Supporting Information

Wang *et al*. 10.1073/pnas.0708951105

SI Experimental Procedures

Targeting Constructs and Screening for Positive ESCs. The *Cdx4* knockout parental vector (plox-Cdx4) was a kind gift from Dr. Brian Huntly, in which a 4-kb fragment containing *Cdx4* genomic DNA 40 bp downstream of the first exon was used as 3′-arm of homologous recombination. We made the following modifications to this vector: (*i*) the 3′arm was extended to 5.8 kb by cloning a 1.8-kb fragment of *Cdx4* amplified from genomic DNA into plox-Cdx4; (*ii*) a 2-kb fragment upstream ATG codon of *Cdx4* was amplified from mouse genomic DNA was used as 5′ arm of homologous recombination. A GFP cDNA was fused in frame with ATG of *Cdx4* to express GFP under *Cdx4* promoter. Cdx4-GFP fragment was cloned into KpnI/ApaI digested plox-Cdx4 from step *i*; (*iii*) a DTA cDNA driven by PGK promoter was inserted after the 3′arm of *Cdx4* disruption construct.

Two parental ESC lines were used, one is V6.5 (a kind gift from Dr. Rudolf Jaenisch, F1 ESCs from 129xC57bl/6), another is tetracycline inducible *Cdx4* ESCs (icdx4), in which a *Cdx4* cDNA was inserted into a tetracycline inducible locus flanked by loxP sites (1). Both ESC lines are from male embryos (40+XY). The genomic locus of *Cdx4* is at X-chromosome.

ESCs were electroporated with NotI linearized targeting vector, and treated with 140 μg/ml hygromycin B at 24 h after electroporation. Hygromycin B resistant ESC clones were screened for single correct insertion of GFP cDNA by Southern blot analysis. Two correct icdx4-derived and four V6.5-derived ESC lines were obtained after screening. Because the *Cdx4* was located at X-chromosome, single correct insertion of GFP in these ESC lines resulted in a total disruption of the *Cdx4* genomic locus ($Cdx4^{GFP/y}$). These ESC clones were subjected to second round of electroporation with either CRE or Fle expression vectors with puromycin-resistant gene to excise the CMV-hygromycin-GFP to prevent the formation of truncated *Cdx4*. ESCs were treated with 2 μg/ml puromycin 24 h after electroporation for 2 days only, and individual clones were then selected and tested for hygromycin sensitivity and GFP^{-}. Correct clones were further confirmed by Southern blot analysis and PCR to confirm the complete excision of the CMV-hygromycin-GFP. Because *Cdx4* cDNA was flanked by loxP sites in icdx4-derived clones, Fle-excision would only loop out CMV-hygromycin with the *Cdx4* at the inducible locus intact. Therefore, the expression of Cdx4 could be induced with doxycycline treatment. All PCR primer and PCR conditions used to make the targeting vector are available upon request. In summary, following are the *Cdx4* knockout ESC lines used in this study:

KO1, KO2: icdx4-derived, contain no endogenous or inducible *Cdx4*

KO3: icdx4-derived; with endogenous *Cdx4* locus disrupted, but contains inducible *Cdx4* locus.

KO4: V6.5-derived

Genomic DNA Isolation and Southern Blot Analysis. Genomic DNA was isolated according to the manufacturer's protocol (Gentra Systems) and Southern blot analysis was performed according to standard procedures. *Cdx4* 5′ and 3′probes were PCR products amplified from mouse genomic DNA outside of the homologous region. Hygromycin probe was amplified from the parental vector plox-cdx4. Probe labeling and DNA hybridization were performed according to standard protocol and manufacture's recommendation (Roche, DIG labeling kit and DIG Wash and Block Buffer Set). The first probe was stripped from membrane by washing twice in striping solution (0.2M NaOH and 0.1% SDS), and rinsed with 2xSSC before hybridization with next probe. All PCR primers and PCR conditions used to make these probes were available upon request.

Establishing and Genotyping *Cdx4* Knockout/GFP Knockin Mice. *Cdx4* knockout/GFP knockin mice were generated from four V6.5-derived *Cdx4* knockout ESC lines by traditional blastocyst injection (BALB/c as blastocyst donor) or standard tetraploid complementation (with CD1 as tetraploid blastocyst donor). Chimeric mice generated from traditional blastocyst injection were crossed back to BALB/c to check for germline transmission. The *Cdx4* heterozygote ($Cdx4^{GFP/+}$), homozygote ($Cdx4^{GFP/GFP}$) and hemizygote ($Cdx4^{GFP/y}$) mice were obtained by crossing the founder mice with C57BL/6. *Cdx4* genotyping used similar protocol for *Cdx1* described in the text (experimental procedures). Genotyping primers: Cdx4-F: 5′-TGACATGACCT CCCCAGTTTTCGGATC-3; AS3: 5′- TACCCTTACT CCCTTGACTTATCTGG-3′; GFP-F: 5′-TCATCTGCACCACCGGCAA-3′; GFP-R: 5′-GTTGTAGTTG-TACTCCAGCT-3′, wild-type and mutant allele gave a 500bp and a 300bp band respectively. All experimental procedures were approved by the Animal Care and Use Committee of Children's Hospital Boston, and conducted in accordance with the Animal Welfare Act and Public Health Service Policy.

Microarray Hybridization and Data Processing. Total RNA was extracted as described above from $Cdx2^{-/-}$ and $Cdx2^{+/+}$ day 6 EBs in three independent experiments to get biological triplicates. Gene expression analysis was conducted using Agilent Mouse Genome 4 × 44K arrays according to the manufacturer's protocol (Agilent Technologies). Briefly, 500 ng of total RNA from each sample was amplified and labeled with Cy3 using the Agilent Low RNA Input Fluorescent Linear Amplification Kit. 1.5 μg of Cy3-labeled cDNAs were then fragmented and hybridized to the slide for 17 h in a rotating hybridization oven at 65°C using the Agilent *In Situ* Hybridization Kit. Slides were then washed and scanned with an Agilent Scanner. Data were finally obtained with the Agilent Feature Extraction software (v9.1), using defaults for all parameters, and then deposited into Rosetta Resolver (Rosetta Biosoftware). The resultant profiles were combined into ratio experiments. Intensity plots were generated for each ratio experiment and gene alteration was considered statistically significant if the *P* value was less than 0.001. The function and network analysis was performed by using knowledge-based network software, Ingenuity Pathway Analysis (Ingenuity Systems).

RNA Interference. Lentivirus-based shRNA against luciferase, *Cdx1* or *Cdx2* were generated as described in Fig. S4A. To select hygromycin-resistant ESCs, 140 μg/ml hygromycin B were used 24 h after viral infection for seven days. EBs made from hygromycin B-resistant ESCs were treated continuously with 80 μg/ml hygromycin B through EB development. Hygromycin-negative ESCs were tested under same conditions and no EBs could be formed with 50 μg/ml hygromycin B treatment. In the case of combined gene knockdown, ESCs were infected sequentially with viral supernant containing shRNA against each target genes.

1. Wang Y, Yates F, Naveiras O, Ernst P, Daley GQ (2005) Embryonic stem cell-derived hematopoietic stem cells. *Proc Natl Acad Sci USA* 102:19081–19086.

