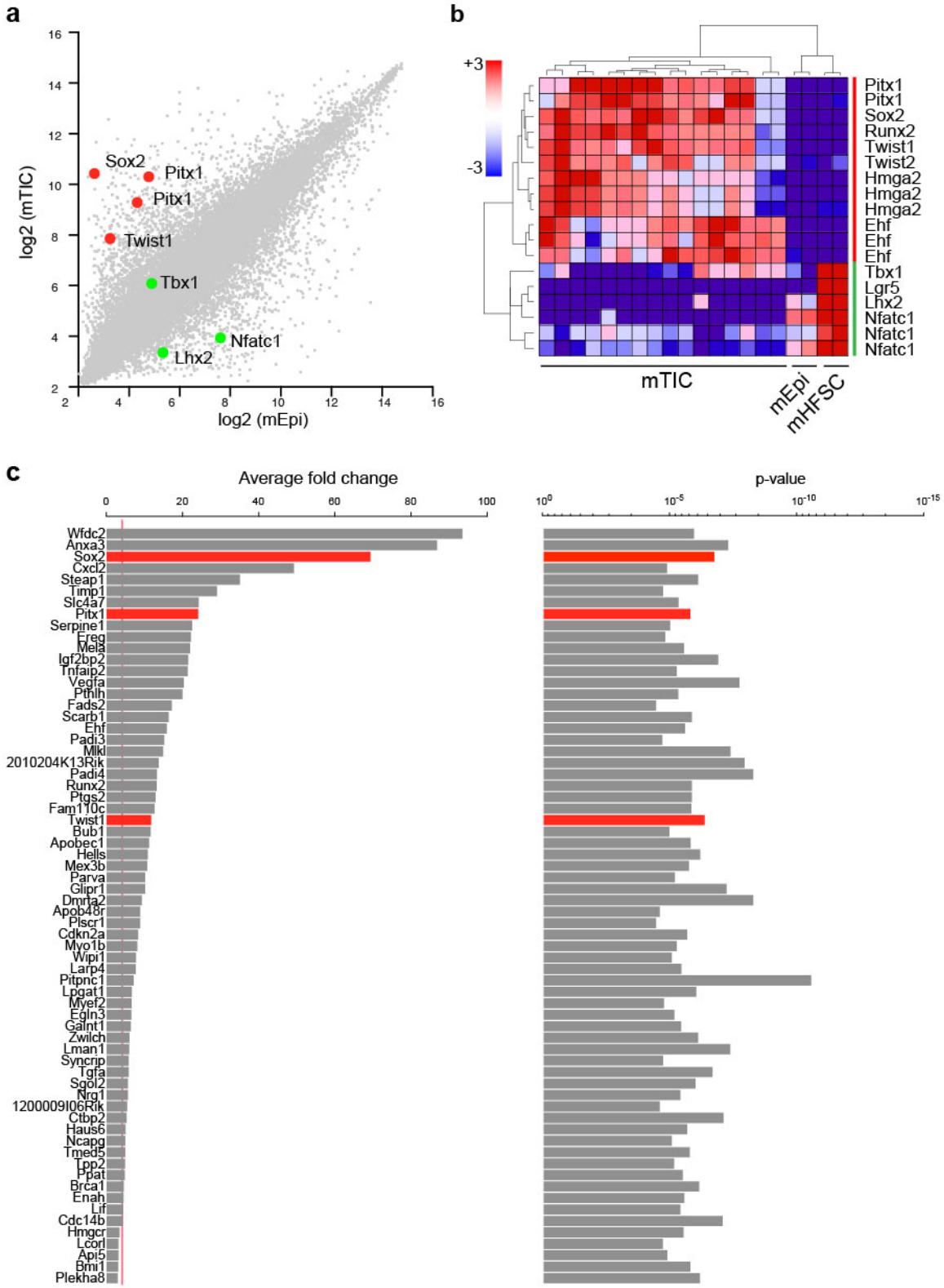
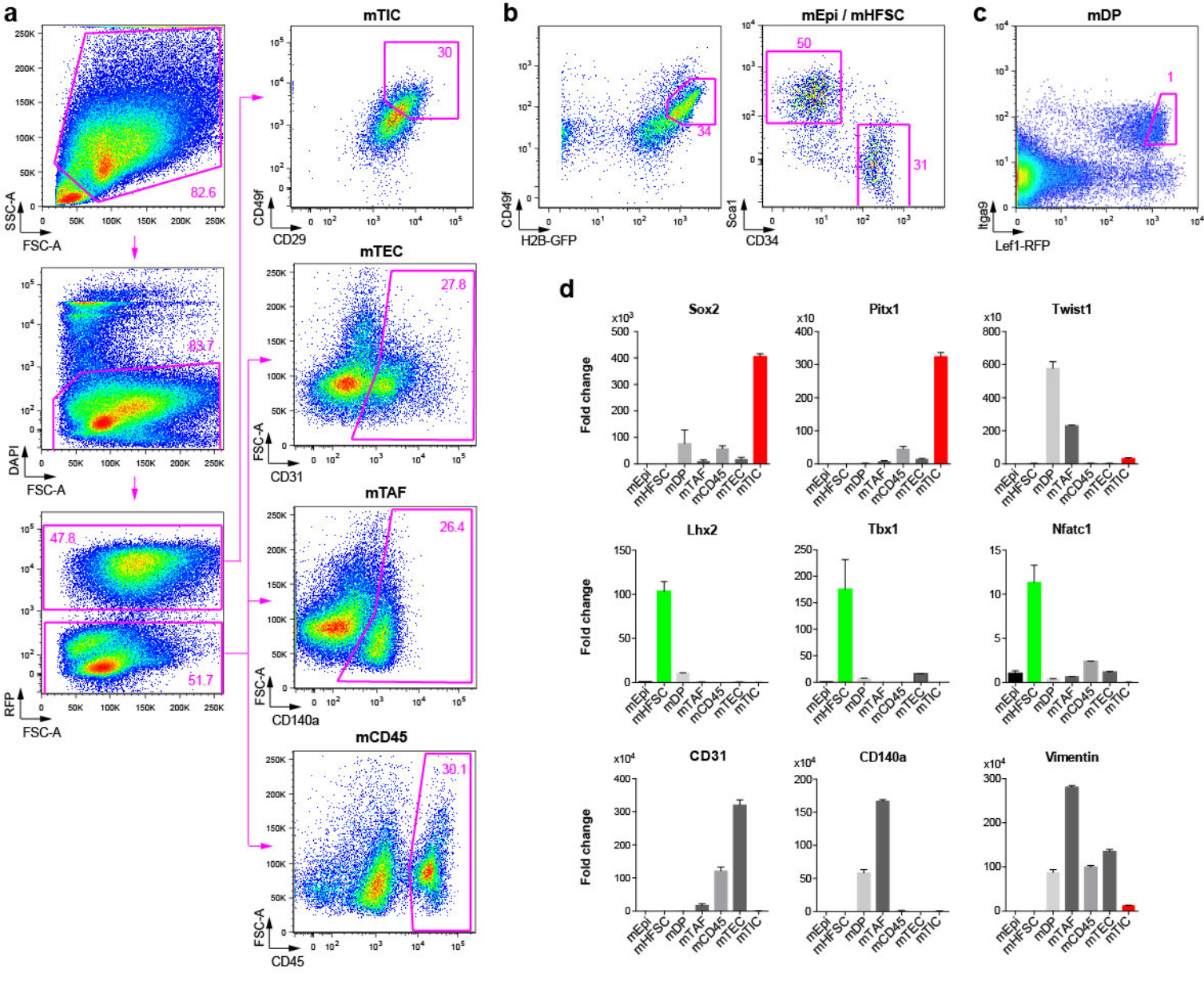


