Supporting Information

Cattoglio et al. 10.1073/pnas.1505569112

SI Materials and Methods

Pull-Down Assays. pFLAG-CMV-5a and p3XFLAG-CMV-10 constructs expressing OCT4, SOX2, RAD23B, XPC, or XPC truncations were transfected by calcium phosphate into 293T cells at ∼70% confluency in 10-cm dishes. Cells were scraped from plates in ice-cold PBS 48 h after transfection, pelleted, and flash-frozen in liquid nitrogen. Cell pellets were thawed on ice, resuspended to 1 mL/10-cm plate of co-IP lysis buffer (0.2 M NaCl, 25 mM Hepes, 1 mM $MgCl_2$, 0.2 mM EDTA, 0.5% Nonidet P-40, and protease inhibitors) and passed through a 25-G needle five times. Whenever benzonase treatment was required, cell pellets were resuspended to a low-salt co-IP buffer (0.1 M NaCl, 25 mM Hepes, 1 mM $MgCl_2$, 0.2 mM EDTA, 0.5% Nonidet P-40, and protease inhibitors), with or without 125 U/mL of benzonase (Novagen), passed through a 25-G needle, rocked at 4 °C for 1 h, and added to NaCl to a final concentration of 0.2 M. Lysates were then rocked at 4 °C for 30 min and centrifuged at 16,100 × *g* at 4 °C. Supernatants were quantified by Bradford and 250–500 μg of total protein precleared with protein A/G-Sepharose beads for 2 h at 4 °C (GE Healthcare Life Sciences) before overnight immunoprecipitation with 1–2 μg of antibody. BSA-precleared protein A/G Sepharose beads were then added to the samples and incubated at 4 °C for 2 h. After extensive washes, beads were boiled in 2× SDS-loading buffer and analyzed by SDS/PAGE and Western blot. When anti-HA agarose was used for IP instead of unconjugated antibodies, the resin was precleared with BSA for at least 2 h at 4 °C before overnight incubation with cell lysates.

For co-IP experiments with the human SCC complex and SOX2 (Fig. 6*E*), pCMV-5a-hSOX2-HA or control pLKO.1-FLAG-HA-RFP expression plasmids were transfected by calcium phosphate into 293T cells. Then 7.5×10^6 cells were seeded on 15-cm dishes 24 h before transfection and scraped 48 h posttransfection in co-IP buffer [0.2 M KCl, 50 mM Hepes pH 7.6, 0.1% Nonidet P-40, 10% (vol/vol) glycerol, 0.1 mM EDTA, 1 mM $MgCl_2$, 1 mM DTT, 1 mM TCEP, and protease inhibitors]. Lysates were passed through a 25-G needle three times, cleared, and incubated overnight at 4 °C with ∼50 μg of human recombinant His6-XPC:FLAG-RAD23B:CETN2 purified from Sf9 cells, as previously described (1). Anti-HA resin was then added and incubated for an additional 4 h, followed by three washes in co-IP buffer. Resin with immobilized proteins was then evenly divided into separate tubes with either 0.1 mg/mL ethidium bromide (Calbiochem-Millipore), 250 U/mL benzonase (Novagen), or 10 μg/mL RNaseA (Fermentas, Thermo Scientific) in co-IP buffer and incubated at 37 °C for 1 h. After three more washes, proteins were eluted with 0.2 mg/mL HA peptide (Sigma-Aldrich) and analyzed by SDS/PAGE and Western blot.

For co-IP experiments with human purified proteins ± RNA or heparin (Fig. 6*F*), HA-tagged SOX2 was expressed in 293T cells by transient transfection. Cells were lysed in 1 mL/10 cm plate of high-salt lysis buffer (HSLB) 48 h posttransfection (0.5 M NaCl, 50 mM Hepes pH 7.6, 1% Nonidet P-40, 10% glycerol, 1 mM DTT, 1 mM TCEP, and protease inhibitors), cleared by centrifugation, and precleared overnight on protein A resin. For each IP, 400 μL of precleared SOX2 lysate were incubated for ∼2 h with 10 μL of HA-resin (Sigma-Aldrich) preblocked overnight at 4 °C with 1% BSA in HSLB. Resin was then washed four times with HSLB and twice with co-IP buffer (0.2 M NaCl, 25 mM Hepes pH 7.6, 0.1% Nonidet P-40, 10% glycerol, 0.1 mM EDTA, 1 mM MgCl2, 1 mM DTT, and 1 mM TCEP). Next, ∼2 μg of purified recombinant SCC was added to the resin with 150 μL of co-IP buffer and either RNA buffer, 2 μg of TriZOL-extracted 293T or *E. coli* RNA, or increasing amounts of heparin (10-fold concentration range from 0.5 to 500 μg), and incubated overnight at 4 °C. Resin was washed extensively with 200 μL co-IP buffer, and proteins were eluted into 10 μL of co-IP buffer containing 0.3 mg/mL HA peptide (Sigma-Aldrich) and analyzed by SDS/PAGE and Western blot.