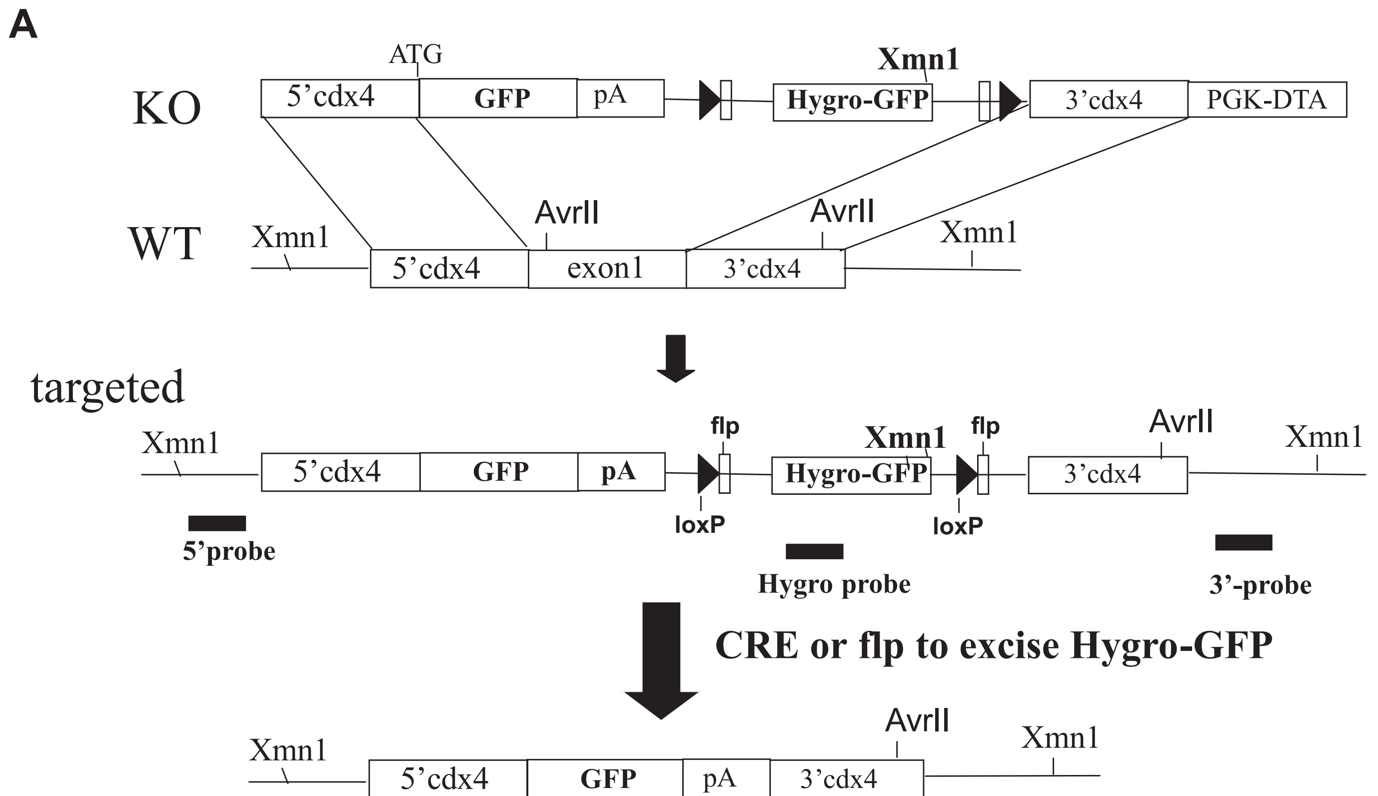


Fig. S1. Establishment and characterization of *Cdx4* knockout ESCs and mice. (*A* and *B*) *Cdx4* gene disruption construct, targeting strategy, and screening of ESC clones. The entire first exon of *Cdx4* gene was replaced by a GFP cDNA, a poly(A) adenylation signal (to prevent the formation of truncated *Cdx4* protein) and a CMV-driven hygromycin-GFP (Hygro-GFP) fusion gene, allowing positive selection of hygromycin B-resistant ESCs. Hygro-GFP was flanked by loxP and flp sites, which allowed Hygro-GFP to be excised by cre or fle. PGK-DTA (diphtheria toxin) would be deleted upon homologous recombination to enable negative selection of ESCs with random integration of the targeting construct. Correctly targeted ESC clones after hygromycin B selection could be distinguished by the size difference of the XmnI fragment from wild-type controls, detected by Southern blot (the upper two panels of *B*). The ESC clones with a correct and single integration were electroporated again with a Cre-puromycin or a fle-puromycin expressing vector to excise the CMV-hygromycin-*GFP* gene to prevent the formation of truncated *Cdx4* from CMV promoter. GFP^{-} and puromycin-resistant ESC clones were selected and confirmed by Southern blot analysis after XmnI/AvrII digestion of genomic DNA (of *B*). (*C*) The expression of *Cdx4* and *GFP* during EB differentiation in $Cdx4^{+/y}$ and/or $Cdx4^{GFP/y}$ ESCs was examined by real-time RT-PCR. *Cdx4* transcripts were not detectable by real-time RT-PCR in *Cdx4*-deficient ($Cdx4^{GFP/y}$) ESCs. GFP expression in *Cdx4* knockout/*GFP* knockin ESCs mimicked that of endogenous Cdx4 during EB development. (*D*) GFP expression in *Cdx4* knockout/*GFP* knockin mice mirrored that of endogenous Cdx4 during embryonic development. *GFP* expression was first detected at the base of the allantois at 7 dpc and then appeared in the caudal region of the embryo after 8.5 dpc, finally disappearing after 10 dpc. (*E*) Flow cytometry analysis of surface antigens on EB-derived cells collected at different time points. Cells expressing these markers were present at similar levels and time points in $Cdx4^{GFP/y}$ (KO1, KO3, and KO4) and $Cdx4^{+/y}$ EBs. Error bars represent one standard deviation calculated from duplicate samples. Results were reproduced from two independent experiments. (*F–H*) Hematopoietic colony formation from yolk sac (YS) at 7.5 dpc and 9 dpc (*F*), fetal liver at 14.5 dpc (*G*) or bone marrow (*H*) of *Cdx4* knockout ($Cdx4^{GFP/y}$ and $Cdx4^{GFP/GFP}$) mice compared to wild-type controls. Error bars represent one standard deviation. *P* values were calculated with Student’s *t* test. (*G* and *H*) *P* value >0.05 for all comparisons, suggesting that no statistically significant differences were observed.

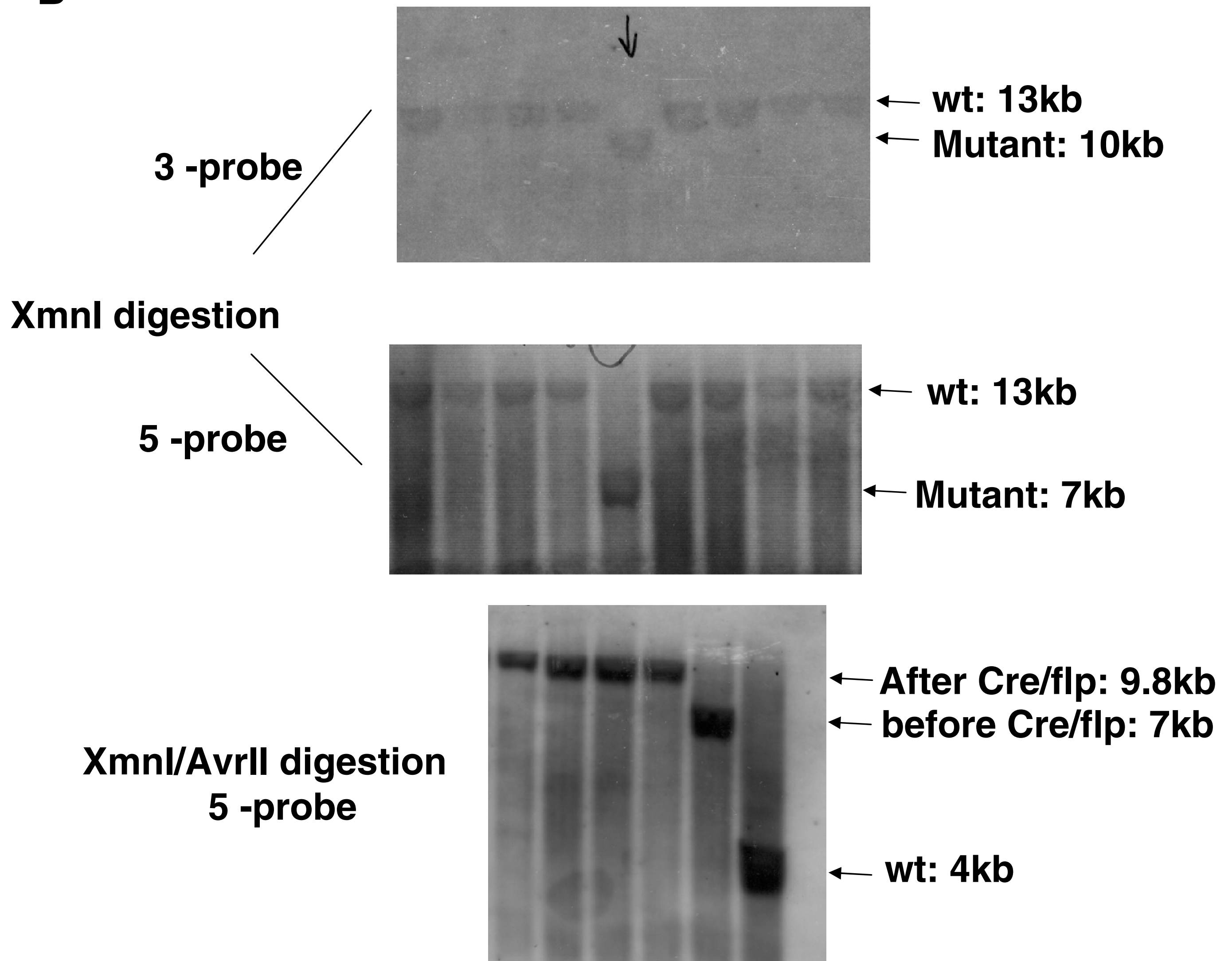


Fig. S1B. *Fig. S1 continued.*

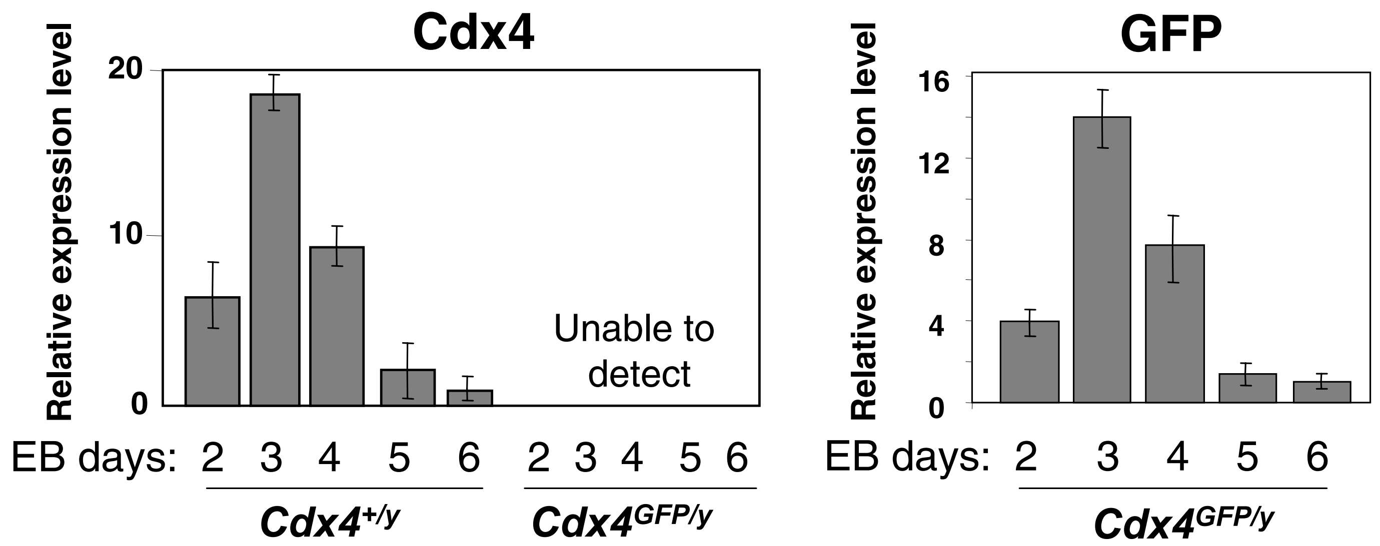
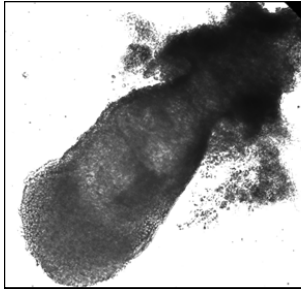
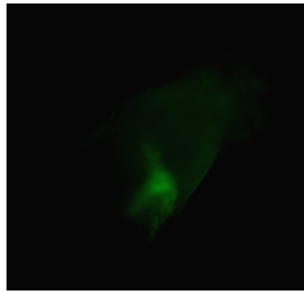


