac, H3K27me3 and H3K9me3 signals in both WT and shMSX2 hTSCs based on counted reads considering all peaks detected in any of the 23 samples.

(*D*) Venn diagram depicting overlap between MSX2 binding sites, differentially enriched for ARID1A, SMARCA4 or H3K27ac in MSX2-KD vs WT.

(*E*) Genome browser tracks of MSX2, ARID1A, SMARCA4 and H3K27ac signal in WT and MSX2-KD hTSCs at the *PSG* gene cluster. MSX2: green, ARID1A and SMARCA4: purple, H3K27ac: blue. Differentially enriched H3K27ac regions are highlighted in light blue.

SI References

1. B. Egan, *et al.*, An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition. *PLoS One* **11**, e0166438 (2016).
2. S. Andrews, *FastQC: A quality control tool for high throughput sequence data*.
3. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
4. Y. Liao, G. K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47–e47 (2019).
5. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550–21 (2014).
6. R. Kolde, *pheatmap: Pretty Heatmaps*.
7. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
8. Y. Zhang, *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137-9 (2008).
9. K. Krismer, Y. Guo, D. K. Gifford, IDR2D identifies reproducible genomic interactions. *Nucleic Acids Res.* **48**, e31–e31 (2020).
10. L. J. Zhu, *et al.*, ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
11. W. J. Kent, *et al.*, The human genome browser at UCSC. *Genome Res* **12**, 996–1006 (2002).
12. E. Afgan, *et al.*, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**, W537–W544 (2018).
13. S. Durinck, *et al.*, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).

14. J. R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
15. J. Larsson, *Area-Proportional Euler and Venn Diagrams with Ellipses.* (2020).
16. C. S. Ross-Innes, *et al.*, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
17. M. Cy, *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28** (2010).
18. H. Wickham, *Elegant Graphics for Data Analysis* (Springer).
19. E. Y. Chen, *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
20. M. V. Kuleshov, *et al.*, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).
21. Z. Xie, *et al.*, Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).