1. Fong YW, et al. (2011) A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell* 147(1):120–131.

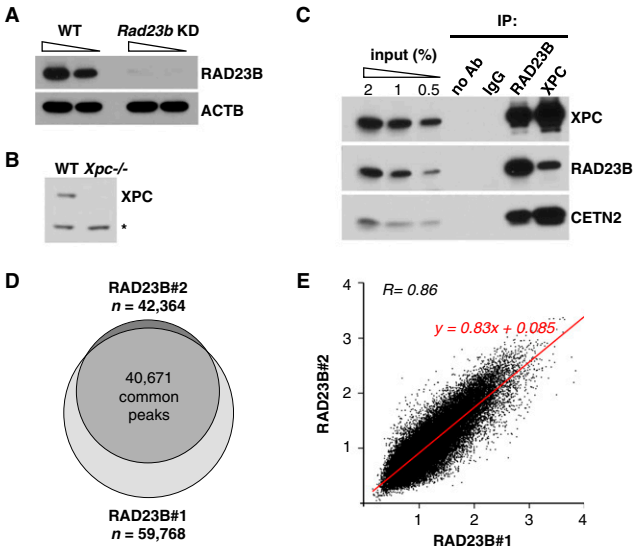

Fig. S1. Specificity and reproducibility of RAD23B ChIP-seq. (*A*) Immunoblotting of whole-cell extracts from WT and *Rad23b* knockdown (KD) D3 mESCs with the RAD23B antibody used in *C* and for the ChIP-seq experiments in *D* and *E*, and β-actin as a loading control. (*B*) Immunoblotting of whole-cell extracts from WT and *Xpc* knockout ($^{-/-}$) mESCs with the XPC antibody used in *C*; a nonspecific band (*) works as a loading control. (*C*) Western blot of D3 mESCs whole-cell extracts immunoprecipitated (IP) with RAD23B and XPC antibodies shows efficient pull-down of the whole SCC complex. Pull-down with no antibody and normal IgG control for antibody specificity. (*D*) Venn diagram showing the extensive peak overlap between two independent RAD23B ChIP-seq experiments (#1 and #2). *n* is the total number of peaks identified in each replicate. (*E*) Scatter plot comparing the enrichment over background ($\log_2$[scaled fold-change]) of the 40,671 peaks in common between the two independent RAD23B ChIP-seq experiments (#1 and #2). The correlation between the two replicates is linear and high, as measured by Pearson's correlation coefficient (*R*).

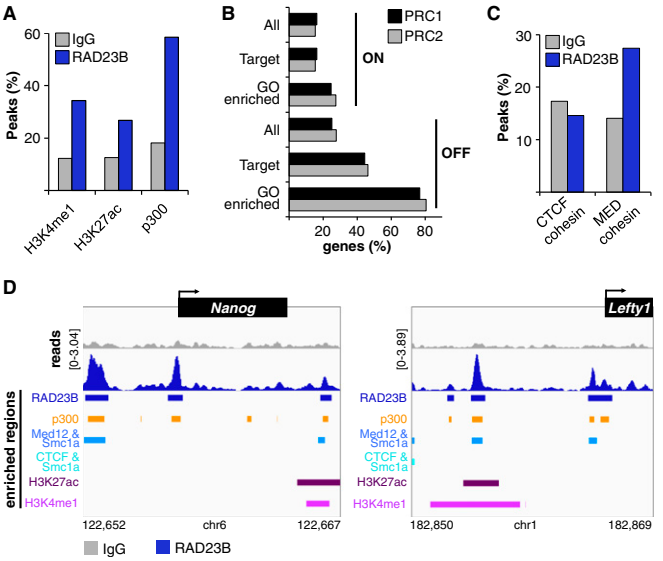

Fig. S2. Overlap between RAD23B binding sites and relevant factors in mESCs. (*A*) Percentage of control (IgG) and RAD23B ChIP-seq peaks overlapping with p300 binding sites and H3K4me1 and H3K27ac enhancer chromatin marks in mESCs (1). (*B*) Percent of Polycomb (PRC1 and 2)-bound active (ON) and inactive (OFF) genes (2) in D3 mESCs (All), and among all (Target) or GO-enriched RAD23B target genes in Fig. 1*D*. (*C*) Percentage of control (IgG) and RAD23B ChIP-seq peaks overlapping with CTCF/cohesin and Mediator (MED)/cohesin binding sites in mESCs (3). (*D*) Genomic regions enriched in RAD23B, p300, Mediator/cohesin (Med12 and Smc1a subunits) and CTCF/cohesin (Smc1a subunit) binding at *Nanog* (*Left*) and *Lefty1* (*Right*) gene loci evaluated by ChIP-seq. H3K27ac and H3K4me1 histone modifications are also shown. ChIP-seq tracks for normal IgG and RAD23B (*Upper*) were computed by IGV (4, 5), and plotted as (number of reads) × [1,000,000/(total read count)]; data range is the same for both tracks and specified on the *y* axis. Genomic coordinates are in kilobases.

1. Creyghton MP, et al. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 107(50):21931–21936.
2. Morey L, et al. (2012) Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* 10(1):47–62.
3. Kagey MH, et al. (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467(7314):430–435.
4. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
5. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.