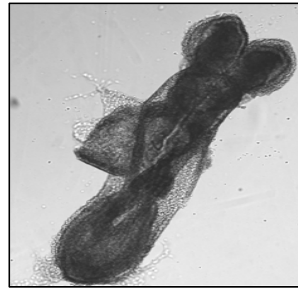
Fig. S1C. *Fig. S1 continued.*

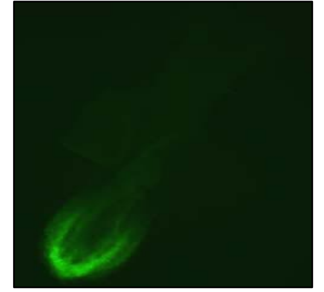
D



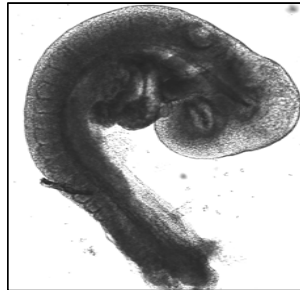


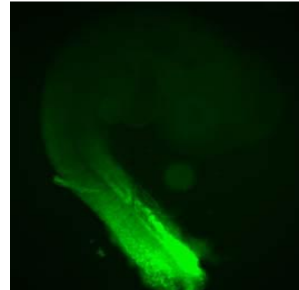
Day7.5





Day8.5





Day9

Fig. S1D. *Fig. S1 continued.*

E

Flk1, day 4 EB

% of positive cells

45
40
35
30
25
20
15
10
5
0

Cdx4$^{+/y}$ KO1 KO3 KO4

Cdx4$^{GFP/y}$

CD41, day 6 EB

% of positive cells

45
40
35
30
25
20
15
10
5
0

Cdx4$^{+/y}$ KO1 KO3 KO4

Cdx4$^{GFP/y}$

CD41/c-kit, day 6 EB

% of positive cells

10
8
6
4
2
0

Cdx4$^{+/y}$ KO1 KO3 KO4

Cdx4$^{GFP/y}$

CD45, day 8 EB

% of positive cells

30
25
20
15
10
5
0

Cdx4$^{+/y}$ KO1 KO3 KO4

Cdx4$^{GFP/y}$

Fig. S1E. *Fig. S1 continued.*

F

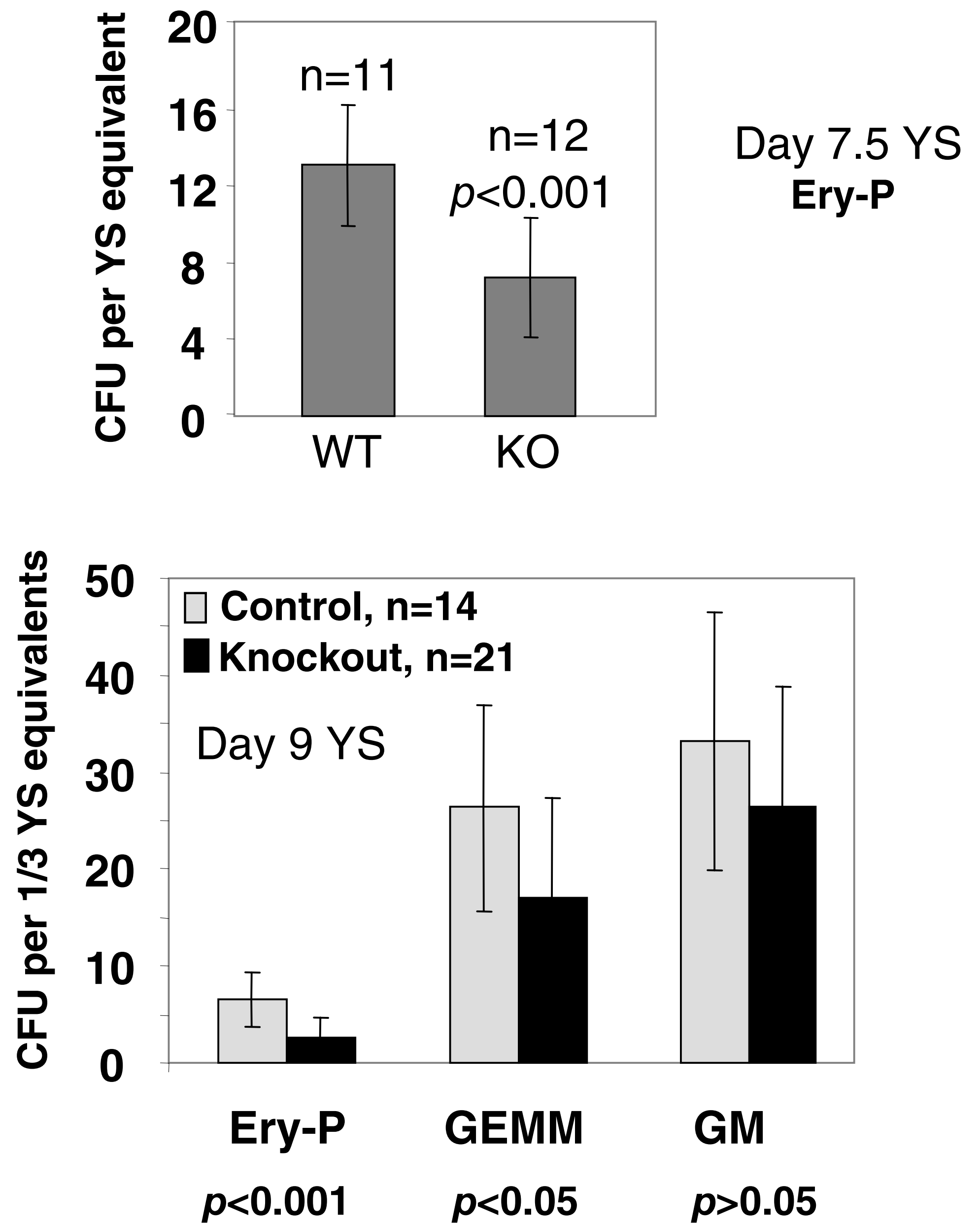
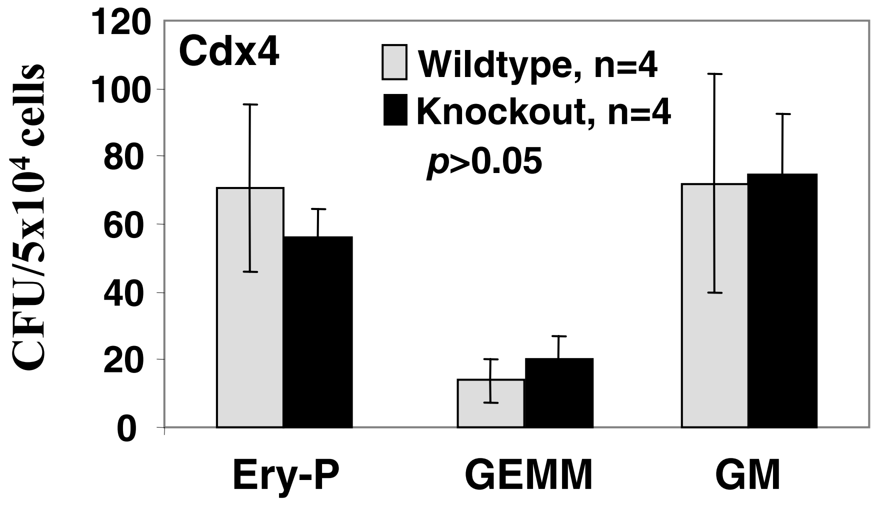


