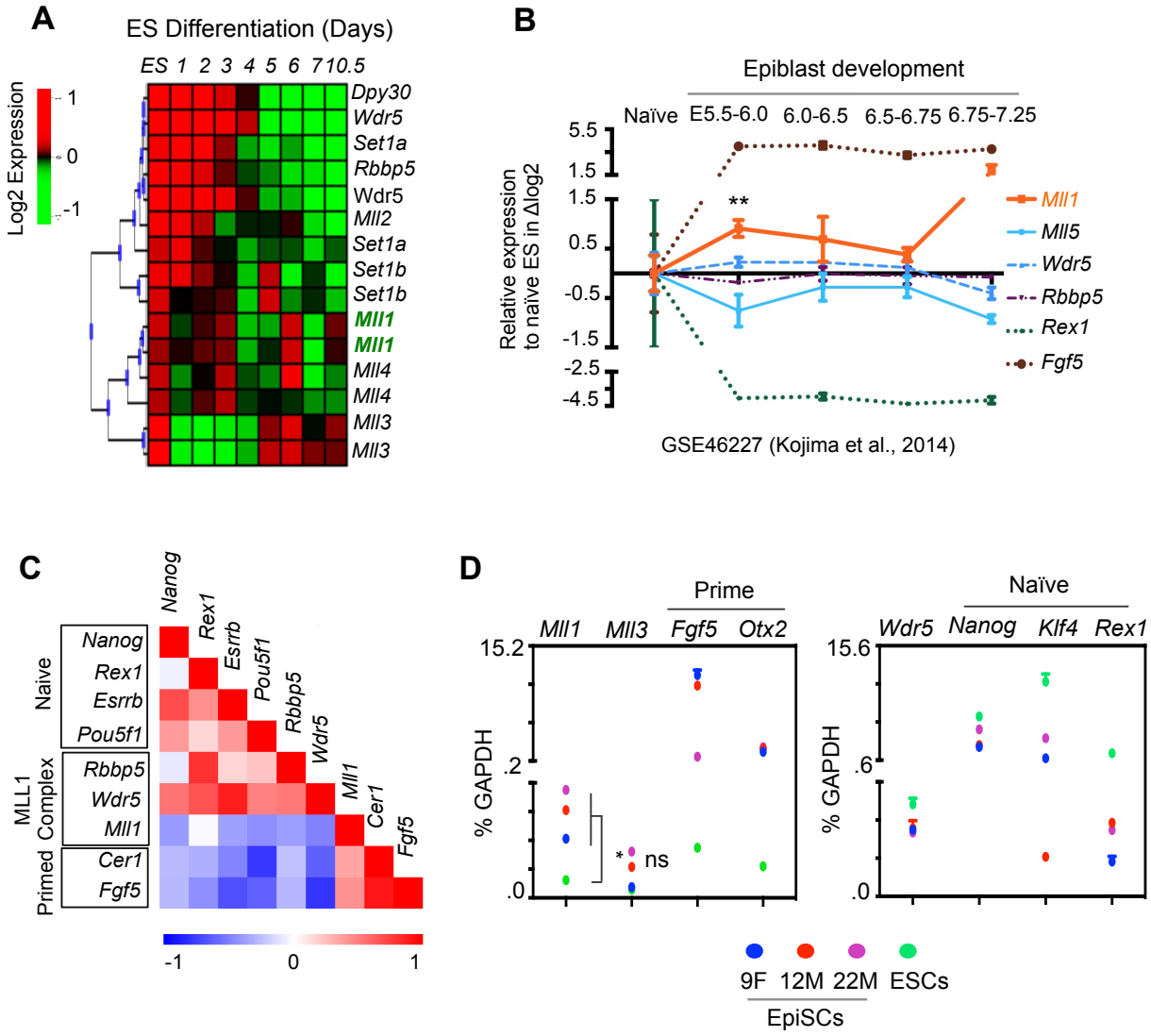
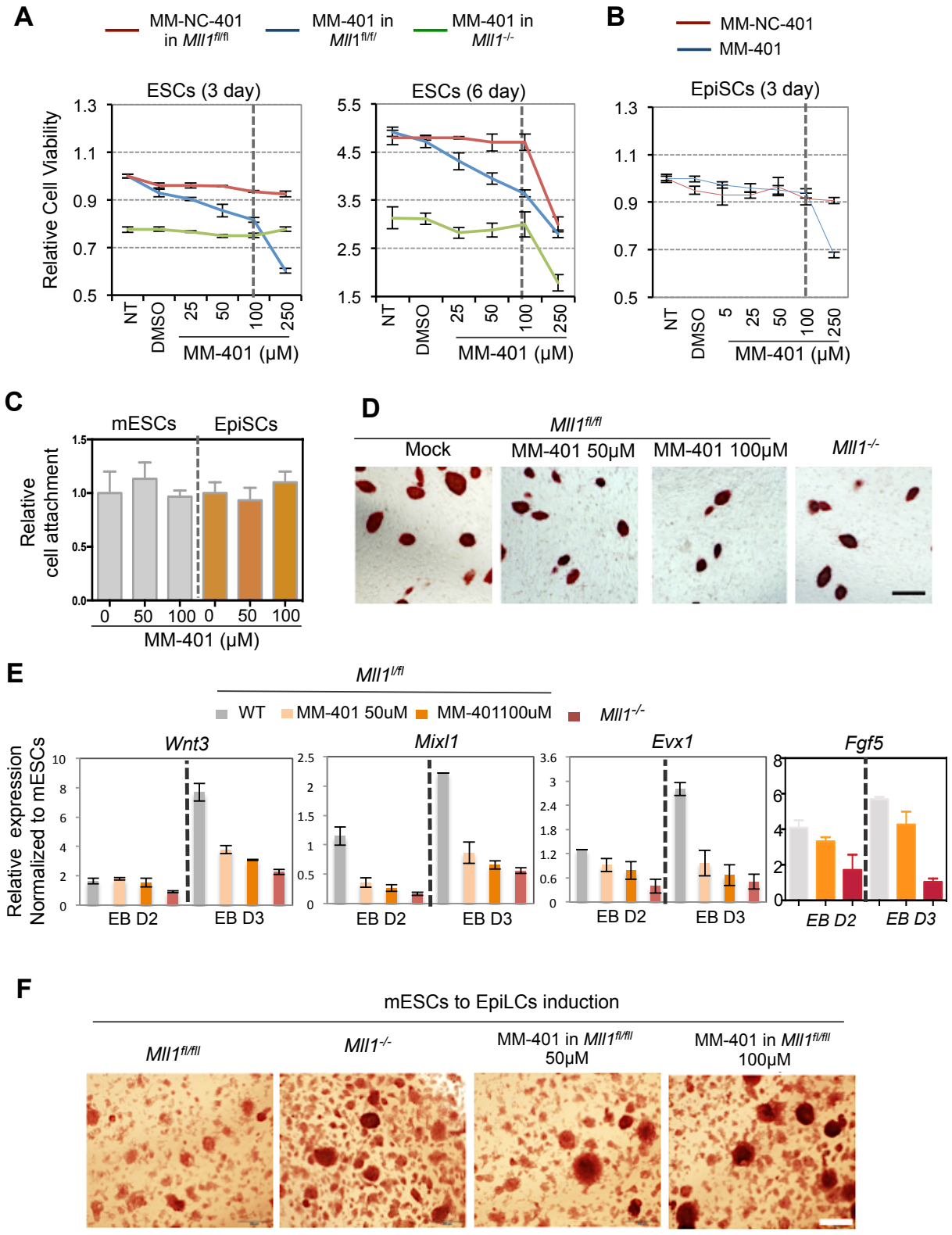
Supplemental Figures