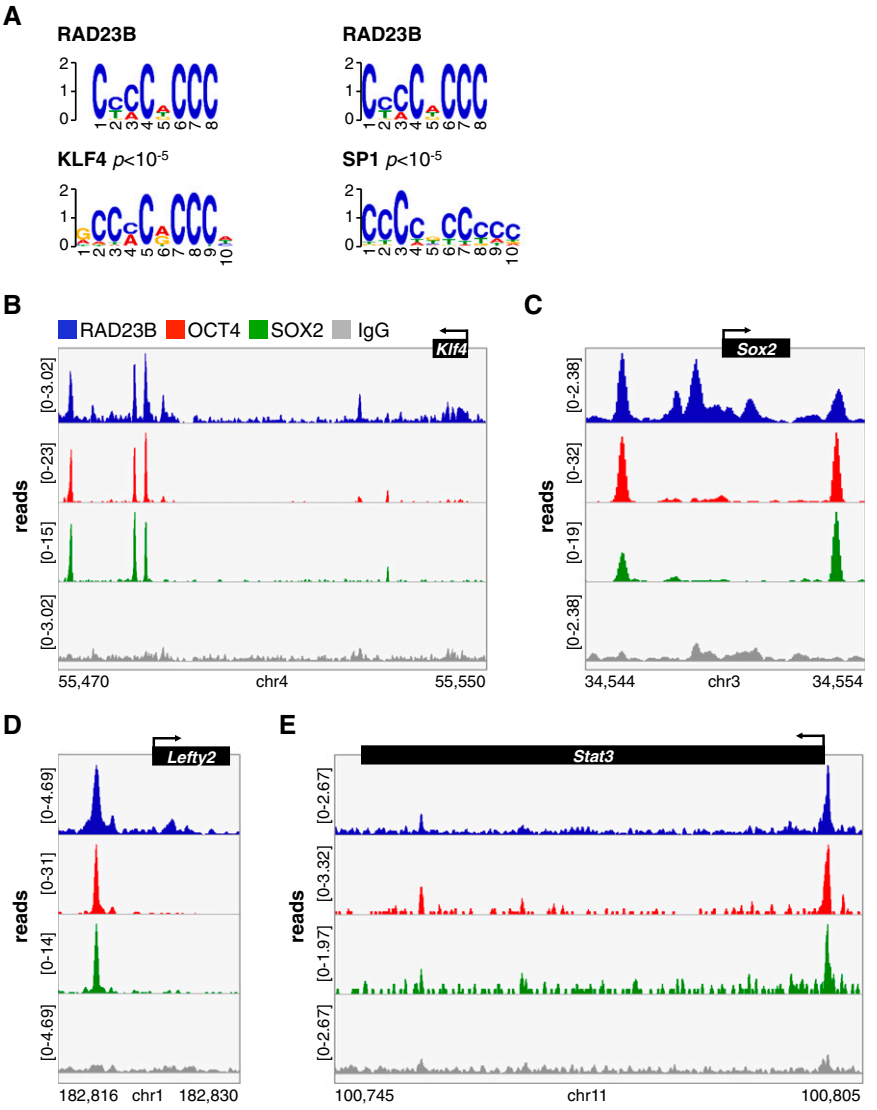

Fig. S3. RAD23B and OCT4/SOX2 extensively colocalize in mESCs. (*A*) Sequence analysis of OCT4/SOX2-overlapping RAD23B binding sites (peak midpoint ±125 bp) by MEME and TOMTOM motif discovery and comparison tools. Shown is a MEME secondary motif to the one in Fig. 2*B*. (*B–E*) Co-occupancy of RAD23B (blue), OCT4 (red), and SOX2 (green) proteins at *Klf4* (*B*), *Sox2* (*C*), *Lefty2* (*D*), and *Stat3* (*E*) loci evaluated by ChIP-seq. IGV-computed ChIP-seq tracks (1, 2) are plotted as (number of reads) x [1,000,000/(total read count)]; data range for each track is specified on the *y* axis. Genomic coordinates are in kilobases. Normal IgG (gray) controls for specificity.

1. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
2. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.