Fig. S1F. *Fig. S1 continued.*

G



Day 14.5 Fetal liver

Fig. S1G. *Fig. S1 continued.*

H

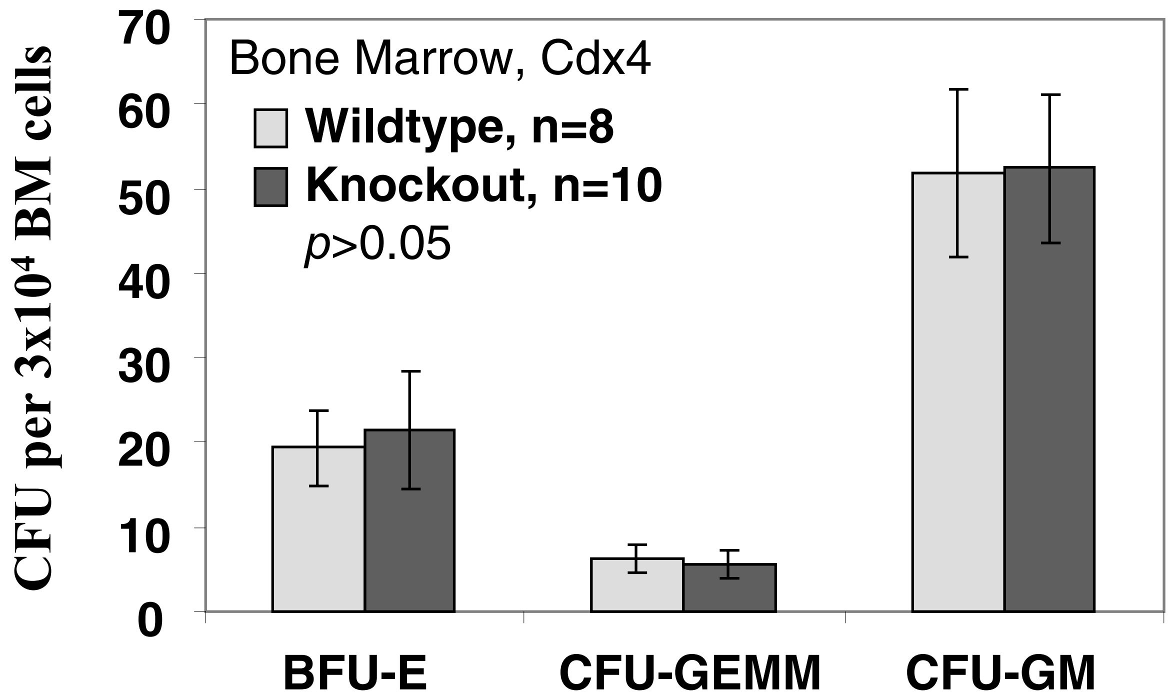


Fig. S1H. *Fig. S1 continued.*

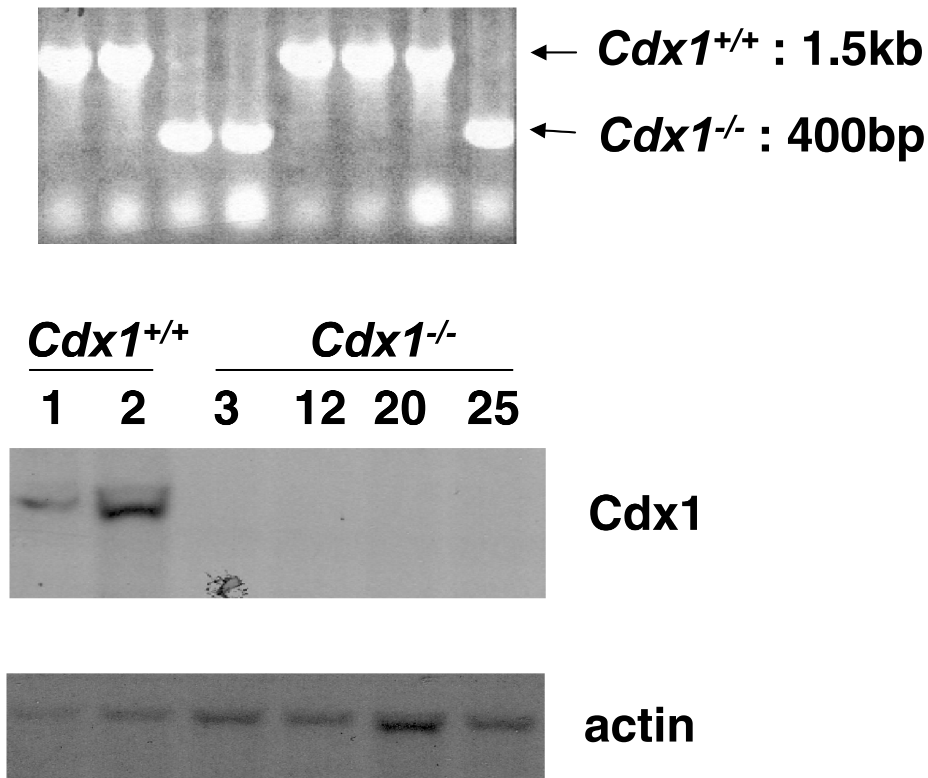


Fig. S2. Characterization *Cdx1*$^{-/-}$ ESCs and mouse model. (*A*) *Cdx1*$^{-/-}$ and *Cdx1*$^{+/+}$ ESC lines were confirmed by PCR (*Upper*, genomic DNA isolated from day 6 EBs) and Western blot (*Lower*, absence of Cdx1 in *Cdx1*$^{-/-}$ day 4 EBs). (*B*) Hematopoietic differentiation from *Cdx1*$^{-/-}$ day 6 EBs was normal, as determined by flow cytometry analysis of surface antigens that recognize hematopoietic cells. *Cdx1*$^{+/+}$: WT; *Cdx1*$^{-/-}$: KO. (*C*) Expression levels of *Hox* genes in day 6 *Cdx1*$^{-/-}$ EBs, relative to wild-type control (WT), as measured with real-time RT-PCR. (*B* and *C*) Error bars represent one standard deviation calculated from triplicates in a representative experiment. Results were reproduced from two independent experiments. (*D–F*) No significant difference in hematopoietic colony formation from yolk sac at 7.5–9.5 dpc (*D*), fetal liver at 14.5 dpc (*E*), or bone marrow (*F*) of *Cdx1* heterozygote (*Cdx1*$^{-/+}$) or homozygote (*Cdx1*$^{-/-}$) null mice from wild-type controls (*Cdx1*$^{+/+}$). Error bars represent one standard deviation. *P* values were calculated with Student's *t* test. *P* value $>$0.05 for all comparison.

B

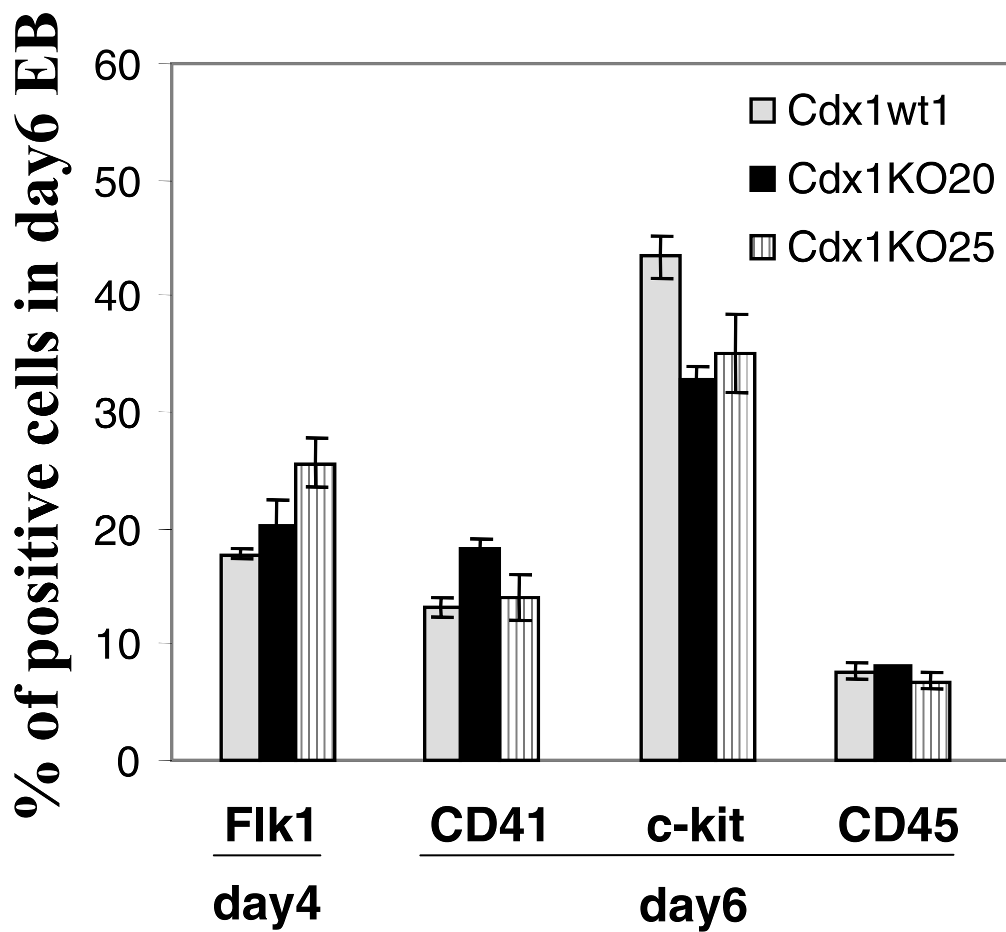


Fig. S2B. *Fig. S2 continued.*

C

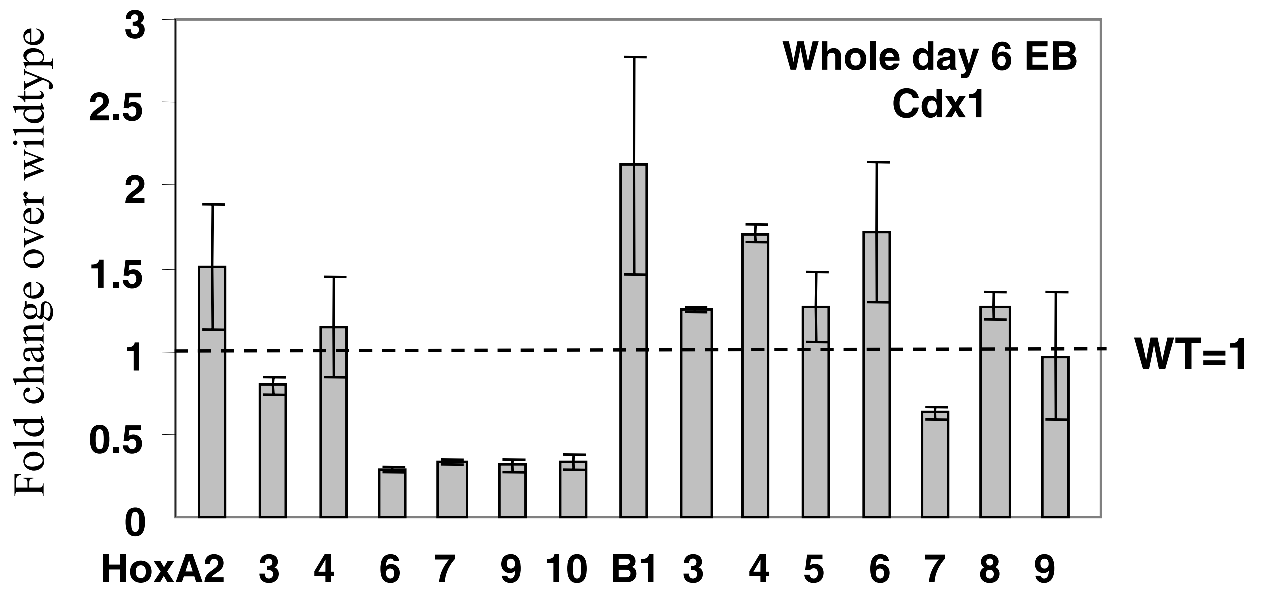
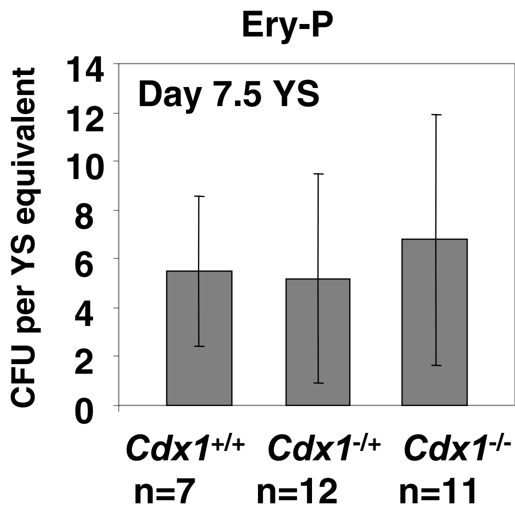
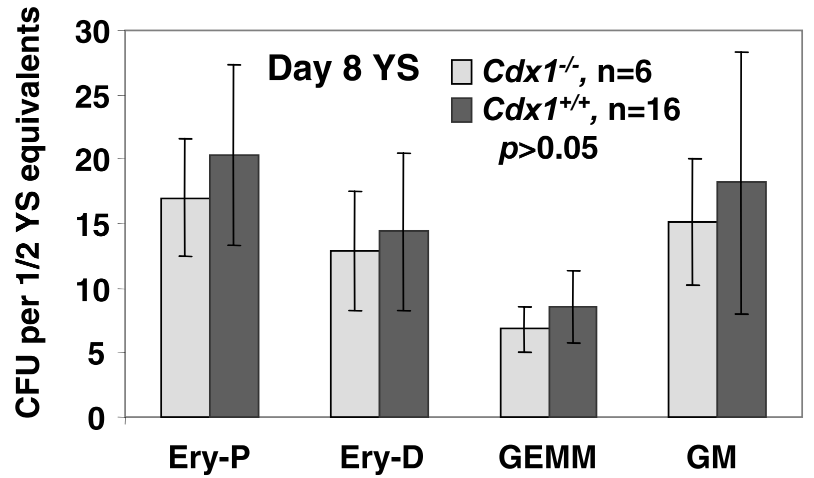