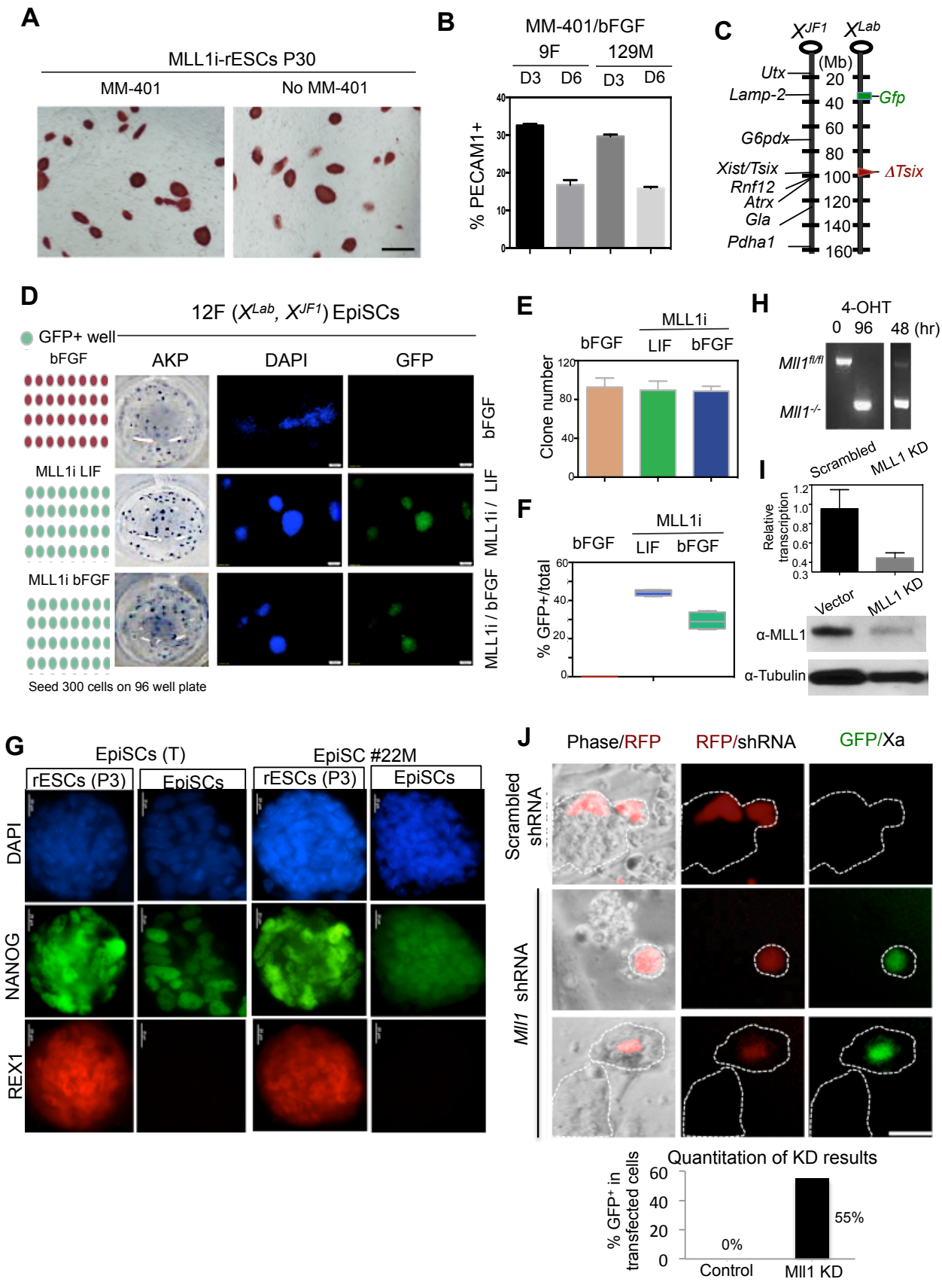
Supplemental Figure 1



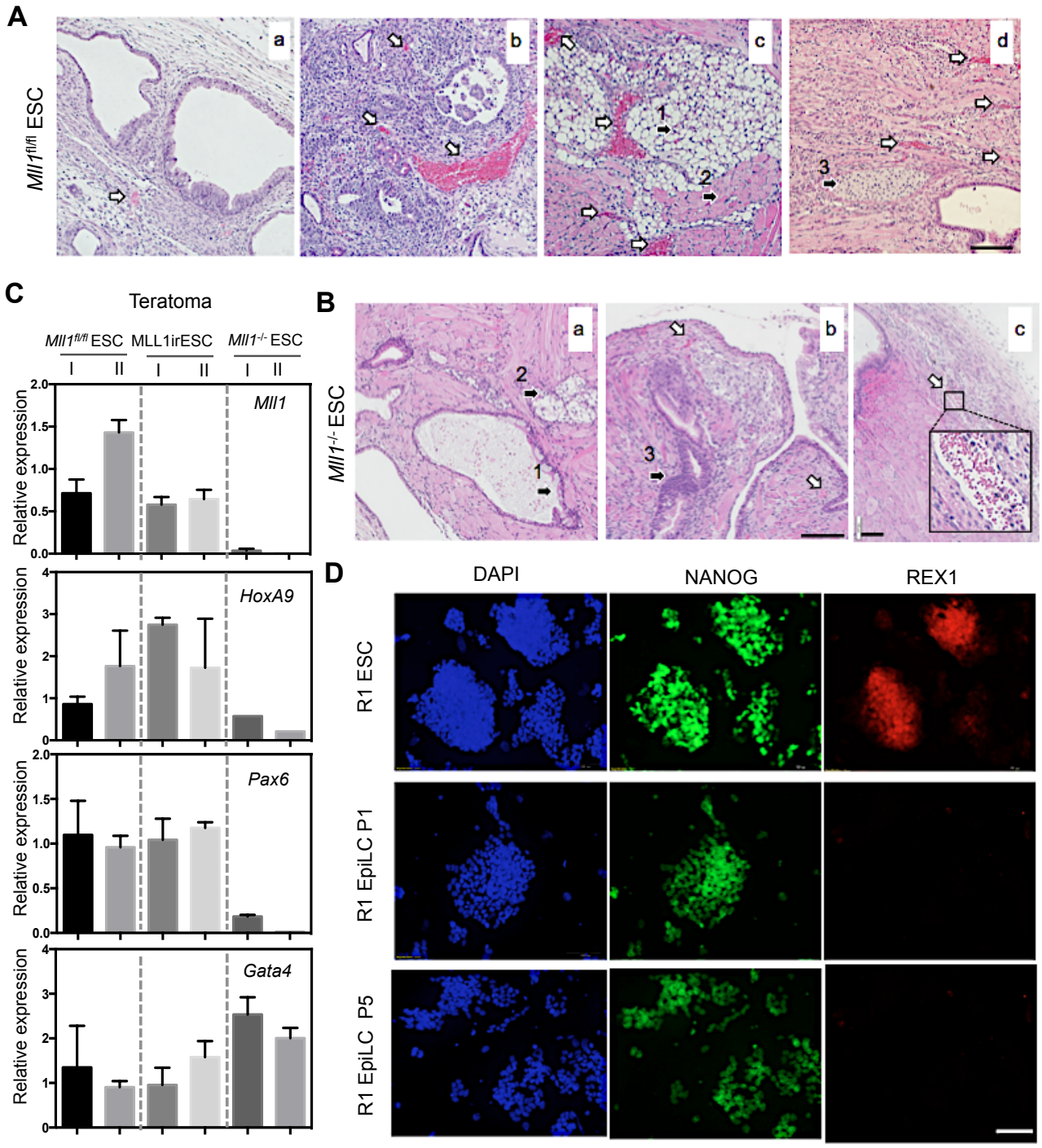
Supplemental Figure 2



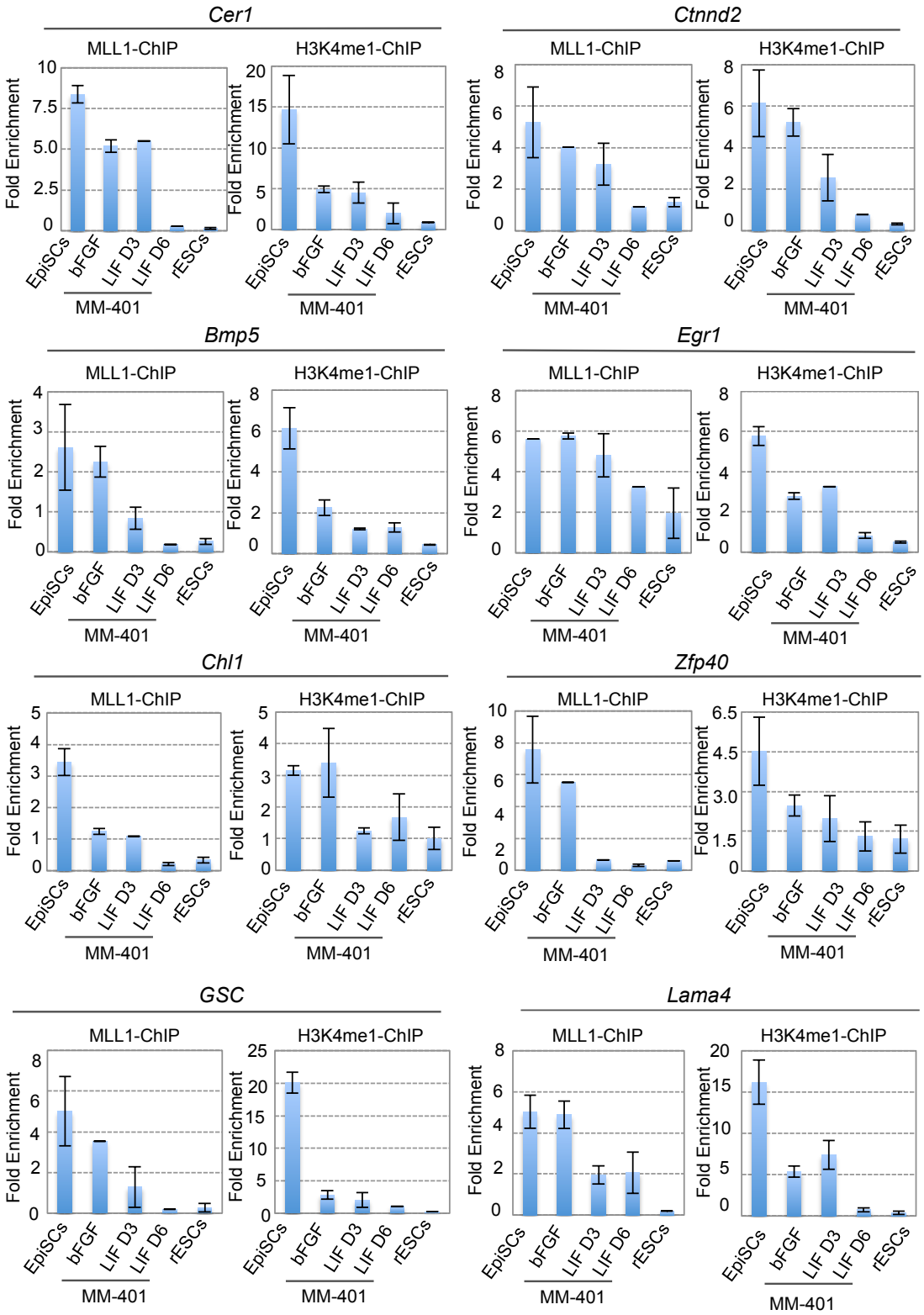
Supplemental Figure 3



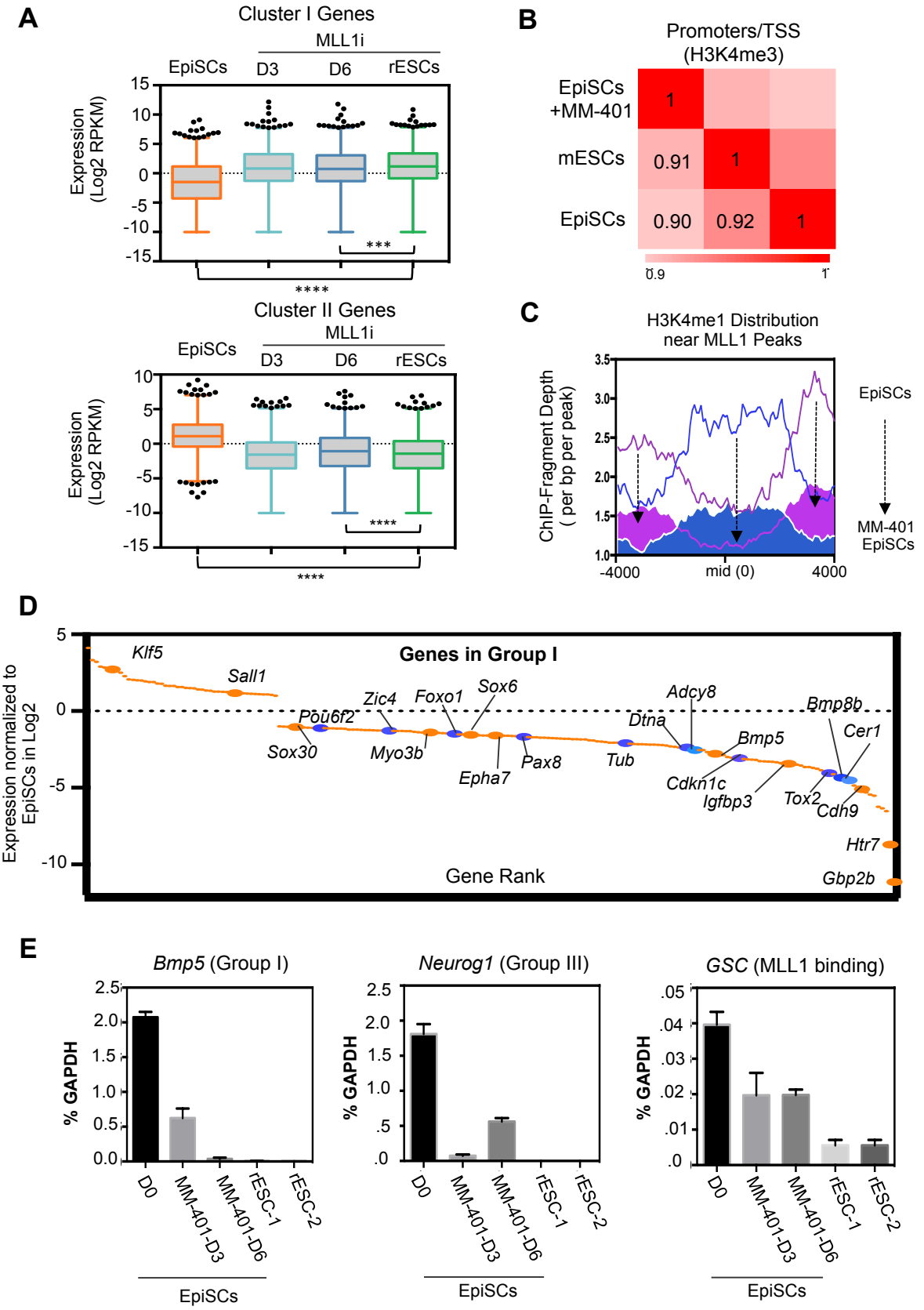
Supplemental Figure 4



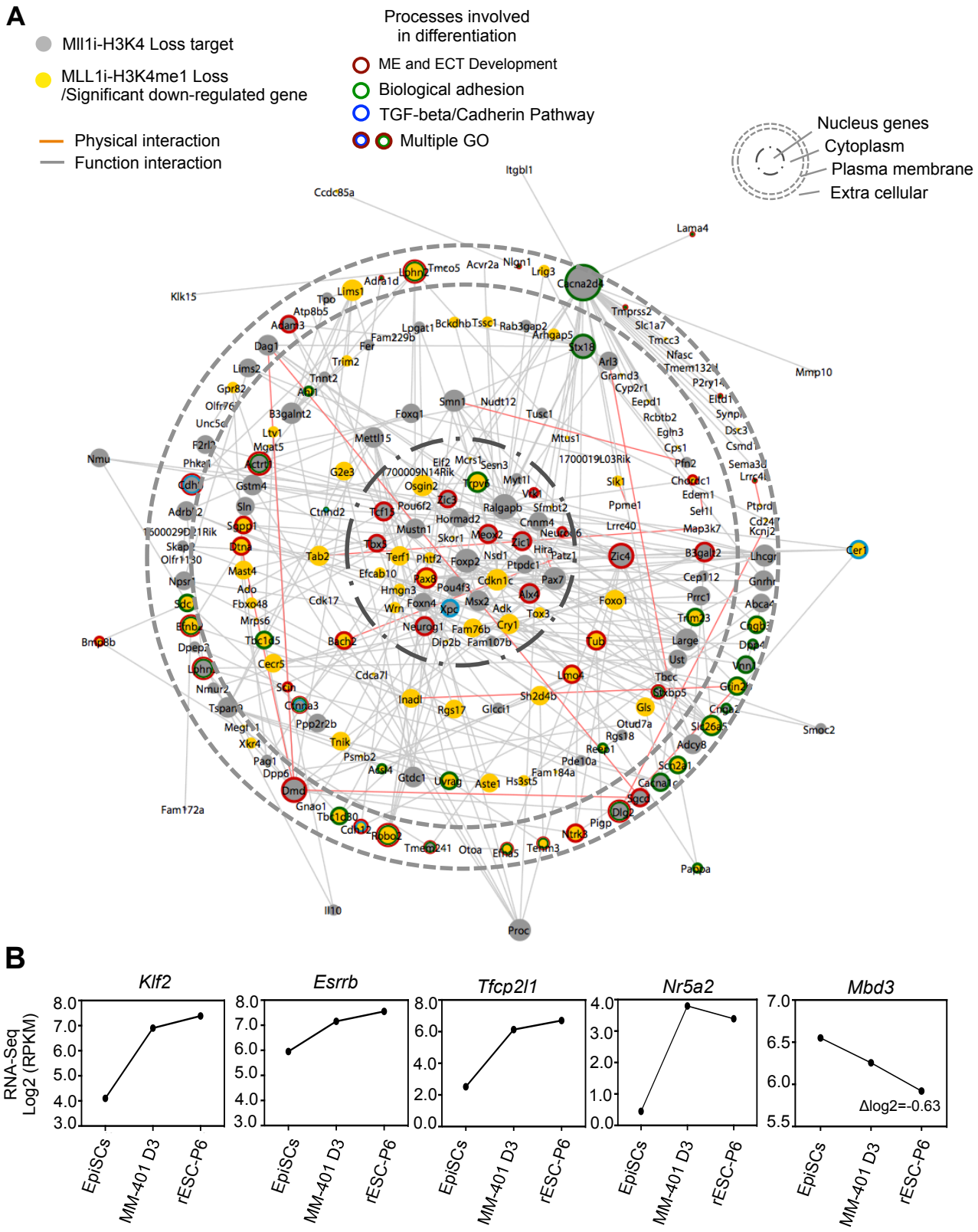
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure Legends

Supplemental Figure 1 (related to Figure 1). Dynamic regulation of *Mll1* expression during ESC differentiation and development.

A. Heat map for expression of MLL family HMTs as well as core components of the MLL complexes. Visualization and hierarchical clustering were done using FunGene database (http://biit.cs.ut.ee/fungenes/). Two different Affix IDs for each member were included except *Dpy30*, *Rbbp5* and *Mll2*. **B.** Relative expression of selected genes (indicated on right) during epiblast development *in vivo*. The microarray data was reported by (Kojima et al., 2014). Y-axis, log2 expression change relative to levels in naïve ESCs, which was arbitrarily set as 0. **C.** Pearson correlation matrix for expression of selected genes as indicated on left. Heat map key represents Pearson correlation coefficient. RNA-seq data used for analyses include 4 replicates of ESCs and EpiSCs from Factor et al., 2014 and duplicates of RNA-seq we performed for EpiSCs. **D.** Relative expression of genes in different cell lines as indicated on bottom. Average expression from three independent experiments was presented after normalization against Gapdh in each cell line. Data was presented as mean ± SD (error bars) from three independent experiments. *Mll1* expression in ESCs was significantly different from its average expression in three EpiSC lines. *, $p<0.05$; n.s., no significant difference from student-*t* test.