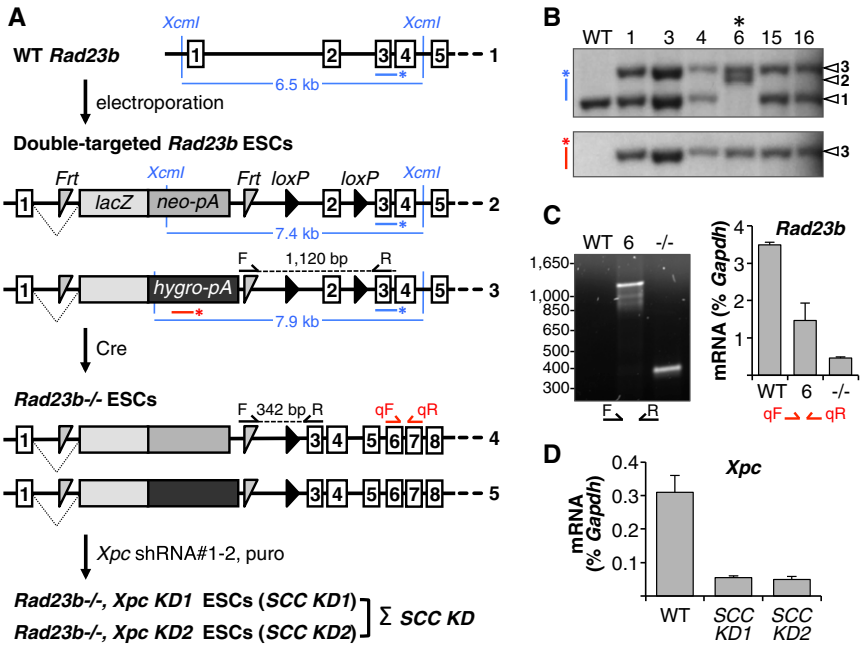

Fig. S4. Generation and characterization of *SCC* knockdown mESCs. (*A*) *Rad23b* genomic locus in WT mESCs (1) and in mESCs with neomycin (*neo*, 2, KOMP ID: 40373) or hygromycin (*hygro*, 3) targeting cassettes floxing exon 2. *lacZ* gene is preceded by a splicing acceptor site (dashed lines). In blue are *XcmI* restriction sites used for the Southern blot in *B* and expected genomic band sizes; asterisked lines indicate the probe position (blue, 3′ arm probe; red, hygromycin probe). *Frt*, Flp recombinase recognition site. *pA*, poly-adenylation signal. *loxP*, Cre recombinase recognition site. *Rad23b*$^{-/-}$ mESCs are obtained upon Cre-mediated excision of exon 2 (4 and 5). Primers used for the genotyping (F-R) and qPCR analyses (qF-qR) in *C* are specified. *Rad23b*$^{-/-}$ mESCs are finally infected with lentiviral vectors expressing two independent shRNA constructs targeting *Xpc* (shRNA#1–2) and selected with puromycin to obtain *SCC KD1-2* mESC lines. (*B*) Southern blot analysis of *Rad23b* WT and targeted mESC clones, hybridized with either a 3′ arm probe (blue, *Upper*) or an hygromycin probe (red, *Lower*). Expected band sizes are specified in *A*. Asterisk denotes a clone (#6) targeted on both alleles. (*C*, *Left*) Genomic PCR on *Rad23b* WT mESCs and double-targeted clone #6 before (6) and after ($^{-/-}$) Cre recombination. Expected band sizes are indicated in *A*. (*Right*) *Rad23b* transcript levels in WT mESCs and double-targeted clone #6 before (6) and after ($^{-/-}$) Cre recombination. Values are average and SD of qPCR technical duplicates, normalized to *Gapdh*. (*D*) *Xpc* transcript levels in WT mESCs and in *Rad23b*$^{-/-}$ mESCs after shRNA-mediated *Xpc* knockdown (*SCC KD1*, *SCC KD2*). Values are average and SD of qPCR technical duplicates, normalized to *Gapdh* (for protein levels, see Fig. S5*C*).