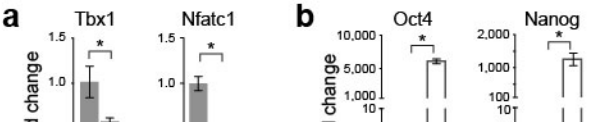
Supplementary Figure 1: Mouse tumor-initiating cells express a set of genes that are not expressed by normal skin epithelial stem and progenitor cells.

a, Scatter plot illustrates the average gene expression value of 45,101 transcripts in cancer cells with tumor-initiating potential (TIC) compared to epidermal stem and progenitor cells (mEpi). Red dots indicate transcription factors unique to mouse TICs. Green dots indicate transcription factors unique to HFSCs. **b**, Heat map visualizes cluster analysis of differentially expressed transcription factors in TICs (n=16), epidermal stem and progenitor cells (Epi, n=2), and HFSCs (n=2). **c**, Bar graph illustrating the mean expression levels and corresponding p-values for 66 candidate genes that we consistently detect in TICs of murine squamous cell carcinomas (SCCs) (16/16), while their expression is not detected in HFSCs or Epi (0/4). Red bars highlight the transcription factors Sox2, Pitx1 and Twist1.

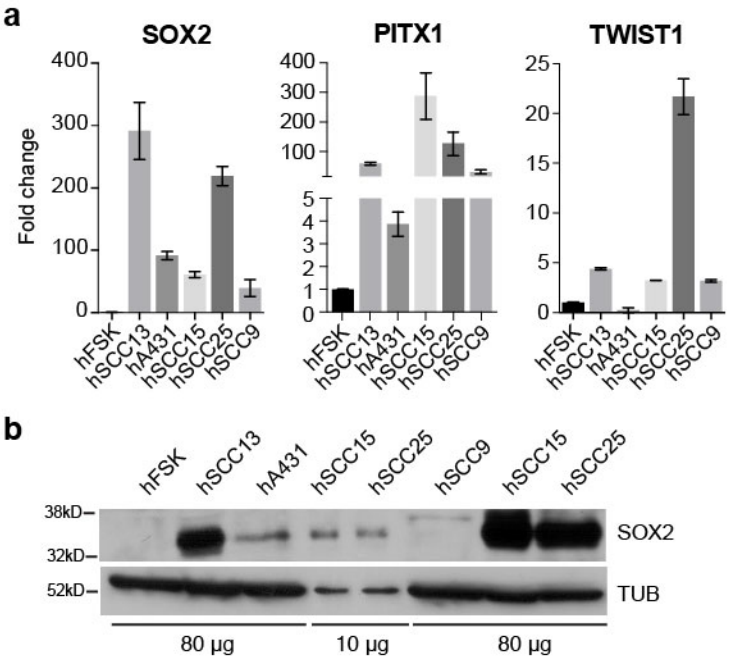


Supplementary Figure 2: Expression of TIC specific transcription factors in mouse TICs, tumor-associated mesenchyme and normal skin epithelium.