Fig. S2C. *Fig. S2 continued.*

D





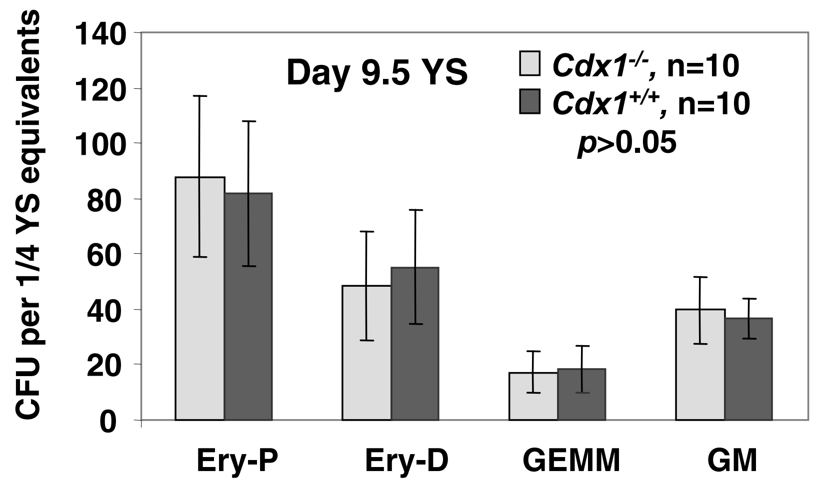


Fig. S2D. *Fig. S2 continued.*

E

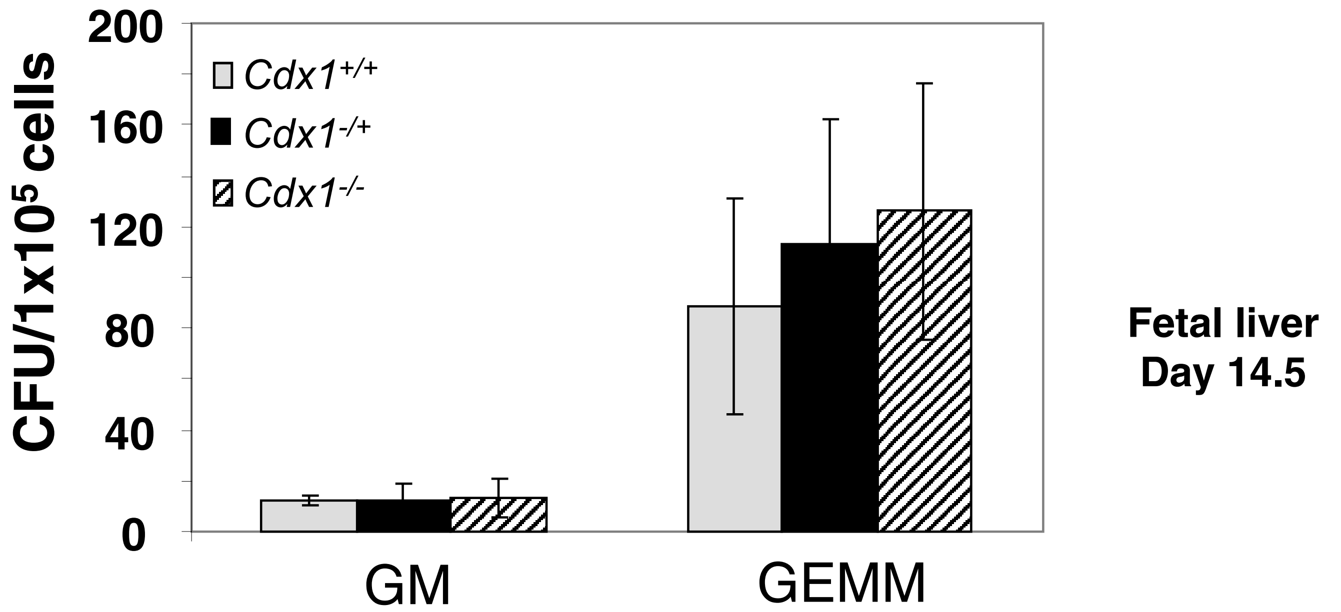


Fig. S2E. *Fig. S2 continued.*

F

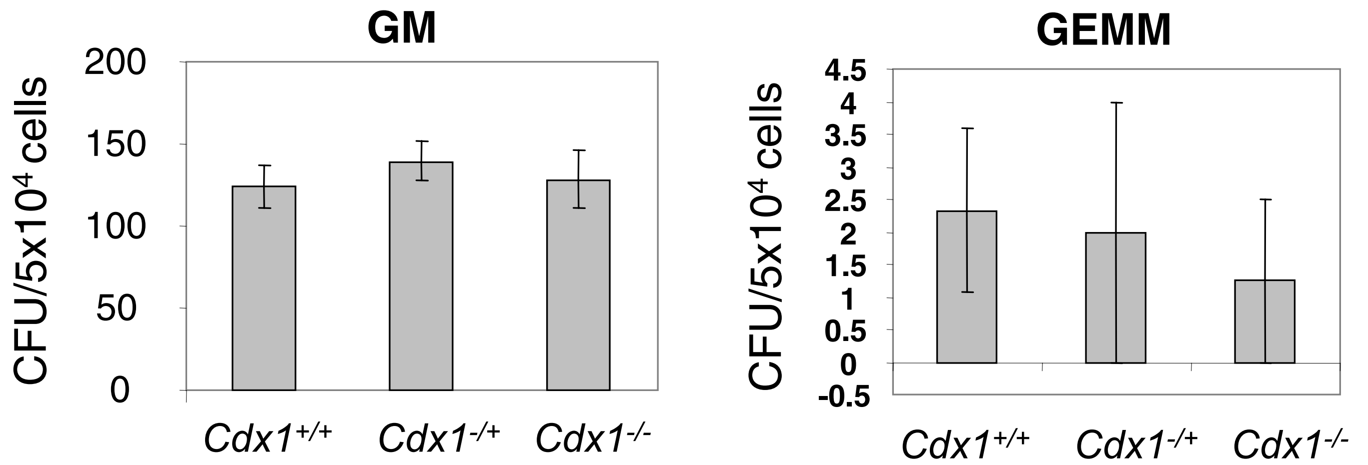


Fig. S2F. *Fig. S2 continued.*

A

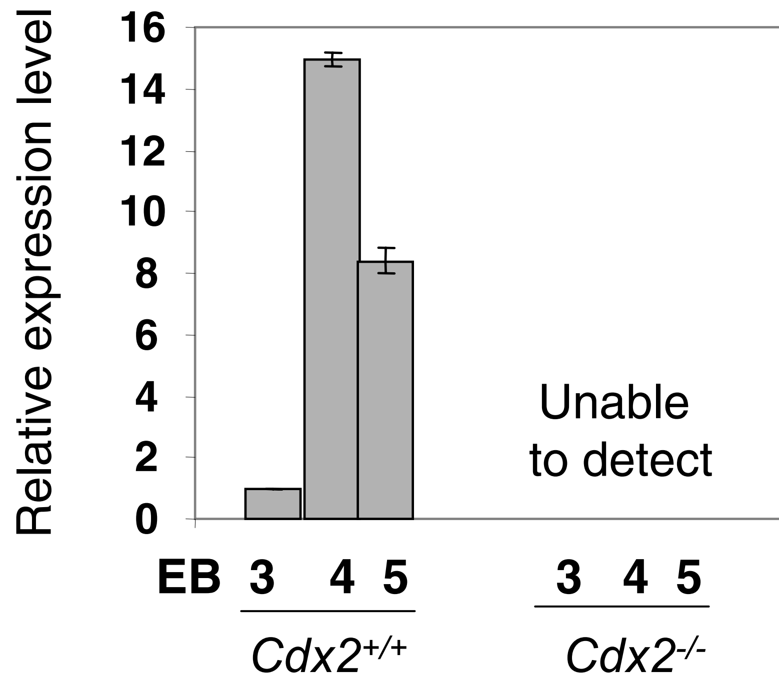


Fig. S3. Characterization *Cdx2*$^{-/-}$ ESCs. (*A*) Expression of *Cdx2* during EB development in *Cdx2*$^{-/-}$ and *Cdx2*$^{+/+}$ ESCs, analyzed by real-time RT-PCR. *Cdx2* expression is under detection in *Cdx2*$^{-/-}$ cells during EB development. (*B*) Representative graphs of CD45 staining on day 8 EBs from *Cdx2*$^{+/+}$ or *Cdx2*$^{-/-}$ ESCs. The percentage of CD45^{+} cells was significantly reduced in *Cdx2*$^{-/-}$ EBs. (*C*) Progenitor colony activity of yolk sac from chimeric embryos at 9 dpc generated with either *Cdx2*$^{-/-}$ or *Cdx2*$^{+/+}$ESCs. Donor blastocysts are lacZ^{+}. The percentage of lacZ^{-} (ESC-contributed)/total colonies was measured. Data represent average ± one standard deviation. *P* value was calculated with Student's *t* test: * *P* values $<$0.001; ** *P* values $<$0.01. (*D*) Correlations of microarray analyses among biological triplicates. Rosetta Resolver was used to visualize the correlation among biological triplicate arrays. All weighted correlation coefficients among these triplicates were higher than 0.98, which indicated good similarity among triplicate samples. A representative graph of two arrays of triplicate *Cdx2*$^{-/-}$ samples (KO1) were visualized in a correlation plot. These two biological replicates are highly correlated, with a weighted correlation coefficient of 0.98844. (*E*) Two *Cdx2*$^{-/-}$ EBs (KO1 and KO2) shared more similarity than either with wild-type control, demonstrating a valid biological correlation between these two different *Cdx2*$^{-/-}$ ESC lines. Clustering analysis using Rosetta Resolver was run for genes with *p*-values lower than 0.001 in both KO1 vs. WT and KO2 vs. WT comparisons. A divisive algorithm was obtained with a Euclidean distance metric type. Green indicates the decreased expression, and red represents the up-regulated expression upon *Cdx2* deficiency. The darkness of the color indicates the magnitude of change. The expression of a total of 6,239 sequences was statistically significantly different upon *Cdx2* deficiency as compared with wild-type control. (*F*) Network interaction of JAK, STAT, and genes involved in hematopoietic development, generated by IPA analysis. Blue lines indicate the components of the canonical Jak/Stat signaling pathway. (*G*) Canonical signal transduction pathway of sonic hedgehog in transcription regulation. The subcellular locations of the components along the pathway were shown. In *E* and *F*, solid and dashed lines represent valid direct and indirect interactions between factors respectively. Green color of the factor means decreased expression, and red color represents up-regulated expression upon *Cdx2* deficiency. The darkness of the color indicates the magnitude of change.