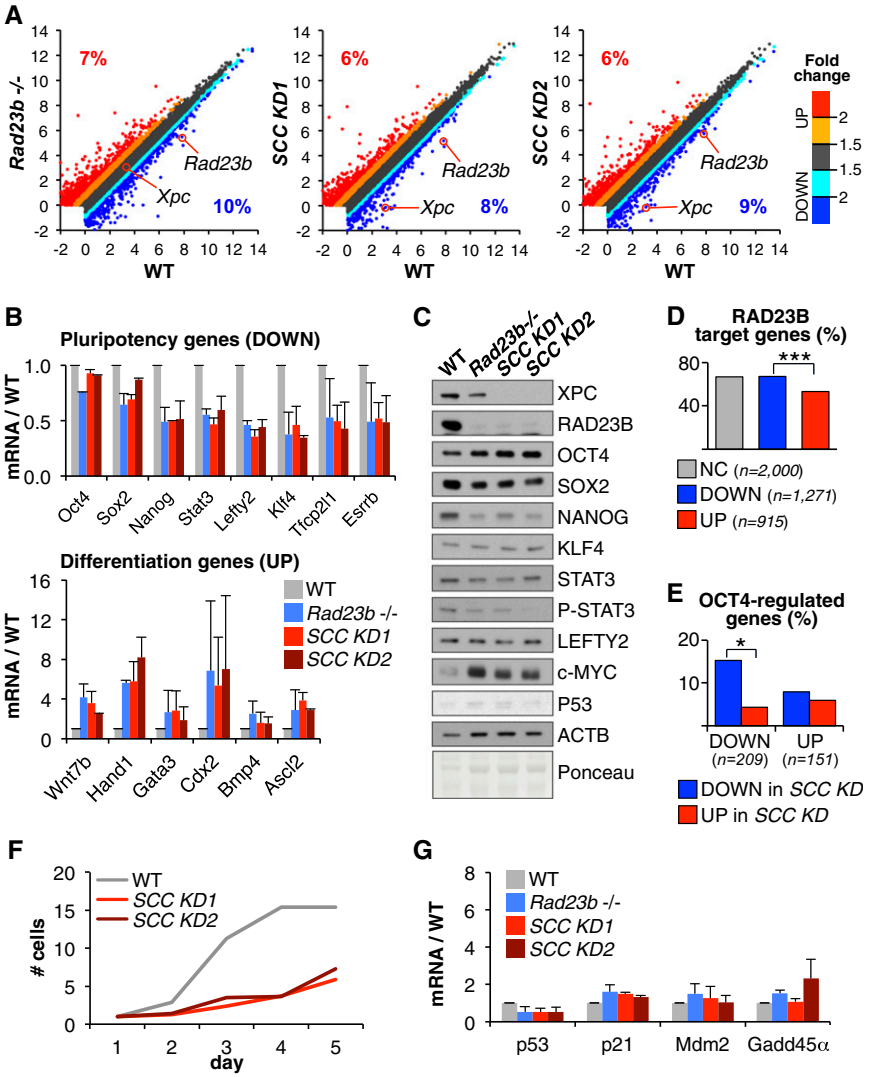

Fig. S5. Impaired stemness in SCC KD mESCs. (*A*) Gene-expression differences between WT and *Rad23b*$^{-/-}$ (*Left*), *SCC KD1* (*Center*), or *SCC KD2* (*Right*) mESCs evaluated by RNA-seq. Plotted are $\log_2$ (RPKM) values for Ensembl protein-coding genes, excluding genes with RPKM < 1 in both samples. The color code indicates the fold-change in gene expression over WT cells (gray, not changed). Total percentages of up-regulated (red) and down-regulated (blue) genes are also specified for each pairwise comparison. *Rad23b* and *Xpc* genes are highlighted. (*B*) RT-qPCR on selected pluripotency and differentiation markers in *Rad23b*$^{-/-}$ (blue), *SCC KD1* (red), and *SCC KD2* (cayenne) mESCs relative to WT cells. Values are average and SD of two independent experiments, normalized to *Gapdh* transcript levels. (*C*) Protein levels of selected genes in WT, *Rad23b*$^{-/-}$, *SCC KD1*, and *SCC KD2* mESCs (relative to Fig. 4*B*). β-Actin (ACTB) and Ponceau staining serve as loading controls. P-STAT3, antibody detecting active, Tyr705-phosphorylated STAT3. *SCC KD1* and *KD2* signals from this blot and an independent one are averaged in Fig. 4*B*. (*D*) Percentage of genes bound by RAD23B within 5 kb from their transcription start site among genes either not changed (NC), down-regulated (DOWN) or up-regulated (UP) upon *SCC KD*. The number of genes in each category is specified (*n*). ***$P < 10^{-3}$, two-sample test for equality of proportions with continuity correction. (*E*) Percentage of genes down-regulated (DOWN) or up-regulated (UP) in mESCs upon OCT4 knockdown (1) that are also deregulated in *SCC KD* mESCs. *$P = 0.05$, two-sample test for equality of proportions with continuity correction. (*F*) Growth curve of WT, *SCC KD1*, and *SCC KD2* mESCs. For each sample we seeded four multi-24 wells with 50,000 cells each (day 1), and detached and counted the whole well every day for the following 4 d. Cell numbers are relative to day 1. (*G*) RT-qPCR on p53 and selected p53 targets in *Rad23b*$^{-/-}$ (blue), *SCC KD1* (red), and *SCC KD2* (cayenne) mESCs relative to WT cells. Values are average and SD of two independent experiments, normalized to *Gapdh* transcript levels.