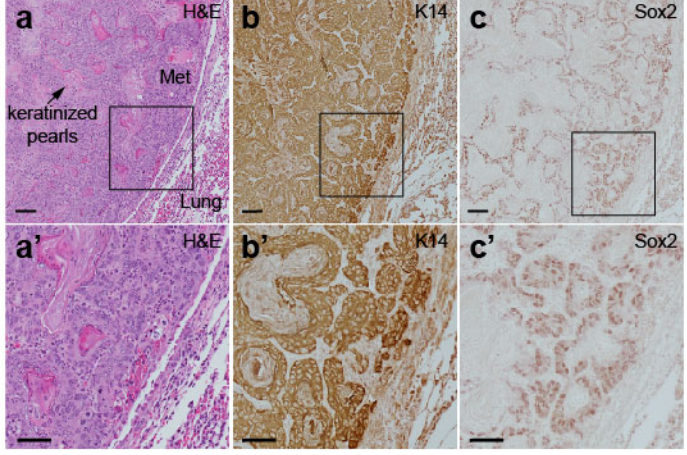
a-b, FACS-sorting strategy of murine (m) TIC, associated mesenchyme (fibroblasts, endothelial and inflammatory cells), Epi, HFSC and dermal papilla (DP) populations. **a**, TIC populations were collected from single, live (DAPI), RFP positive cells (tumor parenchyma labeled with retrovirus expressing RFP) with high α6- and β1-integrin expression (CD49f and CD29). Tumor-associated endothelial cells (TECs) were isolated from single, live (DAPI), RFP negative and CD31 positive cell populations. Tumor-associated fibroblasts (TAFs) were sorted from single, live (DAPI), RFP negative and CD140a positive cell populations. Inflammatory cells were sorted from single, live (DAPI), RFP negative and CD45 positive cell populations. **b**, mEpi and mHFSCs were sorted from single, live (DAPI), lineage positive (K14-H2BGFP), CD49f-positive cells as the Sca1- and CD34-positive cell populations, respectively. **c**, mDP cells were isolated from single, live (DAPI), lineage positive (Lef1-RFP) and Itga9-positive cell populations. **d**, qRT-PCR analyses of the transcription factors Sox2, Pitx1 and Twist1, as well as standard markers for murine HFSCs (Lhx2, Tbx1, Nfatc1), endothelial cells (CD31) and fibroblasts/DP (CD140a, Vimentin) on FACS purified mEpi, mHFSC, mDP, mTAF, mCD45, mTEC, and mTIC populations. Bar graphs show mean with error bars indicating ± s.d (n=3).



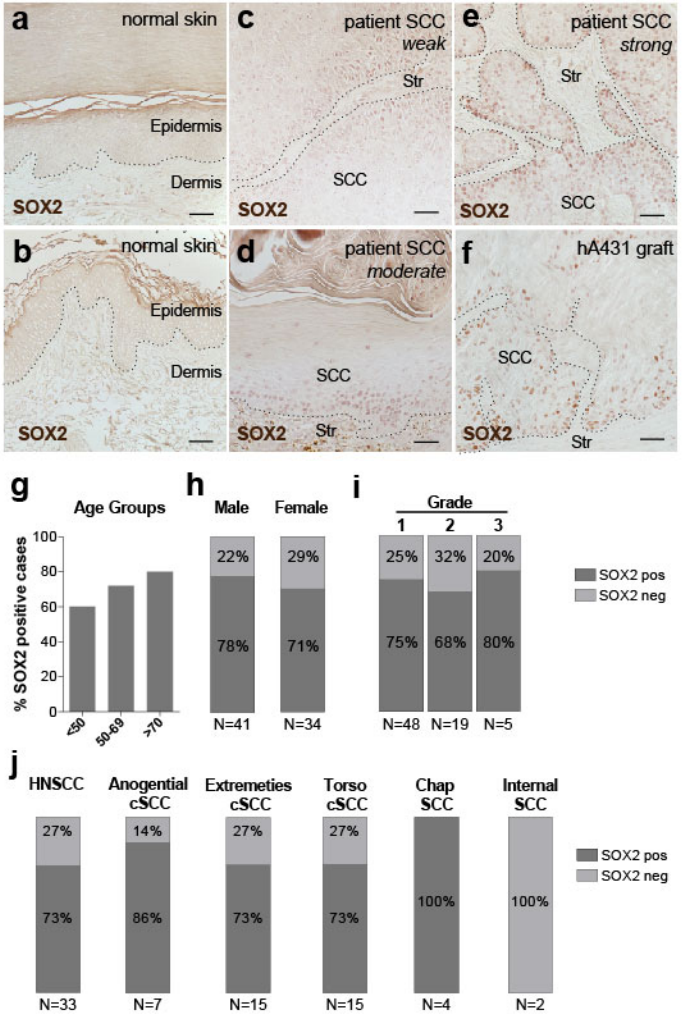
Supplementary Figure 3: Expression of HFSC and ESC markers in TIC cultures.
a, qRT-PCR analysis of Tbx1 and Nfatc1 on mTIC and mHFSC cultures. **b**, qRT-PCR analysis of embryonic stem cell (ESC) markers Oct4 and Nanog on mTIC, HFSC, and ESC cultures. Bar graphs show mean with error bars indicating ± s.d. (n=3, *P<0.05 Student's t-test).