Supplemental Figure 2 (related to Figure 1). MM-401 treatment delays ESC-to-EpiSC differentiation

A. Cell viability assay for $Mll1^{-/-}$ and $Mll1^{f/f}$ ESCs treated with different concentrations of MM-401 or MM-NC-401 as indicated. Y-axis, relative viability against untreated $Mll1^{f/f}$ ESCs at Day 3, which was arbitrarily set as 1. **B.** Relative cell viability assay for EpiSCs treated with different concentrations of MM401 or MM-NC-401 as indicated on bottom. Viability of untreated EpiSCs was arbitrarily set as 1. For **A** and **B**, NT, no treatment; DMSO, mock treatment. Means and standard deviations (as error bars) from at least three independent experiments were presented. **C.** Relative cell attachment for ESCs and EpiSCs treated with different concentrations of MM-401. Attachment rate was normalized against untreated ESCs, which was arbitrarily set as 1. Means and standard deviations (as error bars) from at least three independent experiments were presented. **D.** Representative AKP stained $Mll1^{-/-}$ or $Mll1^{flox/flox}$ ESC clones were shown. Scale bar, 500μm. **E.** Real-time PCR for expression of selected genes at day 2 (D2, left) or day 3 (D3, right) of embryoid body formation (EBs) as indicated on top. Gene expression was presented as fold change to their levels in ESCs, which was arbitrarily set as 1. Means and standard deviations (as error bars) from at least three independent experiments were presented. **F.**

Representative *AKP*-stained clones after induction of ESC differentiation by bFGF and Activin A. *Mll1*$^{f/f}$, *Mll1*$^{-/-}$ or *Mll1*$^{f/f}$ ESCs treated with MM-401 were indicated on top. Scale bar, 500μm.

Supplemental Figure 3 (related to Figure 1, 2 and 3). MM-401 treatment reprograms EpiSCs to ESCs

A. Representative AKP staining for Mll1i-rESCs at passage 30 (P30) that were maintained with or without MM-401 since P6. Scale bar, 500μm. **B.** Percent of PECAM1^{+} cells in two EpiSC lines, as indicated on top, which were cultured with bFGF at day 3 (D3) or day 6 (D6). Data were represented as mean ± S.D. from three independent experiments. **C.** Schematic for *X*^{Lab} and *X*^{JF1} X-chromosomes in 12F EpiSCs. *Gfp* (green), *Tsix* (red) as well as dosage-compensated genes (e.g. *AtrX*) on X chromosomes are highlighted. Distance from the centromere is labeled by incremental of 20 megabases (Mb). **D.** Quantitation of X-chromosome reactivation (XCR) after 3-day MM-401 treatment. 300 EpiSCs were seeded in each well. Left panels, summary of XCR in 32 wells under conditions shown on top. Red, GFP^{-} well; Green, GFP^{+} well. Middle panels, representative images for AKP^{+} clones in each well. Right two panels, representative DAPI staining and GFP fluorescence for EpiSCs or EpiSCs treated with MM-401/LIF or MM-401/bFGF as indicated on right. Scale bar, 100μm. **E.** Cell attachment assay for EpiSCs cultured under different conditions as indicated on top. Y-axis, clone numbers per 300 seeded single cells. **F.** Quantification of GFP^{+} clones after normalization against total clone numbers, which were obtained from E. Means (middle line) and standard deviations (top/bottom edges of the box) from 32 wells were presented. **G.** Immunofluorescence for NANOG and REX1 (indicated on left) in 129.T (left) and 22M (right) EpiSCs as well as 129.T-rESCs and 22M-rESCs as indicated on top. Scale bar, 20μm. **H.** Genotyping for *Mll1* at different time point after 4-OHT treatments as indicated. **I.** Top, real-time PCR analysis of *Mll1* transcripts in mouse embryonic fibroblast (MEF) upon 3-day *Mll1*-shRNA knockdown Bottom, immunoblot of MLL1 after shRNA treatments as indicated on top. Immunoblot of Tubulin was included as the loading control. Gene expression after normalization against Gapdh was presented as fold change to its respective level in scramble-shRNA treated cells, which was arbitrarily set as 1. Means and standard deviations (as error bars) from three independent experiments were presented (n=3). **J.** Knockdown *Mll1* in 12F EpiSCs. Cells were treated with MM-401 or transfected with *Mll1* shRNA or scramble shRNA as indicated. Cells that were successfully transfected with shRNAs expressed RFP (red). Xi-reactivation was indicated by re-expression of GFP (green). Representative phase contrast and RFP/GFP fluorescence images were presented. Quantification of percentage of GFP^{+} cells in the transfected cells (RFP^{+}) was shown at the bottom.

Supplemental Figure 4 (related to Figure 4). *In vivo* characterization of Mll1i-rESCs and *Mll1*$^{-/-}$ ESCs

A Hematoxylin and Eosin (H&E) staining of tissues derived from *Mll1*$^{flox/flox}$ teratomas. (a) respiratory-like epithelium from endoderm; (b) neural epithelium from ectoderm and (c and d) mesoderm. The arrows indicated: (1) adipose-like tissue; (2) muscle; and (3) cartilage. **B.** H&E staining of tissues derived from *Mll1*$^{-/-}$ teratomas. a) endoderm and mesoderm; b) ectoderm; c) red blood cells and blood vessel (zoom in). The closed arrows indicated (1) gastrointestinal-like epithelium; (2) adipose-like tissue; and (3) neural epithelium. In both **A** and **B**, blank arrow indicated red blood cells and blood vessel, which were less developed in *Mll1*$^{-/-}$ teratomas. Scale bar, 100μm. **C.** Expression of selected genes (as indicated in box) in *Mll1*$^{flox/flox}$, MLL1i-rESCs and *Mll1*$^{-/-}$ teratomas**.** Gene expression after normalization against *Gapdh* was presented as fold change to their respective levels in *Mll1*$^{flox/flox}$ teratomas. Averaged expression from two *Mll1*$^{flox/flox}$ teratomas was arbitrarily set as 1. Means and standard deviations (as error bars) from three independent experiments were presented (n=3). **D.** Immunofluorescence for NANOG and REX1 (as indicated on top) using R1-ESCs and two independent derived R1-EpiLCs at different time points after MM-401 treatment as indicated on left. P1, passage 1, P5, passage 5. Scale bar, 100μM.

Supplemental Figure 5 (related to Figure 6). ChIP confirmation for MLL1 and H3K4me1 ChIP-seq results

A-H. ChIP for MLL1 and H3K4me1 at selected MLL1 targets as indicated on top. Cells used for ChIP assays were indicated on bottom. D3, 3-day MM-401 treatment; D6, 6-day MM-401 treatment. bFGF or LIF indicates cytokines in the culture medium at the time of MM-401 treatment. Signals for each experiment were presented as fold enrichment against those of input, which was arbitrarily set as 1. Means and standard deviations (as error bars) from at least three independent experiments were presented.

