SUPPLEMENTARY INFORMATION

MATERIALS AND METHODS

Procedures for designing and making TALENs

TALENs were designed according to previously published principles[1, 2]. All TALENs used the +63 truncation point for fusion to the obligate heterodimeric FokI cleavage domain[2, 3]. The tandem arrays of TALE repeats were assembled as described[1]. The sequence of the repeat variable di-residues (RVDs) of each TALEN is listed in Supplementary Table 1.

The editing activity of each TALEN was assayed using the Surveyor nuclease (Transgenomic) as described previously[4]. Briefly, genomic DNA from TALEN-treated cells was extracted using DNeasy kit (Qiagen). The targeted loci were PCR amplified with the following primer pairs: Cel-I F(gtctgtctttgtctgtctgtc) and Cel-I R (gggtatttctctctgtgtagg) for the *Sry* locus; Cel-I F (gagttcttcttgcgttcacc) and Cel-I R (aatgagcactttcagagtagg) for the *Uty* locus. PCR products were then denatured, re-hybridized, digested with the Surveyor nuclease, and analyzed by agarose gel electrophoresis. The band quantification was based on relative band intensities using Image J software. The gene modification efficiency was calculated using equation:

% gene modification = $100 \times (1 - (1 - \text{fraction cleaved})^{1/2})$, where the fraction cleaved is the total of the cleavage bands divided by the sum of the cleavage bands and uncut bands.

Transfection and Cell culture

V6.5 mESCs (on a 129/Sv x C57BL/6 F1 hybrid background) were cultured on gelatin-coated plates with standard mESC culture conditions. Cells were transfected with TALEN plasmids alone (for knockout) or together with donor plasmids (for knock-in) using FuGENE HD reagent (Promega), following manufacturer's instructions. For each knockout experiment, 24 hours after transfection, mESCs

were re-plated at a low density on mitomycin C-inactivated mouse embryonic fibroblast (MEF) feeder layers. Individual clones were picked and genotyped by Southern blot analysis. For each knock-in experiment, 24 hours after transfection, mESC were re-plated at a low density on DR4 MEF feeder layers. Puromycin (2 μg/ml) was added one day after replating, and ganciclovir (2 μM) was added three days after replating. Individual colonies were then picked and genotyped by Southern blot analysis.

Southern blotting

Genomic DNA was separated on a 0.7% agarose gel after restriction digests with the appropriate enzymes, transferred to a nylon membrane (Amersham) and hybridized with ^{32}P random primer (Stratagene)-labeled probes. Between hybridizations, blots were stripped and checked for complete removal of radioactivity before rehybridization with a different probe.

FACS analysis

GFP-positive cells were measured using the BD LSR II flow cytometer and analyzed with FlowJo data analysis software.

Mice and tetraploid embryo complementation

Tetraploid and diploid embryo injections were performed as previously described[5]. Briefly, blastocyst-stage embryos (on a C57BL/6 x DBA/2 F2 hybrid background) were collected 94-98 hours post hCG, and injected with 10-12 mESCs using a 16 μm needle. Twenty embryos were transferred into the uteri of pseudopregnant females. C-sections were performed 17 days later. All pups were fostered to lactating ICR mothers. *Sry*$^{dl1Rlb/x}$;Tg(*Sry*)2Ei mice were obtained from Jackson Laboratory (Stock number: 010905). All experiments involving mice were approved by the Committee on Animal Care at the Massachusetts Institute of Technology.

Whole-mount immunohistochemistry

Gonads were dissected from E14.5 embryos and fixed with 4% paraformaldehyde at 4°C overnight. After blocking and permeabilization with 3% BSA/5% donkey serum/0.1% Triton X-100/PBS for another night, gonads were incubated with antibodies against FOXL2 (ab5096, Abcam), SOX9 (AB5535, Millipore) and PECAM (550274, BD Biosciences) at 4°C overnight. After washing with 0.1% Triton X-100/PBS, tissues were incubated with donkey secondary antibodies conjugated with FITC, Rhodamine Red X, or DyLight 649 (Jackson ImmunoResearch) at 4°C overnight. Antibodies were diluted 1:100 in 1% BSA/0.1% Triton X-100/PBS solution. After washing, gonads were preserved in SlowFade Gold Antifade reagent (Life Technologies). Images were taken using an LSM710 confocal microscope (Zeiss).

Fluorescence *In Situ* Hybridization (FISH) analysis

All assays were performed on V6.5 or *Sry*-targeted mESCs. Extended metaphase FISH and interphase FISH were performed as previously described[6, 7]. BAC clones, 291C04 and 146P03, from the RPCI-24 library, were used for probing the mouse Y short arms and long arms, respectively.

RT-PCR

Embryonic tissues, ESCs and MEFs were collected, submerged in TRIzol (Life Technologies), and then stored at -80°C until genotyping was completed. Total RNA was prepared according to the manufacturer's instructions and DNase-treated using DNA Free Turbo (Ambion). A total of 300 ng of total RNA was reverse transcribed. PCRs were then performed, and the products were separated on an agarose gel. For qPCR, the reactions were performed with SYBR Green dye. Results were analyzed using the delta-delta Ct method with *Gapdh* as a normalization control.

Prediction of off-target sites

The position weight matrices (PWMs) of *Sry* TALENs were constructed using model 3 as described in Moscou and Bogdanove[8] with a pseudoweight of 0.99 for observed di-residues. A PWM $\{P_{ij}\}$ of width w (where i is a position from 1 to w and j is a nucleotide from {A,C,G,T}) was converted into a position score matrix (PSM) $S_{ij}=10^3\log(P_{ij})$. The score of a potential binding site is defined as the sum of the scores of each position of the binding site against the PSM. A relative score was then calculated by subtracting the score of the site by the score of the consensus sequence. The relative score represents the relative affinity (in logs) for binding to that site compared to a site with the consensus sequence. All *Sry* Left binding sites (hits) with a minimum relative score of -13,288 (corresponding to ~10^5 weaker predicted binding affinity) were found in the mouse genome (mm9). Binding sites of *Sry* Right with minimum relative score of -13,288 were found with a spacer of 11-26 bp from each *Sry* Left hit. The total relative score of a paired hit is defined as the sum of the *Sry* Left and the *Sry* Right relative scores and was used to rank the predicted target sites.

References

1. Cermak, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011).
2. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-148 (2011).
3. Doyon, Y. et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* **8**, 74-79 (2011).
4. Guschin, D.Y. et al. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**, 247-256 (2010).
5. Wang, Z. & Jaenisch, R. At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation. *Dev Biol* **275**, 192-201 (2004).
6. Ye, M.H., Saito-Ohara, F. & Ikeuchi, T. High-resolution chromosome

R-banding in lymphoblastoid cell lines by the combined use of cell synchronization and ethidium bromide treatment. *Jpn J Hum Genet* **41**, 203-208 (1996).

7. Saxena, R. et al. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet* **14**, 292-299 (1996).

8. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).