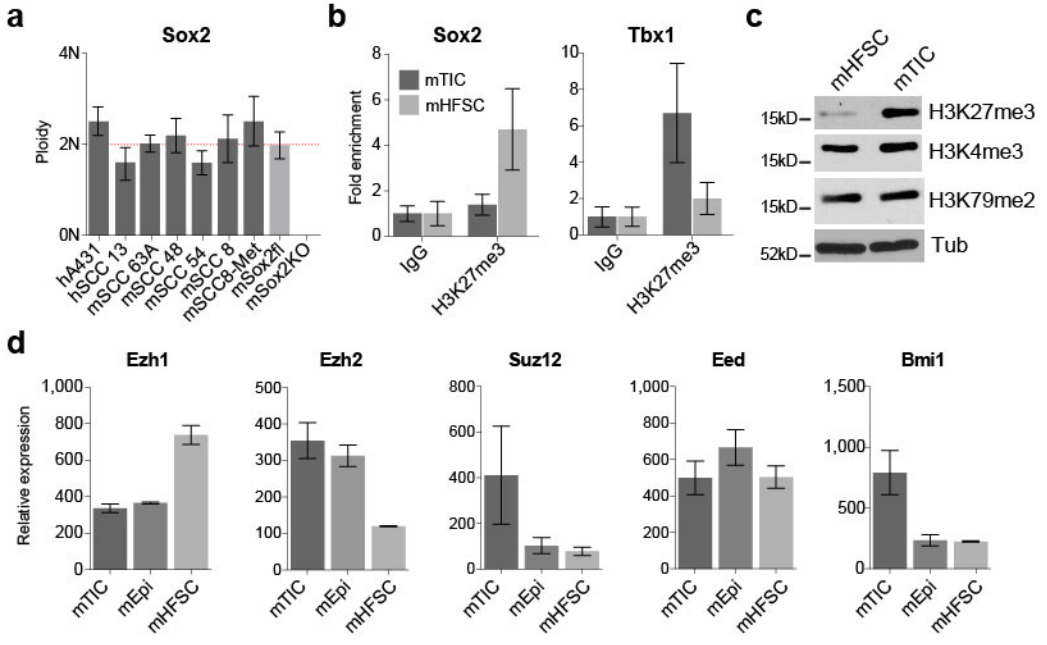
Supplementary Figure 4: SOX2 is expressed in human SCC cell lines from different origins.
a, qRT-PCR analyses for TIC-specific transcription factors on human (h) SCC cell lines compared to primary human foreskin keratinocyte (hFSK) cultures. Bar graphs show mean with error bars indicating ± s.d. (n=3). **b**, Western blot analysis of SOX2 on total protein lysates from human foreskin keratinocytes (hFSK) and human SCC cell lines. Tubulin serves as a loading control. Total protein amounts loaded per lane are indicated.



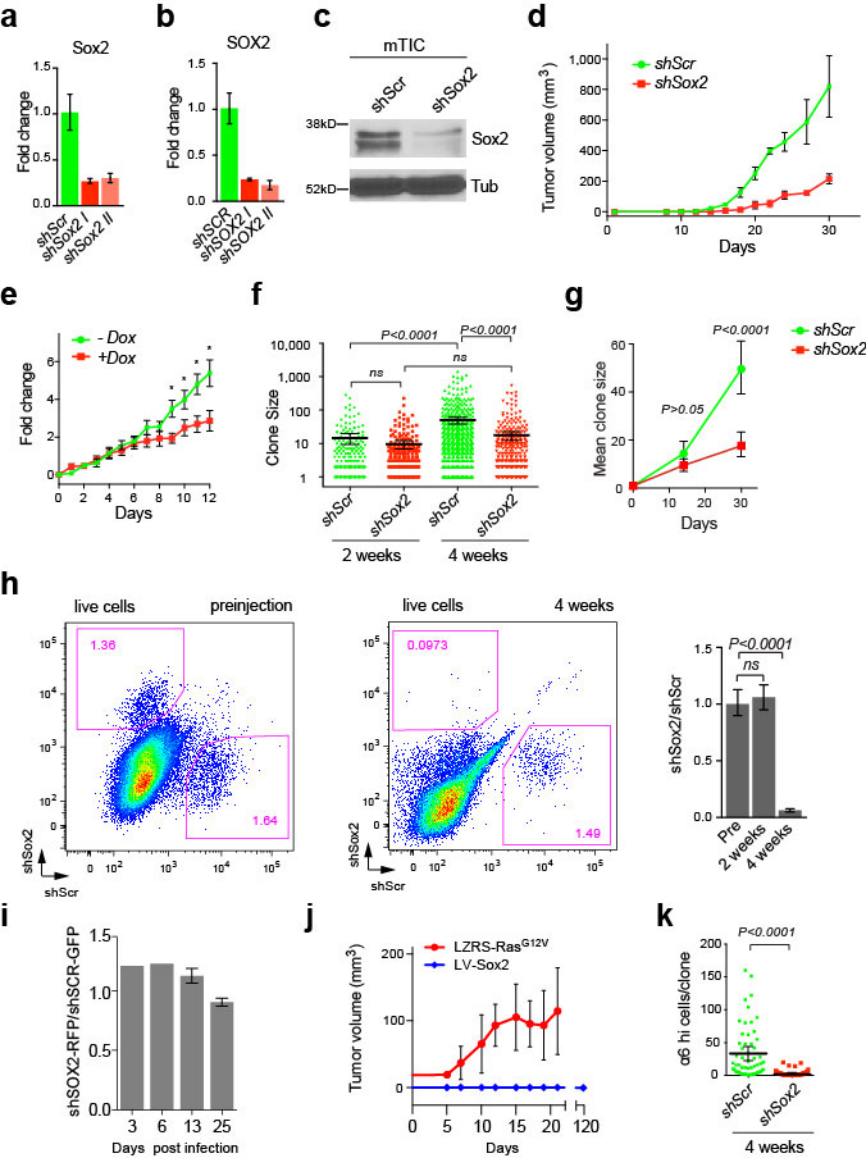
Supplementary Figure 5: Sox2 is expressed in spontaneous lung metastasis of mouse cutaneous SCCs. **a-c,** Representative images of spontaneous lung metastasis that developed in 5% of DMBA-treated mice bearing primary cutaneous SCCs. **a, a',** Hematoxylin and eosin (H&E) staining shows an example of lung metastasis (Met) from cutaneous SCC. Note characteristic keratin pearls. **b, b',** Immunohistochemistry for keratin 14 (K14), a marker for cutaneous SCCs. **c, c',** Immunohistochemistry for Sox2 on spontaneous lung Met. Scale bars are 50 μm.