Supplemental Figure 6 (related to Figure 5, 6 and 7). Gene expression and H3K4me changes during EpiSC reversion

A. Box plots showing expression (log2 RPKM) of cluster I (Top) and cluster II (Bottom) genes that are defined in Figure 5D at different time points of EpiSC reversion as indicated on top. The central mark represents median and edges of the box represented the 25th and 75th percentiles of the range in relative expression, respectively. The whiskers extended to 5th to 95th percentile and outliers were plotted in dot. Statistical significance was calculated by student *t*-test. ***,

p<0.001, ****, *p*<0.0001. **B.** Pearson correlation matrix of H3K4me3 at gene promoters. n= 34,000 retrieved from UCSC. H3K4me3 in ESCs (GSE47949) and EpiSCs (GSE57407) are from public database. Analyses details see Supplemental Methods. **C.** H3K4me1 levels before (line) and after (solid) MM-401 treatment near MLL1 peak centers, which were shown as midpoint 0. Pink and blue colors represented two groups of genes that had distinct H3K4me distribution patterns. ChIP-Seq signals of H3K4me1 were calculated by tag count per bp per peak and presented as ChIP-fragment depth in Y-axis. **D.** Gene rank of group I MLL1 targets based on Δlog2 expression change in ESCs vs. EpiSCs. Only genes that have Δlog2 expression change >1 or <-1 were included in the rank list. Selected genes were highlighted in blue. **E**. Real-time PCR for selected genes as indicated on top. Gene expression was normalized and presented as %*Gapdh* in different cells as indicated on bottom. Means and standard deviations (as error bars) from at least three independent experiments were presented.

Supplemental Figure 7 (related to Figure 6 and 7). RNA-Seq results for expression of selected genes during EpiSCs reversion

A. The MLL1-dependent gene network based on physical interactions (red line) or phenotypic associations (grey line). The size of the node was proportional to respective contribution to the gene network. The diagram for subcellular localization of proteins was indicated on top right corner. Genes that had lower expression (log2 fold change < -1) after MM-401 treatment were highlighted in yellow. In this layout, different subcellular localizations are presented as concentric circles. Genes belong to specific pathways were indicated by green (biological adhesion), red (developmental process) or blue (TGF-β/Cadherin Pathway) line circles. **B.** Expression of selected naïve stem cell markers during EpiSC reversion. All genes shown here were previously reported to promote EpiSC reversion. MM-401-D3 and MM-401-D6 are EpiSCs treated with MM-401 for three or six days, respectively. Average Log2 (RPKM) from RNA-seq duplicates was presented after normalization.

Supplemental Tables

Supplemental Table 1 (related to Figure 5). Complete list of gene in cluster I and cluster II in Figure 5.

Supplemental Table 2 (related to Figure 6). Complete lists of MLL1 binding sites in EpiSCs.

Supplemental Table 3 (related to Figure 6). Complete lists of H3K4me1 sites in EpiSCs after MM-401 treatment.

Supplemental Table 4 (related to Figure 6). Complete lists of the H3K4me3 sites in EpiSCs after MM-401 treatment.

Supplemental Table 5 (related to Figure 6). Complete lists of group I, II and III genes.

Supplemental Table 6 (related to Figure 6 and Figure S5). Primers for ChIP-qPCR experiments.

Supplemental Table 7 (related to Figure 5 and Figure S1, S4 and S6). Primers for reverse transcription-qPCR.

Supplemental Table 6 (related to Figure 6 and Figure S5) Primers for ChIP-qPCR confirmation.

Cer1_mll1F	CGAGTTCCTTGCCCAACCTA
Cer1_mll1R	ACACACTGCTCAGTAGCACA
Ctnnd2_mll1F	CCTGTGCACACACATACCCA
Ctnnd2_mll1R	ACAGGCACAGTTTTGGGGAG
Bmp5_mll1F	ACTTGTGCATGTAATGAGTCCT
Bmp5_mll1R	CACTTAGCTTTGAAAGGAATACCA
Egr1_mll1F	CTGCCGGTGGGGATATAGGA
Egr1_mll1R	CCACTGTGCTGGGATGTCTA
Chl1_mll1F	ACAACTACAACACAGACACTACCA
Chl1_mll1R	ACTCATGAGGGACAAAGTATCTGG
Zfp40_mll1F	TTTTGAGATCCGGGGCTGGT
Zfp40_mll1R	TGGAAGGAAAGTGAATTGCGTC
GSC_mll1F	AGGCCCATTGTACTTCTGGC
GSC_mll1R	CACCTGCAGAGCTGGTGTAA
Lama4_mll1F	CAGCACAGTGATCTGGATCTCAA
Lama4_mll1R	AGTGCAAGTTTGGGACTTCCG

Supplemental Table 7 (related to Figure 5 and Figure S1, S4 and S6). Primers for testing RNA expression by real-time PCR.

Gadph F	TGGATTTGGACGCATTGGTC
Gadph R	TTTGCACTGGTACGTGTTGAT
Wdr5F	AGGAAGACCTTGACTCCACGA
Wdr5R	AAGACGAACACATCTCCGTCC
Mll1R	GCAGATTGTAAGACGGCGAG
Mll1R	GAGAGGGGGTGTTCCTTCCTT
Mll1exon9-10F	ATGCAGGCACTTTGAACATCC
Mll1exon9-10R	GTTATGGGGACAGAGGTCAGG
Rex1r	ACGAGGTGAGTTTTCCGAAC
Rex1f	CCTCTGTCTTCTCTTGCTTC
Klf4F	GTGCCCCGACTAACCGTTG
Klf4R	GTCGTTGAACTCCTCGGTCT
Fgf5 F	AAGTAGCGCGACGTTTTCTTC
Fgf5 R	CTGGAAACTGCTATGTTCCGAG
Cer1F	CTCTGGGGAAGGCAGACCTAT
Cer1R	CCACAAACAGATCCGGCTT
otx2F	TATCTAAAGCAACCGCCTTACG
otx2R	AAGTCCATACCCGAAGTGGTC
Wnt3F	CTCGCTGGCTACCCAATTTG
Wnt3R	CTTCACACCTTCTGCTACGCT
Mixl1F	ACGCAGTGCTTTCCAAACC
Mixl1R	CCCGCAAGTGGATGTCTGG
Evx1F	GAGAGCCGAAAGGACATGGTT
Evx1R	CTGCCTGCTAGTCCATCGAC
HoxA9	CCCCGACTTCAGTCCTTGC
HoxA9	GATGCACGTAGGGGTGGTG
Pax6F	GCAGATGCAAAAGTCCAGGTG
Pax6R	CAGGTTGCGAAGAACTCTGTTT
Gata4R	CCCTACCCAGCCTACATGG
Gata4F	ACATATCGAGATTGGGGTGTCT
GoosecoidF	CAGATGCTGCCCTACATGAAC
GoosecoidR	TCTGGGTACTTCGTCTCCTGG
Bmp5F	TTACTTAGGGGTATTGTGGGCT
Bmp5R	CCGTCTCTCATGGTTCCGTAG
Neurog1F	CCAGCGACACTGAGTCCTG
Neurog1R	CGGGCCATAGGTGAAGTCTT