1. Matoba R, et al. (2006) Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PLoS ONE* 1:e26.

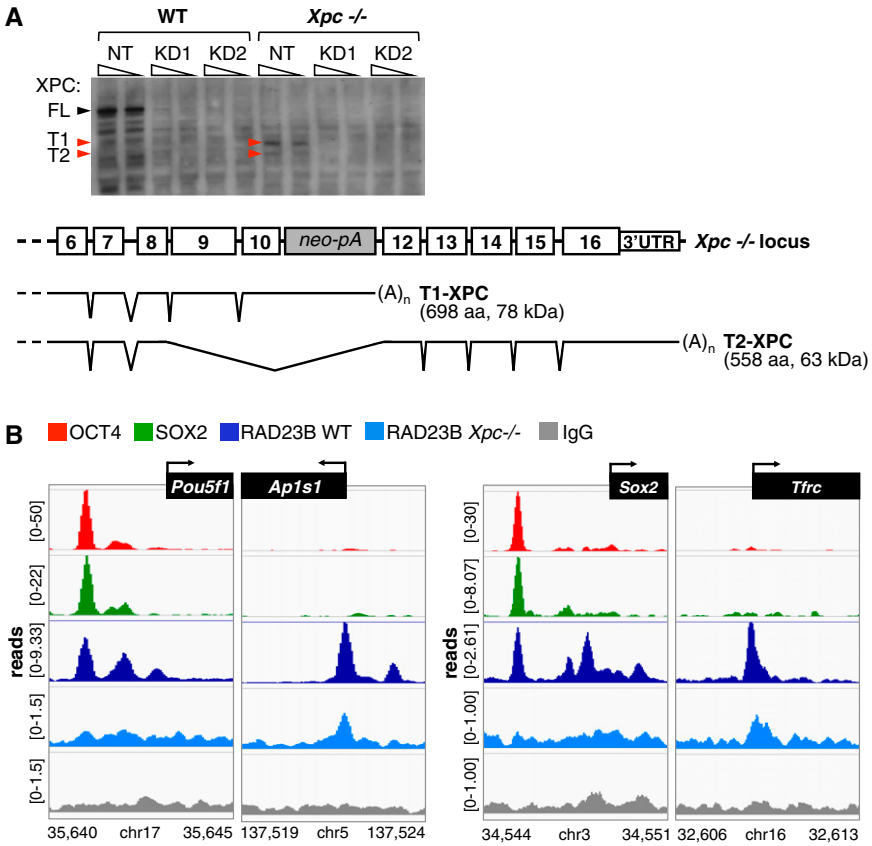

Fig. S6. XPC is required for SCC interaction with OCT4/SOX2 genome-wide. (*A*) Western blot analysis of total cell extracts from $Xpc^{-/-}$ ESCs (1) identified two XPC truncated proteins (T1 and T2), which could be specifically ablated by two independent shRNA constructs against XPC (KD1 and KD2 of Fig. S4). RT-PCR confirmed that the truncated products are encoded by two aberrant transcripts, one including exon 10 and the neomycin cassette replacing exon 11 (T1), and the other splicing in frame from exon 8 to exon 12, and excluding exons 9, 10, and the neomycin cassette (T2). Predicted protein length and molecular weight are specified. WT, wild type ESCs; FL, full-length XPC; NT, nontargeting shRNA construct. (*B*) Co-occupancy of OCT4, SOX2, and RAD23B proteins at *Oct4/Pou5f1* and *Ap1s1* loci (*Left*), and at *Sox2* and *Tfrc* loci (*Right*) evaluated by ChIP-seq. IGV-computed ChIP-seq tracks (2, 3) are plotted as (number of reads) × [1,000,000/(total read count)]; data range for each track is specified on the *y* axis. Genomic coordinates are in kilobases. Normal IgG (gray) controls for specificity.

1. Cheo DL, et al. (1997) Characterization of defective nucleotide excision repair in XPC mutant mice. *Mutat Res* 374(1):1–9.
2. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
3. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.