Supplementary Figure 6: Nuclear SOX2 is expressed in the majority of primary human patient SCCs. **(a-f)** Immunohistochemistry of SOX2 on normal skin **(a,b)**, primary patient SCC samples on tissue microarrays **(c-e)** and A431 xenografts **(f)**. Representative image of primary human patient SCC samples with weak (**c**), moderate (**d**) and strong (**e**) nuclear SOX2 staining. Note that SOX2 localizes predominantly to the tumor-stroma interface. **(f)** Xenograft of human SCC cell line A431 served as a positive SOX2 staining control. **(a-f)** Scale bar measures 50 µm. Str, stroma. **(g-j)** Percentage of human patient SCC samples positive for nuclear SOX2 staining categorized by various patient parameters. **(g)** SOX2 positive SCC cases increase with age. Out of all malignant SCC cases, 60% (9/15) under the age of 50, 74% (26/35) between ages 50-69, and 80% (20/25) over the age of 69 were positive for nuclear SOX2 staining. **(h)** Percentage of SOX2 positive cases are similar between males and females. 70% (24/34) of malignant SCCs from female patients and 78% (32/41) from male patients were positive for nuclear SOX2 staining. **(i)** Nuclear SOX2 staining does not correlate with tumor grade. Out of all malignant SCC cases, 75% (36/48) of Grade 1, 68% (13/19) of Grade 2, and 80% (4/5) of Grade 3 were positive for nuclear SOX2 staining. **(j)** Positive nuclear SOX2 staining by anatomical location of human patient SCC samples. 73% (24/33) of head and neck SCCs, and 86% (6/7) of anogenital SCCs were positive for nuclear SOX2. 73% of cutaneous SCCs on extremities (11/15) and torso (11/15) showed positive nuclear SOX2 staining. 0% (0/2) of internal SCCs and 100% (4/4) of SCCs on chap had positive nuclear staining for SOX2.