B

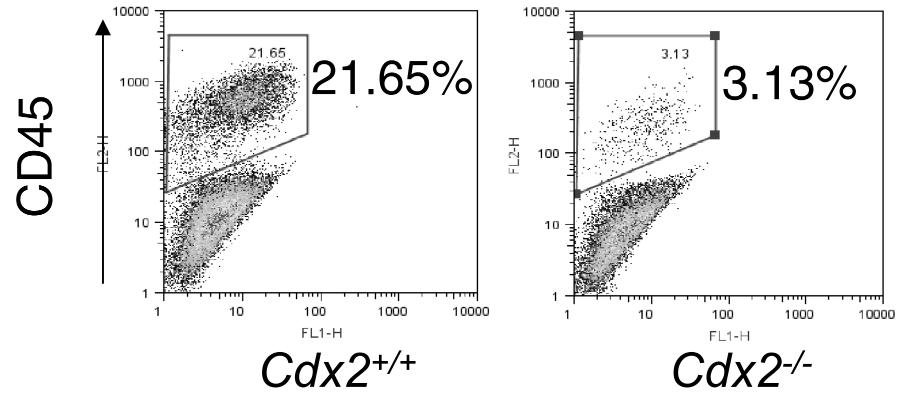


Fig. S3B. *Fig. S3 continued.*

C

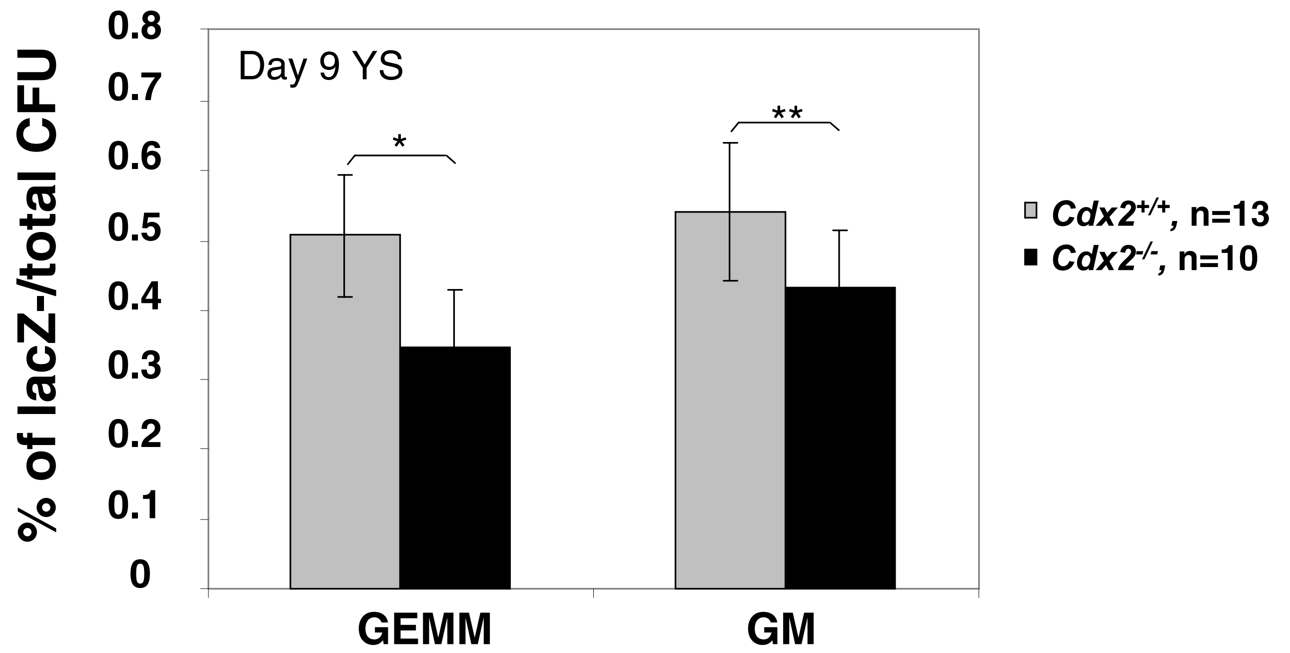
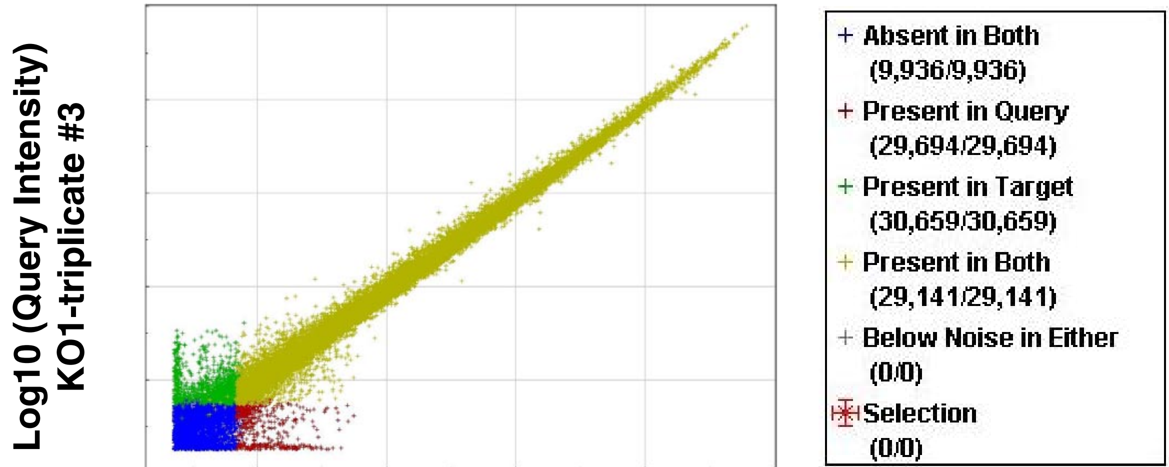


Fig. S3C. *Fig. S3 continued.*

D



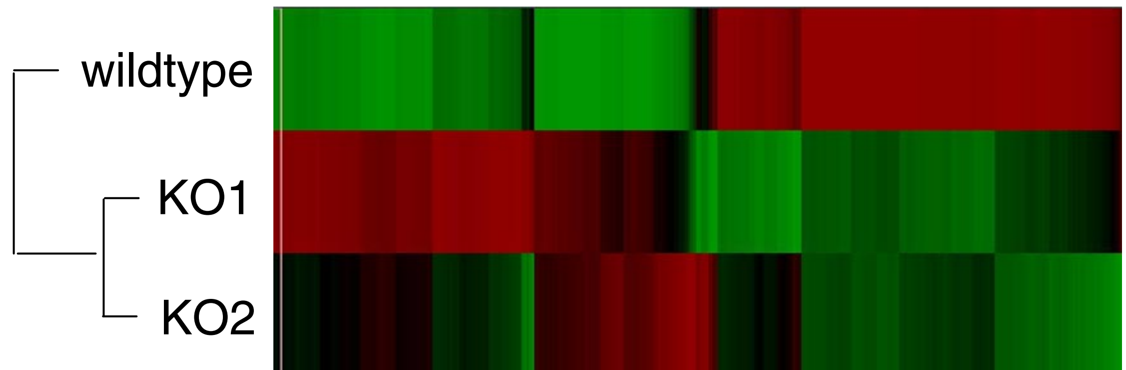
Log10 (Query Intensity)
KO1-triplicate #2

Correlation Coefficients	Non-Weighted	Weighted
Visible	0.99337	0.98844
Selected	N/A	N/A
Present in Both	0.99330	0.98969
Present in Either	0.99332	0.98160
All (no fails/controls)	0.99337	0.98844