Supplemental Experimental Procedures

Establish and culture of *Mll1*$^{-/-}$ ESCs

Mll1$^{f/f}$*; Cre-ER*^{TM} ESCs were derived from ICM of 3.5 d.p.c blastocysts after crossing *Mll1*$^{flox/flox}$ and CAGG Cre-ER^{TM} mice. *Mll1*$^{-/-}$ ESCs were generated upon 5μM 4-Hydroxytamoxifen (4-OHT) treatment for 4 Day. Primers used in genotyping for the *Mll1*$^{flox/flox}$*;* ER^{TM}-Cre and *Mll1*$^{-/-}$ ESCs were E2: 5' GCCAGTCAGTCCGAAAGTAC-3'; F2: 5'-AGGATGTTCAAAGTGCCTGC-3'; G2: 5'-GCTCTAGAACTAGTGGATCCC-3'. Primers E2 and F2 were used to identify the floxed allele (920bp for floxed, or 770 for WT) and E2 and G2 were used to detect the deleted allele (300bp). *Mll1*$^{fl/fl}$ and *Mll1*$^{-/-}$ ESCs were cultured on top of MEF feeder cells in Glasgow Modification of Eagles Medium (GMEM, Fisher11710-035) containing 15% fetal bovine serum (FBS; GIBCO, #104390924) and LIF (Millipore, ESG1107).

Establish and culture EpiSCs

EpiSC were derived from peri-implantation stage mouse embryos following the protocol as described in (Najm et al., 2011) and in (Gayen et al., submitted). Specifically, individual mouse embryos were plated on quiescent mouse embryonic fibroblast (MEF) feeder cells in K15F5 medium containing Knockout DMEM (GIBCO, #10829-018) supplemented with 15% Knockout Serum Replacement (KSR; GIBCO, #A1099201), 5% fetal bovine serum (FBS; GIBCO, #104390924), 2 mM L-glutamine (GIBCO, #25030), 1X nonessential amino acids (GIBCO, #11140-050), and 0.1 mM 2-mercaptoethanol (Sigma, #M7522). After 6 days, blastocyst outgrowths were dissociated with 0.05% trypsin (Invitrogen, #25300-054) very gently to make the partial dissociation. The partial dissociates were plated individually into a 1.9 cm^2 well containing MEF feeder cells and cultured for additional 4-6 days in K15F5 medium. The culture was then passaged by brief exposure (2–3 min) to 0.05% trypsin/EDTA with gentle pipetting to prevent complete single-cell dissociation of pluripotent clusters, and plated onto a 9.6 cm^2 well containing MEF feeders in K15F5 medium. Morphologically distinct mouse EpiSC colonies became evident over the next 4–8 days. They were continuously cultured in medium containing Knockout DMEM supplemented with 20% KSR, 2 mM Glutamax (GIBCO, #35050061), 1X nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 10 ng/ml bFGF (R&D Systems, #233-FB). EpiSCs were passaged by pipetting into small clumps in the presence of 1.5 mg/ml collagenase type IV (GIBCO, #17104-019) every third day.

Induction of epiblast like cells (EpiLCs)

Induction of EpiLC differentiation was adapted from previous protocols (Buecker et al., 2014; Schulz et al., 2014). R1, *Mll1*$^{fl/fl}$ or *Mll1*$^{-/-}$ ESCs were trypsinized and plated on 25,000cm^2 tissue

culture dish pretreated with 5 μg/ml Fibronectin (Sigma, #F1141) in N2B27-based medium containing 15% KSR (KSR; GIBCO, #A1099201), 10ng/ml bFGF (R&D Systems, #233-FB) and 20 ng/ml Activin A (R&D System, #338-AC). Successful induction of EpiSCs was confirmed by immunofluorescence using anti-REX1 (Thermo Scientific PA5-27567) and anti-NANOG (ReproCELL, RCAB002P-F) antibodies. The established EpiLCs were maintained in EpiSC medium as described above.

Generation of reverted naïve ESCs (rESCs)

EpiSCs were passaged in single cell or small clumps on MEF feeder cells as described above. For reversion, MM-401 (50-100uM final concentration) was added to culture medium immediately or 2-3 days after EpiSCs forming clones. MM-401 is replenished to reach 50-100μM concentration during each passage. rESC lines can be established by clone picking or en masse passaging. rESCs for chimera test were derived clonally. rESC line can be expanded in KSR/KO DMEM supplanted with LIF without MM401 since passage 6.

Alkaline Phosphatase Staining of ESCs

The Vector™ Alkaline Phosphatase (AP) Staining Kit was used for the detection of the AP activity according to the manufacturer's instructions.

Fluorescence *in situ* hybridization (RNA FISH).

Double-stranded RNA FISH (dsRNA FISH) were performed as described (Kalantry et al., 2009, Maclary et al., 2014). Briefly, dsRNA FISH probes were created from DNA templates by random priming using BioPrime DNA Labeling System (Invitrogen, #18094011). Probes were labeled with Fluorescein-12-dUTP (Invitrogen) and Cy3-dCTP (GE Healthcare, #PA53021). Labeled probes were precipitated in a 3M sodium acetate solution (Teknova, #S0298) along with 300 mg of yeast tRNA (Invitrogen, #15401-029), 15 mg of mouse COT-1 DNA (Invitrogen, #18440-016) and 150 mg of sheared, boiled salmon sperm DNA (Invitrogen, #15632-011). The solution was centrifuged at 15,000 rpm for 20 min at 4°C. The resulting pellet was washed with 70% ethanol, then washed with 100% ethanol, dried, and re-suspended in deionized formamide (ISC Bioexpress, #0606). Probe was denatured via incubation at 90°C for 10 min followed by immediate 5 min incubation on ice. A 2X hybridization solution consisting of 4X SSC, 20% Dextran sulfate (Millipore, #S4030), and 2.5 mg/ml purified BSA (New England Biolabs, #B9001S) was added to the denatured solution. The probe was then pre-annealed by incubation at 37°C for 1 hr. The probes were then stored at -20°C. Coverslips having the cells were dehydrated through 2 min incubations in 70%, 85%, 95%, and 100% ethanol solutions and

subsequently air-dried. The probe was hybridized overnight in a humid chamber at 37°C. The samples were washed 3 times while shaking at 39°C with 2XSSC/50% formamide, 2 times with 2X SSC, and 2 times with 1X SSC. A 1:250,000 dilution of DAPI (Invitrogen, #D21490) was added to the third 2X SSC wash. The embryos were then mounted in Vectashield (Vector Labs, #H-1200).