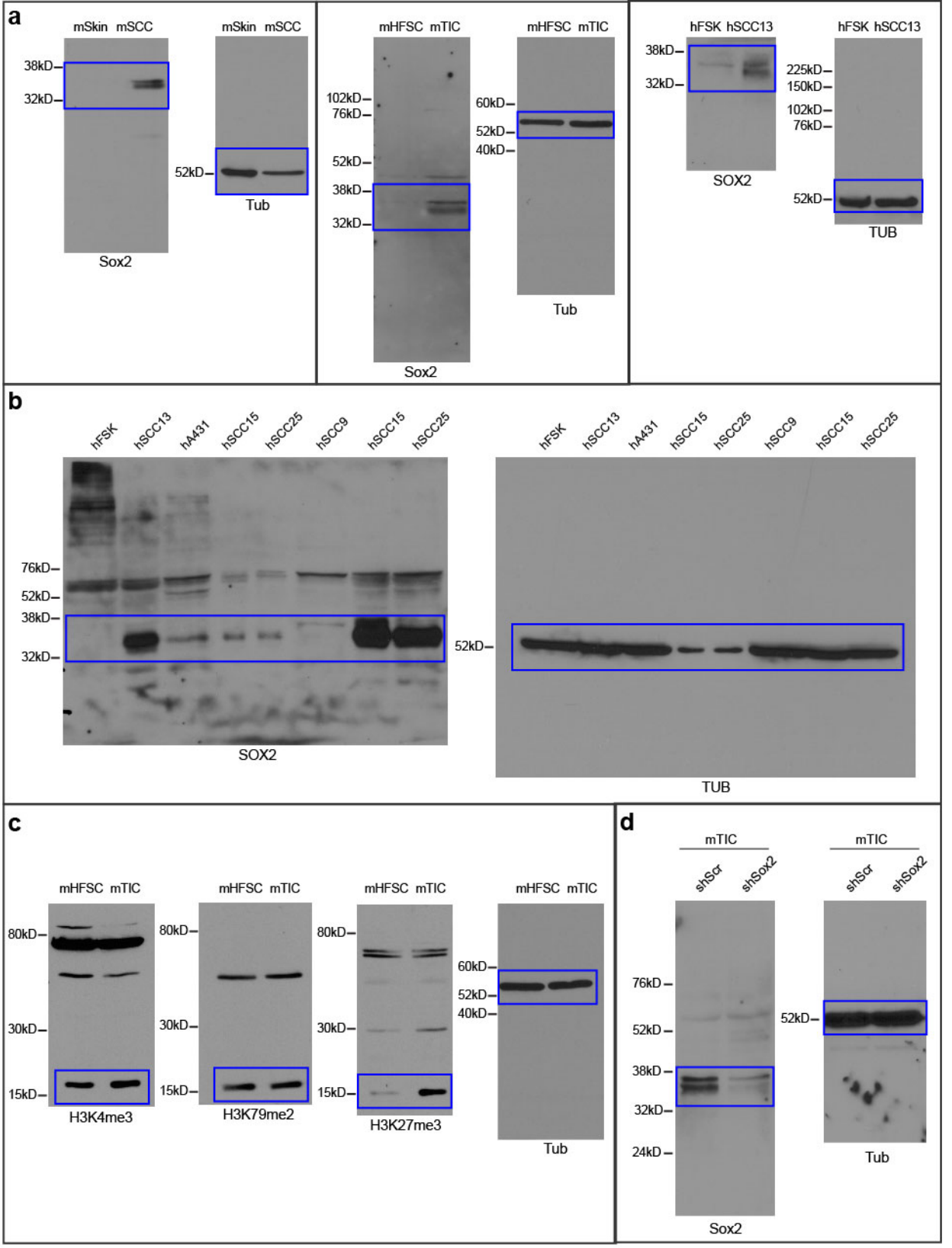


Supplementary Figure 7: Sox2 de novo expression is mediated via epigenetic remodeling.
(a) TaqMan copy number analysis of SOX2 on human (h) and mouse (m) SCC lines. Normal Sox2-floxed and Sox2-knock out keratinocytes were used as reference. **(b)** qRT-PCR analyses of chromatin samples from cultured mTICs and HFSCs after immunoprecipitation with anti-H3K27me3 and IgG control antibodies. Bar graphs showing average Fold enrichment of Sox2 and Tbx1 promoter sequence with error bars indicating + s.d. (n=3, *P<0.05 Student's t-test). **(c)** Western blot analyses of H3K27me3, H3K4me3, H3K79me2 expression on total protein extracts from mHFSC and mTIC cultures. β-tubulin served as a loading control. **(d)** Relative expression levels of Polycomb Repressive Complex components in freshly isolated mouse tumor initiating cells (TIC), skin epithelial cells (Epi, α6-integrin high/CD34 low) and hair follicle stem cells (HFSC, α6-integrin high/CD34 high). Bar graphs show mean Affymetrix expression values with error bars indicating + s.d (n=16 mTIC; n=2 mEpi; n=2 mHFSC).


a
Sox2
Fold change
shScr
shSox2 I
shSox2 II
b
SOX2
Fold change
shSCR
shSOX2 I
shSOX2 II
c
mTIC
shScr
shSox2
38kD
Sox2
52kD
Tub
d
shScr
shSox2
Tumor volume (mm3)
Days
e
- Dox
+Dox
Fold change
Days
f
P<0.0001
P<0.0001
ns
ns
Clone Size
shScr
shSox2
shScr
shSox2
2 weeks
4 weeks
g
P<0.0001
shScr
shSox2
P>0.05
Mean clone size
Days
h
live cells
preinjection
1.36
1.64
shSox2
shScr
live cells
4 weeks
0.0973
1.49
shSox2
shScr
P<0.0001
ns
shSox2/shScr
Pre
2 weeks
4 weeks
i
shSOX2-RFP/shSCR-GFP
3
6
13
25
Days post infection
j
LZRS-RasG12V
LV-Sox2
Tumor volume (mm3)
Days
k
P<0.0001
α6 hi cells/clone
shScr
shSox2
4 weeks

Supplementary Figure 8: Sox2 downregulation reduces cutaneous SCC growth and clonal expansion

(a) qRT-PCR analysis of Sox2 on mouse tumor initiating cells (mTICs) infected with lentivirus expressing short hairpin RNA (shRNA) against Sox2 along with nuclear red fluorescent protein (H2B-RFP) or scrambled control shRNA along with nuclear green fluorescent protein (H2B-GFP). **(b)** qRT-PCR analysis of Sox2 on human SCC cells infected with shSCR;H2B-GFP or two independent shSOX2;H2B-RFP expressing lenti viruses. **(a-b)** Bar graphs show average fold change with error bars indicating ± s.d. **(c)** Western blot analysis of Sox2 protein on shScr and shSox2 transduced mouse mTICs. β-Tubulin (Tub) served as loading control. **(d)** Tumor growth curve of mouse TICs infected with lentivirus expressing shSox2;H2B-RFP or scrambled shScr;H2BGFP control followed by transplantation onto Nude recipient mice. Measurements are mean with error bars indicating ± s.e.m. (n=4). **(e)** Doxycycline-inducible knockdown of Sox2 in established mouse SCC grafts. Data are represented as mean with error bars indicating ± s.e.m. (n=10, *P<0.05, Student's t-test). **(f-g)** Analysis of clonal growth competition assay at 2 and 4 weeks after intradermal transplantation when 1-2% of mTICs have been transduced with shSox2;H2B-RFP and shSCR;H2B-GFP (n=6). **(f)** Scatter plots illustrate clone size distributions. Horizontal lines represent mean with error bars indicating ± 95% CI. **(g)** Line graphs showing average clone size as a function of time (± s.e.m). **(f-g)** P values were determined by Mann-Whitney non-parametric t-tests. **(h)** Representative scatter plot of flow cytometry analysis of clonal competition at different time points (pre-injection and 4 weeks post-transplantation). Bar graphs show mean ratio of shSox2;H2B-RFP to shScr;H2B-GFP cells per tumor as measured by flow cytometry. **(i)** Flow cytometry analysis of clonal competition at different time points in human A431 SCC cultures. Bar graphs show mean ratio with error bars indicating ± s.e.m. (n=6). **(j)** Tumor growth curves of hair follicle stem cells (HFSCs) after infection with LZRS-RasG12V or lentivirus expressing Sox2 followed by intradermal injection into Nude mice. **(k)** Scatter plot showing the number of α6-integrin expressing SCC cells per shScr-GFP and shSox2-RFP clone 4 weeks after transplantation. Horizontal bar shows mean with error bars indicating ± 95% CI. Statistical significance was determined by Mann-Whitney non-parametric t-test.