P<=0.01

Fig. S3D. *Fig. S3 continued.*

E



6239 sequence sets

Fig. S3E. *Fig. S3 continued.*

F

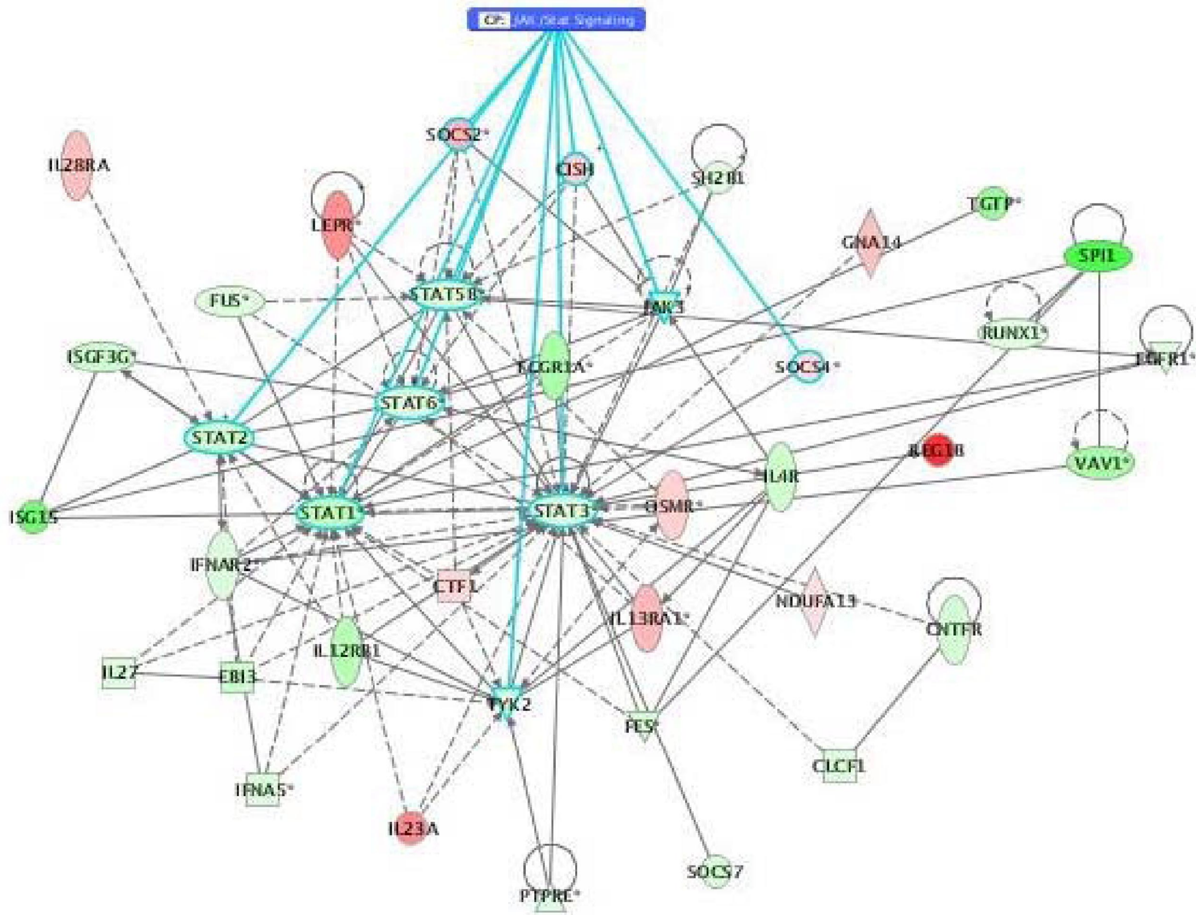


Fig. S3F. *Fig. S3 continued.*

G

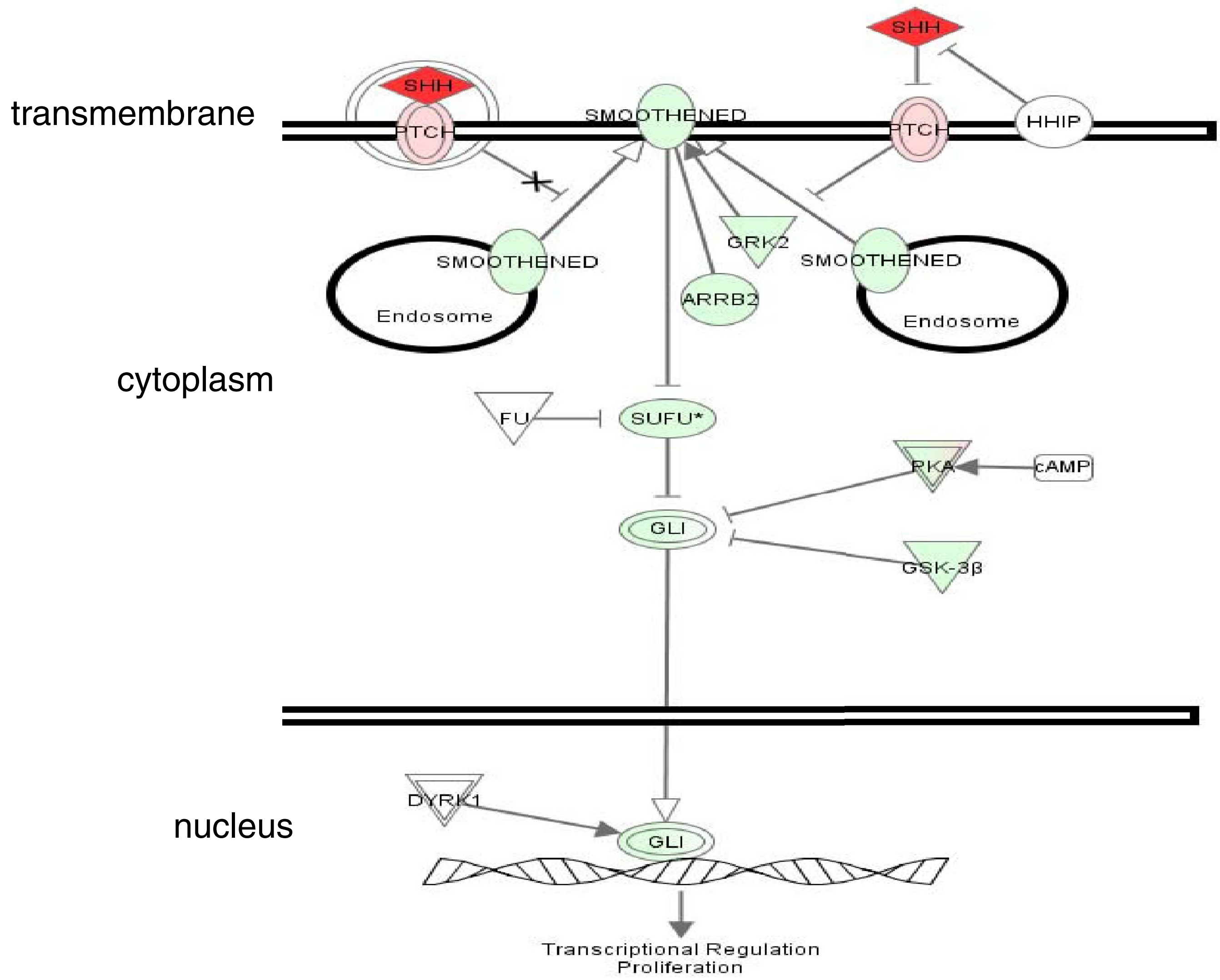
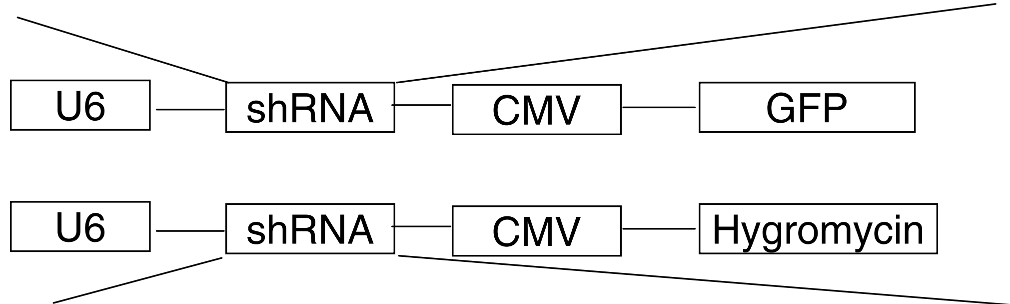


Fig. S3G. *Fig. S3 continued.*

A

Cdx1 (X1): T**GTCTCAATCACCTGGAGAT**TTCAAGAGA**ATCTCCAGGTGATTGAGAC**TTTTTTC



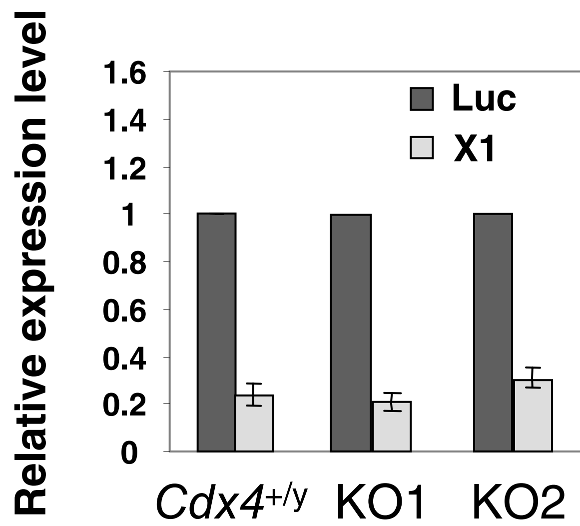
Luciferase (Luc): T**GAGCTGTTTCTGAGGAGCC**TTCAAGAGA**GGCTCCTCAGAAACAGCTC**TTTTTTC

Cdx2 (X2): T**GTATGTCTGTGTTGTAAAT**TTCAAGAGA**ATTTACAACACAGACATAC**TTTTTTC

Sense Antisense

Fig. S4. Functional redundancy among Cdx family members. (*A*) ShRNA constructs. The major difference between the two parental lentiviral vectors used in this experiment was the selection gene driven by CMV, one was GFP, and the other was hygromycin-resistant gene. A shRNA against *Cdx1* was inserted into the vector with CMV-GFP, and EBs were made directly from GFP^+ cells sorted from ESCs infected with this construct. The shRNA against *Cdx2* was constructed into lentiviral vector containing CMV-hygromycin-resistant gene. The cells expressing shRNA against *Cdx2* were selected by hygromycin treatment. Control vectors contained the shRNA against luciferase. (*B*) Hematopoietic colony formation upon combined deficiency of *Cdx1* and *Cdx4*. (*Upper*) Expression of *Cdx1* was analyzed by real-time RT-PCR on day 4 EBs upon gene knockdown against *Cdx1* (X1) as compared with control shRNA against luciferase gene (Luc). (*Lower*) Hematopoietic colony formation from day 6 EBs upon gene knockdown of *Cdx1* (X1) or luciferase (Luc). ESCs expressing the shRNA against *Cdx1* or luciferase were selected by sorting for GFP^+ cells. $Cdx4^{GFP/y}$ ESCs lines: KO1 and KO2. (*C–E*) Hematopoietic colony formation from day 6 EBs upon gene knockdown of *Cdx2* (X2), *Cdx1 and Cdx2* (X1+X2), or luciferase (Luc) in $Cdx4^{+/y}$ or $Cdx4^{GFP/y}$ (KO3) ESCs. Cells were treated with or without (−) doxycycline during different time periods of EB development (as indicated below *x* axis). Data are represented as the averaged fold change in cfu relative to control (value = 1) from duplicates or triplicates in four independent experiments. Each data point denotes five to eight biological replicates as total. Error bars represent ± one standard deviation. *P* value was calculated with Student's *t* test. **P* value <0.05; ***P* value >0.05.