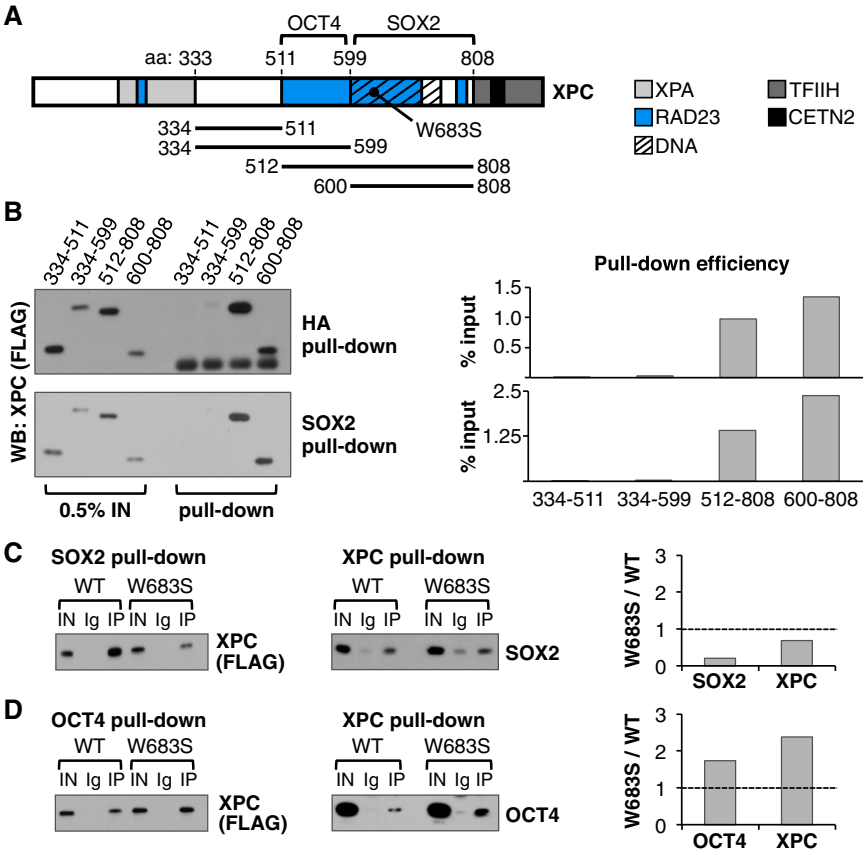

Fig. S7. OCT4 and SOX2 independently interact with XPC. (*A*) Scheme of XPC protein and its functional domains, and of the DNA binding mutant and the fragments used below. OCT4- and SOX2-interaction domains as inferred by truncation experiments are annotated. (*B*) Pull-down assays in 293T overexpressing HA-tagged SOX2 and FLAG-tagged XPC fragments by HA resin (*Upper*) or SOX2 antibody (*Lower*), followed by immunoblotting with anti-FLAG antibody. (*Right*) ImageJ quantification of pull-down efficiencies calculated as ratio to input (IN). (*C*) Pull-down assays in 293T overexpressing SOX2 and either WT or DNA-binding mutant (W683S) XPC FLAG-tagged proteins by SOX2 (*Left*) and XPC (*Center*) antibodies, followed by immunoblotting with anti-FLAG and anti-SOX2 antibodies, respectively. (*Right*) ImageJ quantification of pull-down efficiencies calculated as ratio to input relative to WT signal. IN, 0.5% input; Ig, IgG control pull-downs; IP, specific pull-downs. (*D*) Pull-down assays in 293T overexpressing OCT4 and either WT or W683S mutant XPC FLAG-tagged proteins by OCT4 (*Left*) and XPC (*Center*) antibodies, followed by immunoblotting with anti-FLAG and anti-OCT4 antibodies, respectively. (*Right*) ImageJ quantification as in *B*. IN, 0.5% input; Ig, IgG control pull-downs; IP, specific pull-downs.

Table S1. List of primers used in this study

Name	Sequence (5′–3′)	Figure
Nanog_E_for	GGGCACTCTCACACCAAAAT	2*C*, 3*C*
Nanog_E_rev	TTGACCTCTGTCCACATCCA	2*C*, 3*C*
Nanog_-950_for	GGCAAACTTTGAACTTGGGATGTGGAAATA	2*C*
Nanog_-950_rev	CTCAGCCGTCTAAGCAATGGAAGAAGAAAT	2*C*
Nanog_TSS_for	GGGTCACCTTACAGCTTCTTTTGCATTA	2*C*
Nanog_TSS_rev	GGCTCAAGGCGATAGATTTAAAGGGTAG	2*C*
Nanog_I_for	CGGGCTGATCTCAAATTCTC	2*C*
Nanog_I_rev	AATGGCAGGATGAAGACAGG	2*C*
Pou5f1_-5000_for	GTTTGATTAGGCTTGAGAGGAGACC	2*C*
Pou5f1_-5000_rev	GAGTTGTCCCTGTGCCTCTGTAAAG	2*C*
Pou5f1_E_for	GGAACTGGGTGTGGGGAGGTTGTA	2*C*, 3*C*
Pou5f1_E_rev	AGCAGATTAAGGAAGGGCTAGGACGAGAG	2*C*, 3*C*
Pou5f1_TSS_for	AGCAACTGGTTTGTGAGGTGTCCGGTGAC	2*C*
Pou5f1_TSS_rev	TCCCCAATCCCACCCTCTAGCCTTGAC	2*C*
Pou5f1_I_for	TTCCATTGAAATGGCATTGA	2*C*
Pou5f1_I_for	CGGTTTCTCAGGCTTTTCTG	2*C*
Oct4_mRNA_for	CCAGGCAGGAGCACGAGTGG	3*B*
Oct4_mRNA_rev	CCTGGGACTCCTCGGGAGTTG	3*B*
Sox2_mRNA_for	GAGTGGAAACTTTTGTCCGAGA	3*B*
Sox2_mRNA_rev	GAAGCGTGTACTTATCCTTCTTCAT	3*B*
Rad23b_mRNA_for (qF)	AATTGCAGCCCTGAGAGCCAGC	3*B*, 4 *A*–*C*
Rad23b_mRNA_rev (qR)	GCTGTCGTGGTTGCTGCAGCT	3*B*, 4 *A*–*C*
Gapdh_mRNA_for	TGTGCAGTGCCAGCCTCGTC	3*B*
Gapdh_mRNA_rev	TGAAGGGGTCGTTGATGGCAACA	3*B*
Klf4_E_for	AGGGAGAGGGGAAATGAAGA	3*C*
Klf4_E_for	CCCGTCAAAGCACTAGCTTC	3*C*
Actb_TSS_for	CATGGTGTCCGTTCTGAGTGATC	3*C*
Actb_TSS_rev	ACAGCTTCTTTGCAGCTCCTTCG	3*C*
3′arm_probe_for	ACTAGTGCTAGCAAAGGCAGTGACCAATCCAC	4 *A* and *B*
3′arm_probe_rev	ACTAGTCTCGAGCTCAAGGGCCATCACCTTTA	4 *A* and *B*
Hygro_probe_for	GTGGAGGAGAATCCCGGCCCTGGGATCCCCAGCTTGGCCA	4 *A* and *B*
Hygro_probe_rev	ACTAGTGCGCGCCTATTCCTTTGCCCTCGGACG	4 *A* and *B*
Rad23b_genotype_F	CACACCTCCCCCTGAACCTGAAAC	4 *A*–*C*
Rad23b_genotype_R	GGAACAATTTCATGCACATTAGTG	4 *A*–*C*
Xpc_mRNA_for	TCGATTTCCATGGAGGCTAT	4*D*
Xpc_mRNA_rev	TAGCGGAGTTTCAGCCTCTC	4*D*
Nanog_mRNA_for	CCTCAGCCTCCAGCAGATGC	S5*B*
Nanog_mRNA_rev	CCGCTTGCACTTCATCCTTTG	S5*B*
Stat3_mRNA_for	CAATACCATTGACCTGCCGAT	S5*B*
Stat3_mRNA_rev	GAGCGACTCAAACTGCCCT	S5*B*
Lefty2_mRNA_for	CGCTGGACCTCAAGGACTAC	S5*B*
Lefty2_mRNA_rev	GCCCACACATTCATACGTCA	S5*B*
Klf4_mRNA_for	CAGTGGTAAGGTTTCTCGCC	S5*B*
Klf4_mRNA_rev	GCCACCCACACTTGTGACTA	S5*B*
Tfcp2l1_mRNA_for	CAGCCCGAACACTACAACCAG	S5*B*
Tfcp2l1_mRNA_rev	CAGCCGGATTTCATACGACTG	S5*B*
Esrrb_mRNA_for	GCACCTGGGCTCTAGTTGC	S5*B*
Esrrb_mRNA_for	TACAGTCCTCGTAGCTCTTGC	S5*B*
Wnt7b_mRNA_for	GTCTTCGGGCAAGAACTCC	S5*B*
Wnt7b_mRNA_rev	GTCACAGCCACAATTGCTCA	S5*B*
Hand1_mRNA_for	CGCCTGGCTACCAGTTACAT	S5*B*
Hand1_mRNA_rev	GCGCCCTTTAATCCTCTTCT	S5*B*
Gata3_mRNA_for	GGGTTCGGATGTAAGTCGAG	S5*B*
Gata3_mRNA_rev	CCACAGTGGGGTAGAGGTTG	S5*B*
Cdx2_mRNA_for	CAAGGACGTGAGCATGTATCC	S5*B*
Cdx2_mRNA_rev	GTAACCACCGTAGTCCGGGTA	S5*B*
Bmp4_mRNA_for	TTCCTGGTAACCGAATGCTGA	S5*B*
Bmp4_mRNA_rev	CCTGAATCTCGGCGACTTTTT	S5*B*
Ascl2_mRNA_for	TTTTCGAGGACGCAATAAGC	S5*B*
Ascl2_mRNA_rev	CACTGCTGCAGGACTCCCTA	S5*B*
p53_mRNA_for	GCTCACCCTGGCTAAAGTTCTG	S5*G*
p53_mRNA_rev	AGTCGCTACCTACAGCCAGGAT	S5*G*
p21_mRNA_for	ACGGTGGAACTTTGACTTCG	S5*G*