Supplementary Figure 9: Expanded views of western blot films.
a, Figure 1e-g blots.
b, Supplementary Figure 4b blots.
c, Supplementary Figure 7c blots.
d, Supplementary Figure 8c blots.

Supplementary Table 1: Full sequences of shRNAs

Name	Sequence
shScramble SHC002	CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTT
Sox2 TRCN0000424718	CCGGAGGAGCACCCGGATTATAAATCTCGAGATTTATAATCCGGGTGCTCCTTTTTTG
Sox2 TRCN0000085748	CCGGCGAGATAAACATGGCAATCAACTCGAGTTGATTGCCATGTTTATCTCGTTTTTG
SOX2 TRCN0000231643	CCGGGTACAGTATTTATCGAGATAACTCGAGTTATCTCGATAAATACTGTACTTTTTG
SOX2 TRCN0000257314	CCGGTGGACAGTTACGCGCACATGACTCGAGTCATGTGCGCGTAACTGTCCATTTTTG
NRP1 TRCN0000322980	CCGGTATACTAGAATCACCGCATTTCTCGAGAAATGCGGTGATTCTAGTATATTTTTG

Supplementary Table 2: Antibodies with dilutions

Name	Catalogue Number & Manufacturer	Dilution (Application)
rabbit monoclonal anti-Sox2	ab92494, Abcam	1:1000 (IF & WB) or 1:100 (IHC)
rat anti-mouse CD104/β4	553745, BD Pharmingen	1:1000 (IF)
rat anti-human/mouse CD49f/α6-PerCP-Cy5.5	313617, BioLegend	1:200 (IF & Flow)
rabbit monoclonal anti-Survivin	2808, Cell Signaling	1:1000 (IF)
rabbit anti-phospho Ki67	KI67P-CE, Novo Castra-Leica	1:1000 (IF)
rabbit anti-Casp-3	AF835, R&D Systems	1:1000 (IF)
mouse anti-NuMa	610561, BD Biosciences	1:500 (IF)
anti-CD31	13031982, eBioscience	1:200 (IF)
rabbit anti-K14	PRB-155P, Covance	1:1000 (IHC)
rat monoclonal anti-Nidogen	Sc-33706, Santa Cruz	1:1000 (IF)
anti-phospho-Histone H3 (Ser10)	06-570, Millipore	1:1000 (IF)
rabbit anti-Sox2	2748, Cell Signaling	1:500 (WB)
mouse monoclonal anti-β-tubulin	T5201, Sigma	1:10,000 (WB)
rabbit polyclonal anti-H3K4me3	ab8580, Abcam	1:1000 (WB)
rabbit polyclonal anti-H3K27me3	07-449, Millipore	1:1000 (WB)

rabbit polyclonal anti-H3K79me2	ab3594, Abcam	1:1000 (WB)
hamster anti-rat/mouse CD29/β1-Alexa 700	102218, BioLegend	1:200 (Flow)
mouse anti-human CD29/β1-Alexa 700	303020, BioLegend	1:200 (Flow)
anti-mouse CD45-Biotin	553078, BD Pharmingen	1:200 (Flow)
anti-human CD45-Biotin	13-0459-82, BioLegend	1:200 (Flow)
anti-mouse CD31-Biotin	13-0311-85, eBioscience	1:200 (Flow)
anti-human CD31-Biotin	13-0319-82, eBioscience	1:200 (Flow)
anti-CD11b-Biotin	BAM 1124, R&D Systems	1:200 (Flow)
FITC-Streptavidin	11-4317-87, eBioscience	1:200 (Flow)
rat anti-mouse CD140a-APC	171401-81, eBioscience	1:200 (Flow)
rat anti-mouse CD31-PerCP-eFluor 710	46-0311-80, eBioscience	1:200 (Flow)
APC-Streptavidin	17-4317-82, eBioscience	1:200 (Flow)
Alexa 488 goat anti-rabbit IgG (H+L)	A11034, Invitrogen	1:1000 (IF)
FITC goat anti-rabbit IgG (H+L)	111-095-144, Jackson	1:1000 (IF)
Alexa 568 donkey anti-rabbit IgG (H+L)	A10042, Invitrogen	1:1000 (IF)

DyLight 649 goat anti-rat IgG	405411, BioLegend	1:1000 (IF)
DyLight 405 donkey anti-rabbit IgG (H+L)	711-475-152, Jackson	1:1000 (IF)
biotinylated horse anti-rabbit IgG (H+L)	BA-1100, Vector Laboratories	1:200 (IHC)
HRP donkey anti-rabbit IgG (H+L)	711-035-152, Jackson	1:3000 (WB)
HRP donkey anti-mouse IgG (H+L)	715-035-151, Jackson	1:2000 (WB)