Immunofluorescence

After fixation by 2% paraformaldehyde for 15 min and permeabilization by 0.2% Triton X-100, the cells were washed in PBS for 3 times, 3 min each while shaking. The cover slips were blocked with PBS buffer containing 0.5 mg/mL BSA (New England Biolabs, #B9001S), 50 ug/mL yeast tRNA (Invitrogen, #15401-029), 80 units/mL RNAseOUT (Invitrogen, #10777-019), and 0.2% Tween 20 (Fisher, #BP337-100) in a humid chamber for 30 min at 37°C. The cover slips were incubated with primary antibody diluted in blocking buffer for 1 hr at 37°C. The NANOG antibody (ReproCELL, #RCAB002P-F) was used at a 1:200 dilution; The Oct4 antibody (Santa Cruz SC-5279) was used at a 1:100 dilution. The REX1 antibody (Thermo Scientific PA5-27567) was used at a 1:200 dilution. The slides were washed with PBS/0.2% Tween 20 for 3 times, 3 min each while shaking. The fluorescently-conjugated secondary antibody was used at 1:300 dilution (Alexa Fluor, Invitrogen) for 30 min at 37°C, followed by three washes with PBS/0.2% Tween-20 while shaking for 3 times, 3 min each. The cover slips was mounted with Vectashield with DAPI and processed for imaging.

MLL1 depletion by shRNAs

Lentiviral shRNAs for Mll1 (TACTCTTCTCAATCTTGCA) and scramble-shRNAs were purchased from Dahamacon, GE. The 293FT cell line was used for Lentiviral packaging. Viral transfection was carried out for EpiSCs in suspension at MOI=3 in the presence of polybrene (0.5 μg/ml final concentration, Sigma). The shRNAs were in the TurboRFP expression vectors. The successful transfection can be visualized by RFP fluorescence.

ChIP-seq analyses

H3K4me1 and H3K4me3 ChIP-seq data for ESCs (GSE47949) or EpiSCs (GSE57407) were used for analyses together with ChIP-seq for MLL1, H3K4me1 and H3K4me3 in EpiSCs treated with MM401. ChIP-Seq reads were aligned to the genome assembly (UCSC mm9, http://support.illumina.com/sequencing/sequencing_software/igenome.html) using Bowtie2 (Langmead and Salzberg, 2012). Unique reads (with XS tag in SAM file) that were mapped to a single genomic location were kept and transformed to BAM files for peak identification using

MACS (Zhang et al., 2008). WIG files generated by MACS were used for visualization in UCSC genome browser. HOMER (http://homer.salk.edu) was used for peak annotation as well as calculating tag densities (annotatePeaks.pl) around peak centers, enhancers or TSS. Enhancers were defined by combining H3K4me1 peaks in ESCs (123,336 peaks) and EpiSCs (83,900 peaks) as previously described (Lara-Astiaso et al., 2014). Peak with highest reads intensity was selected if distance between two peaks is shorter than 500bp. Using this approach, 103,748 H3K4me1-marked enhancers were identified and referred to in our study. TSS sites were retrieved from the UCSC website.

ChIP-fragment depth analyses in Supplemental Figure 6C were performed using HOMER. Specifically, reads count was generated for all MLL1 peaks using unique mappers in each sample. Reads within 4kb region of the MLL1 peak center were counted using 100bp bin and normalized against base pair length and peak numbers in addition to the standard normalization against 10 million tags. For clustering analysis, tag densities matrix generated by HOMER was loaded into MEV software (Saeed et al., 2006) and hierarchical clustering was performed using Spearman rank correlation. For generating heat-map, tag densities in regions of enhancer center ±5kb or TSS±5kb were calculated using 100bp bin.

Gene expression analyses by RNA-seq

RNA was extracted by Trizol reagent (Ambion) and further purified by RNeasy Mini kit (QIAGEN) following manufacture's protocol. 10ng of total RNAs were used for preparation of Illumina sequencing library. RNA sequencing was performed on Illumina HiSeq2000 at University of Michigan core facility. Sequenced reads were aligned to mouse reference genome (mm9) using Bowtie and Tophat (version 2.0.3) (Langmead et al., 2009; Trapnell et al., 2009). Mapping results (BAM file) were transformed to SAM files using SAMtools (Heinz et al., 2010). Uniquely mapped reads were extracted using the NH tag. Reads of all annotated genes were counted using featureCounts (Liao et al., 2014) and differential gene expression analysis was performed using DESeq (Anders and Huber, 2010). Fold changes obtained from DESeq were used for subsequent analysis. To measure gene expression level, RPKM (Reads Per Kilobase of exon per Million mapped reads) of genes were normalized against total uniquely mapped reads and total exon length. If different isoforms exist, then all exons are merged before analyses. For Pearson correlation coefficients, LOG_{10} transformed RPKM values of all the selected genes were used as input, and Pearson correlation coefficients between each two samples were calculated using function cor() in R. The matrix of Pearson correlation coefficients was used to generate heat map using MEV software.

Network generation and visualization

Known gene association in 'Mus musculus' for Group II genes was established using GeneMANIA software (genemania.org/). Connections based on physical interactions were established using source file from iRefIndex (irefindex.org) and BioGRID (thebiogrid.org/). Connections based on functional association were established using source file from Mouse Genome Database (MGD, http://www.informatics.jax.org/). Connections based on disease associations were established using source file from OMIM (Hamosh et al., 2005). Network visualization was further refined using Cytoscape (v3.2.1) (Shannon et al., 2003). Sub-cellular localization, for grouping nodes in the network, was determined using information from LOCATE (Fink et al., 2006), Uniprot (UniProt, 2015) and Human Protein Atlas (Uhlen et al., 2015) databases. Node size was set to reflect node degree in the network by Network analyzer module. GO/Pathway annotations were viewed by Panther Functional classification. Genes categorized in biological adhesion (GO:0022610), developmental process (GO:0032502), Wnt signaling pathway (P00057), TGF-beta signaling pathway (P00052) and Cadherin signaling pathway (P00012) were highlight and distinguished by border colors.

Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell stem cell *14*, 838-853.

Factor, D.C., Corradin, O., Zentner, G.E., Saiakhova, A., Song, L., Chenoweth, J.G., McKay, R.D., Crawford, G.E., Scacheri, P.C., and Tesar, P.J. (2014). Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. Cell stem cell *14*, 854-863.

Fink, J.L., Aturaliya, R.N., Davis, M.J., Zhang, F., Hanson, K., Teasdale, M.S., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y.*, et al.* (2006). LOCATE: a mouse protein subcellular localization database. Nucleic acids research *34*, D213-217.

Gayen, S., Maclary, E., Buttigieg, E., Hinten, M., and Kalantry, S. (submitted). The Tsix long non-coding RNA maintains random X-chromosome inactivation patterns.

Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., and McKusick, V.A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic acids research *33*, D514-517.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Kojima, Y., Kaufman-Francis, K., Studdert, J.B., Steiner, K.A., Power, M.D., Loebel, D.A., Jones, V., Hor, A., de Alencastro, G., Logan, G.J.*, et al.* (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell stem cell *14*, 107-120.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S.*, et al.* (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943-949.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Najm, F.J., Chenoweth, J.G., Anderson, P.D., Nadeau, J.H., Redline, R.W., McKay, R.D., and Tesar, P.J. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. Cell stem cell *8*, 318-325.

Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A., and Quackenbush, J. (2006). TM4 microarray software suite. Methods in enzymology *411*, 134-193.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Bluthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell stem cell *14*, 203-216.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research *13*, 2498-2504.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A.*, et al.* (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

UniProt, C. (2015). UniProt: a hub for protein information. Nucleic acids research *43*, D204-212.
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.