B



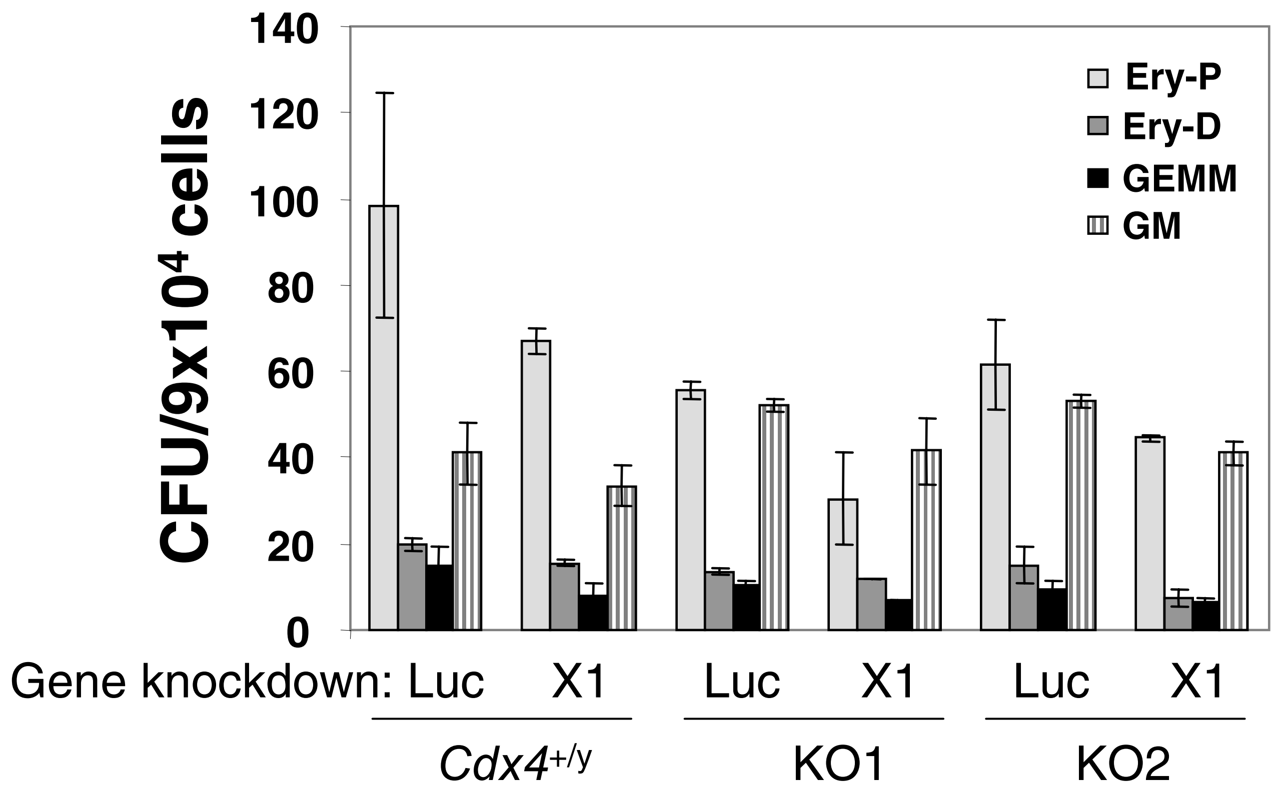


Fig. S4B. *Fig. S4 continued.*

C

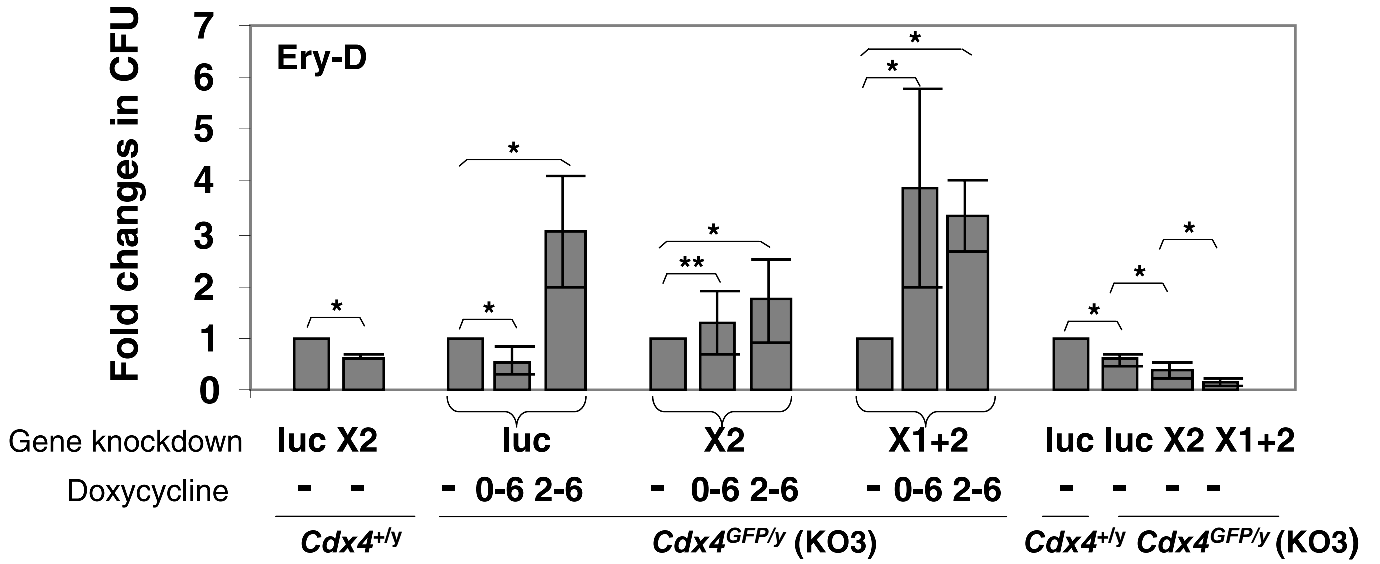


Fig. S4C. *Fig. S4 continued.*

D

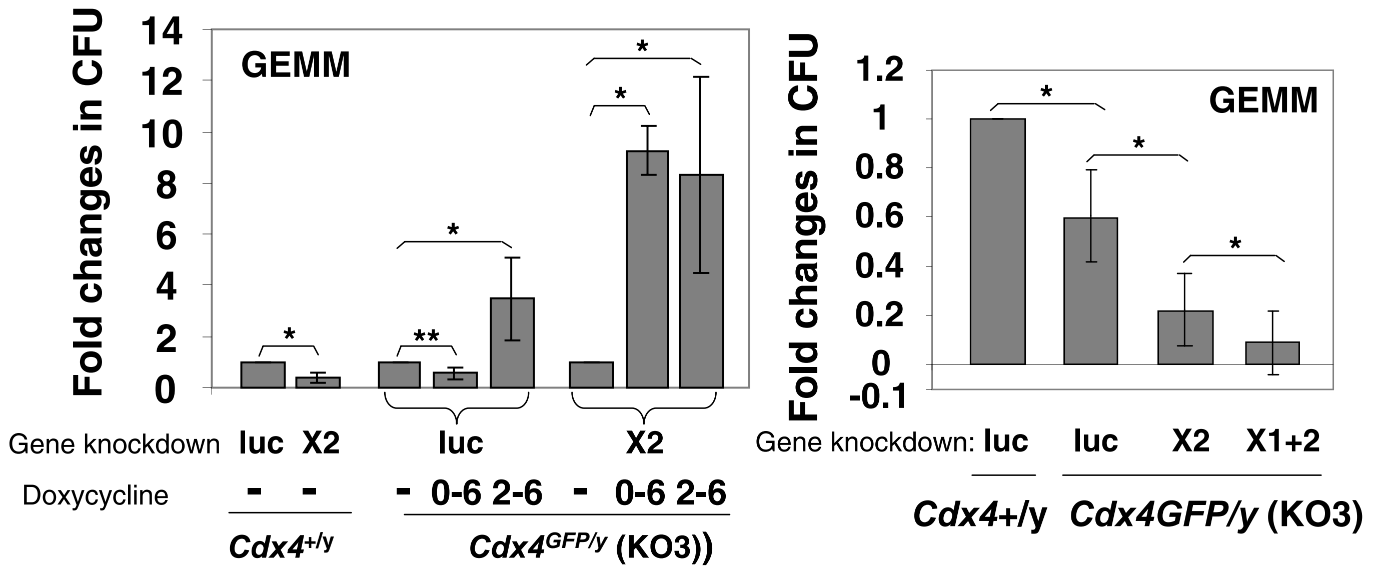


Fig. S4D. *Fig. S4 continued.*

E

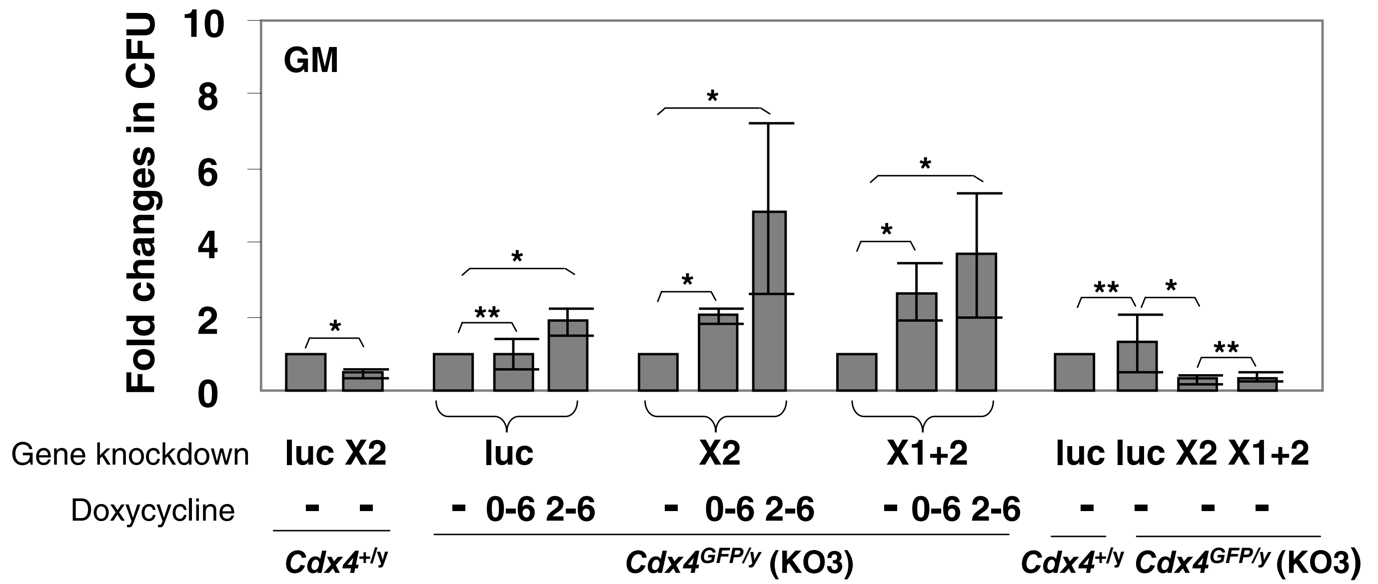


Fig. S4E. *Fig. S4 continued.*

Other Supporting Information Files

Dataset S1(XLS)
Dataset S2(XLS)