Where IF = immunofluorescence; WB = western blotting; IHC = immunohistochemistry; Flow = flow cytometry analysis or sorting

Supplementary Table 3: Sequences of qRT-PCR primers

Name	Sequence
Sox2 (a)	F= 5'GGGGCAGCGGCGTAAGA R= 5'AGCCTCCGGGAAGCGTGTA
SOX2	F= 5'CGGCGGCAACCAGAAAAACA R= 5'TGCCCGCGGGACCACAC
Rplp0	F= 5'GTGCCATCGCCCCGTGTG R= 5'TGGATGATCAGCCCGAAGGAGA
RPLP0	F= 5'GGGGGAATGTGGGCTTTGTGTT R= 5'GGTGCCCCTGGAGATTTTAGTGGT
Vegfa	F= 5'GAAGCTACTGCCGTCCGATTGAGA R= 5'GTGCTGGCTTTGGTGAGGTTTGAT
Nrp1	F= 5'CCACCCCGGCTCGTATGTCAC R= 5'TTCCGAAGGGGGCGATTTAGG
Nrp2	F=5'CTTTGCAGAAGACACCACCA R=5'CAGTTCTGGTGGGAGGGATA
Pitx1	F= 5'CCGGGATGCCAGGAAGAGC R= 5'CACGGGCAGGCGGACAGT
PITX1	F= 5'GGCCGGGGGAGGACGAC R= 5'CAACGCCCGCCCCAACA
Twist1 (a)	F= 5'GGCAAGCGCGGCAAGAAAT R= 5'GAGGGCAGCGTGGGGATGAT
TWIST1	F= 5'GGCAAGCGCGGCAAGAAGT R= 5'GAGGGCAGCGTGGGGATGAT
Ereg	F= 5'CTTCGTCCTTTGTTTGCCTTTGTG R= 5'CCTCCCTGACCTGGTATGTGTTCC
Tgf α	F= 5'CCGGTTTTTGGTGCAGGAAGAG R= 5'GCGAACACCCACGTACCCAGAGT
EgfR	F= 5'CTACCGCCTCCCAGACAGACGAC R= 5'GCGCCGGCAACGACGAG
Spp1	F= 5'CCAGCAGCTCACACTGAAGA R= 5'CCAAACAGGCAAAAGCAAAT
Pitpnc1	F= 5'ATGCTGCTCAAGGAGTACCG R= 5'TAGGCAGCTTGCTGTTGAGA
Igf2bp2	F= 5'CTACTCAAGTCCGGCTACGC R= 5'TGTCCCATATTCAGCCAACA
Sox2 (b)	F= 5'TACTGGCAAGACCGTTTTCGTC R= 5'AAGCAGTTGGTGGTGCAGGATG
Twist1 (b)	F= 5'AGGCCGGAGACCTAGATGTCATT R= 5'CAGCGATGCCTTTCCTGTCA
CD31	F= 5'GTCATGGCCATGGTCGAGTA R= 5'CTCCTCGGCGATCTTGCTGAA
CD140a	F= 5'ACGCATGCGGGTGGACTC R= 5'GATACCCGGAGCGTGTCAGTTAC

Vimentin	F= 5'GTGATGTGCGCCAGCAGTATG R= 5'TGCAGGCGGCCAATAGTGT
Lhx2 [6]	F= 5'CCTACTACAACGGCGTGGGCACTGT R= 5'GTCACGATCCAGGTGTTCAGCATCG
Enpp2	F= 5'ACGGCTAGTCTTCCGGTAGAAATC R= 5'CGCCCTGATGTCCGTGTATCT
Akp2	F= 5'GTGGCGGCGGAAATACA R= 5'ATGTCCCCGGGCTCAAAGA
Wif1	F= CTGCCGAAATGGAGGTAAATGC R= AATTCAGGCCGGCGTTCTAAAG
Nfact1	F= ACATAGCCTCCTGCTGGAAA R= AAGAGGGGTCTGGAGCAAAT
Tbx1	F= CACAGATATCAGCCCCGATT R= GCGTGTCTCCTCAAACACAA

Supplementary Table 4: Sequences of ChIP-PCR primers

Sox2 Binding at	Sequence
Nrp1	F= 5'CCGACTGCCATACAGAAAGC R= 5'TGGCTGTGACACTGTTGTGA
Spp1	F= 5'GGAGCATTGCAAAATGTGAA R= 5'GTAAGCAGTGCCAGGTGTGA
Pitpnc1	F= 5'AGGAGGCAATTAGCAGGACTC R= 5'CCAGCAGGGTAGTCTCTTGC
Igf2bp2	F= 5'CAGTGGTGTAGCCTCTGCAA R= 5'TAACCCATTCATCCCCACTG
Pitx1	F= 5'TCGGGGTTGTTTTGTTTTGT R= 5'GACACCTCAGCCCAGAGC
NRP1	F= 5'TCAGGAGTGAGTAAACAGGCAAT R= 5'GGAAACCTGGGTACCTGGAG

H3K27me3 Binding at	Sequence
Sox2	F= 5' CCCATTTATTCCCTGACAGC R= 5' TTGCAAACACTCTCTTCTCTGC
Tbx1	F= 5' GTCTCCTTTCTCTCGCTCCA R= 5' CGGAAGGGAAGACATGAAAA