Table S1. Cont.

Name	Sequence (5′–3′)	Figure
p21_mRNA_rev	CAGGGCAGAGGAAGTACTGG	S5*G*
Mdm2_mRNA_for	CTGTGTCTACCGAGGGTGCT	S5*G*
Mdm2_mRNA_rev	CGCTCCAACGGACTTTAACA	S5*G*
Gadd45alpha_mRNA_for	GCTCAACGTAGACCCCGATA	S5*G*
Gadd45alpha_mRNA_rev	GCGTCGTTCTCCAGTAGCAG	S5*G*
Xpc_cDNA_for	ACGCGTCGACAACGCCTCTGGCTAGCATGG	6*A*, S7*A*
Xpc_cDNA_808_rev	ACTAGTGGTACCCTCACAGACAATGTAGCC	6*A*, S7*A*
Xpc_cDNA_599_rev	ACTAGTGGTACCGGTCTCAGCCCACCACTC	6*A*, S7*A*
Xpc_cDNA_511_rev	ACTAGTGGTACCTTCAGCACCACTGGAAAC	6*A*, S7*A*
Xpc_cDNA_334_for	ACTAGTGTCGACATGACATCTGTAGAGGGTCCTG	S7*A*
Xpc_cDNA_512_for	ACTAGTGTCGACGAGATGGCAGATAGGAAACC	S7*A*
Xpc_cDNA_600_for	ACTAGTGTCGACATGTTGAGACCCTATCGGAGCC	S7*A*
Sox2_HA_cDNA_for	ACTAGTGTCGACATGTATAACATGATGGAGACGGAGCTGAAGCCG	S7*B*
Sox2_HA_cDNA_rev	ACTAGTGGTACCTACTAAGCGTAATCTGGAACATCGTATGGGTAACCTCCGGATCCCATGTGCGACAGGGGCAG	S7*B*

Dataset S1. ChIP-seq data analysis

Dataset S1

The table contains RAD23B ChIP-seq and Gene Ontology analyses of WT and $Xpc^{-/-}$ mESCs (1–4).

1. Marson A, et al. (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134(3):521–533.
2. Liu Z, Scannell DR, Eisen MB, Tjian R (2011) Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* 146(5):720–731.
3. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.
4. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.

Dataset S2. RNA-seq data analysis

Dataset S2

The table contains Cuffdiff and Gene Ontology analyses of SCC-depleted mESCs (1, 2).

1. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.
2. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.