Supplemental Methods

ES cell derivation.

ES cells were derived from delayed blastocysts harvested from C57BL/6N or 129S5/SvEvBrd females post-ovariectomy; normal male cell lines (40XY) were identified by karyotyping, as previously described (Hogan et al. 1994). The strain of origin of the C57BL/6N cells was confirmed by SNP based Genome Scanning available as a commercial service (The Jackson Laboratory, Bar Harbor, ME). The cell lines were assayed for mouse pathogens using a PCR-based screen available as a commercial service (IMPACT VIII; University of Missouri RADIL).

ES cell culture.

The ES cells were cultured according to standard protocols (Hogan et al. 1994; Matise et al. 2000). Cultures were maintained at 37^{o}C and 5% CO_2. Dulbecco's Modified Eagles Medium (DMEM) and Fetal Bovine Serum (FBS) were qualified for use in ES cell culture by assays measuring plating efficiency, colony morphology, growth rate, and cytotoxicity. ES cells were cultured in M15 (DMEM containing 4.5g/mL D-glucose, with sodium bicarbonate, without sodium pyruvate; 15% FBS; 2mM L-glutamine; 50U/mL penicillin G; 50 μg/mL streptomycin sulfate; 1x10^{-4}M ß-mercaptoethanol). Trypsinization of ES cells was performed by incubating 15-20 minutes in 0.25% trypsin with 0.04% EDTA in phosphate buffered saline (PBS, without calcium and magnesium). ES cells were cryopreserved in DMEM containing FBS at 20% and DMSO at 10% final concentration.

At all times, the ES cell cultures were maintained on mitotically inactivated (by mitomycin C) feeder monolayers of SNL fibroblasts (McMahon and Bradley 1990). These established mouse fibroblast cell lines are antibiotic-resistant (carrying a *neo* transgene), which allows their use in positive selection schemes employing G418 as the antibiotic selection agent. Additionally, these cells are engineered to secrete LIF (Leukemia Inhibitory Factor), a cytokine that prevents differentiation, promoting maintenance of the pluripotent state of the blastocyst-derived ES cells (Nichols et al. 1990). To provide consistency of performance over years of sustained production, active SNL fibroblasts were prepared from a resource of frozen, low passage number, expanded stocks. When used for production of feeder monolayers, the number of passages of the fibroblasts was limited to no more than 18.

Cell culture quality control.

Routine testing for bacteria, yeast and mold contamination by culture of samples in thioglycollate broth (Doyle and Griffiths 1998) was performed at each stage of feeder monolayer production, during weekly expansion of wild-type ES cells for gene trapping procedures, and for every set of replicate 96-well plates at the time of preparation of clones for cryopreservation. Sampling to test for mycoplasma contamination was done weekly for all mitotically active SNL fibroblasts, for all expanded wild-type ES cell cultures prepared for retrovirus infection, and for a sentinel selection of ES cell clone production plates at the time of preparation for cryostorage. Each individual clone expanded for chimera production and germline transmission studies was tested for mycoplasma and other microbial contamination at the time of cryopreservation. Two

mycoplasma assays were used: the MycoAlert enzyme assay (Cambrex Corporation, East Rutherford, NJ) and a PCR assay that allows speciation (Uphoff and Drexler 2004). All ES cell and SNL stocks used have been shown to be mycoplasma-free using both methods.

Infection, selection and picking of ES cell clones.

The method of packaging and titration of viral particles bearing the trapping vector has been detailed elsewhere (Soriano et al. 1991). OmniBankII vectors (accessions EU676801, EU676802, and EU676803) were packaged using the GP+E packaging line (ATCC, Manassas, VA). To maintain high titers of correctly packaged retroviral trapping vectors, viral producing cells (VPCs) were replaced after 10 passages. For each new culture of VPCs, titer was defined as the number of antibiotic resistant ES cell colony forming units per mL. At 24 hours before infection, VPCs at ~70% confluency were fed with a reduced volume of medium (1 mL per 10 cm^2 culture area or 16 mL per 15 cm plate) to concentrate the virus. After 24 hours the culture supernatant was harvested and filtered through a 0.2 μm filter. ES cells (at passage 8-9) were plated at 2 million cells per 10 cm plate; after 24 hours incubation they were infected with a volume of the VPC culture supernatant determined to contain viral particles at a concentration sufficient to produce G418 resistance in up to 200 colonies per 10 cm plate. Polybrene (Hexadimethrine bromide, Sigma-Aldrich, St. Louis, MO) was added to 4 μg/mL, promote infection. Antibiotic selection was begun 24 hours post-infection.

After 10 days of antibiotic selection (G418; 1.25X=180µg/mL, 0.8X = 115µg/mL) isolated colonies were picked into individual wells of 96-well plates where they were trypsinized and vigorously triturated to isolate individual cells. The cell suspensions were transferred to feeder monolayers in 96-well plates. After 3 days of growth the cells were trypsinized and transferred in triplicate to new 96-well feeder monolayer plates. After 3 additional days of growth, one plate was used for sequence tag acquisition and two replica plates were processed for cryopreservation. For cryopreservation, the cells were trypsinized, resuspended in culture medium containing 10% DMSO, transferred to 96-well format cluster tube racks, overlaid with paraffin oil and frozen at approximately 1^{o} per minute at -86^{o}C. For long-term storage, the cluster tube boxes were transferred to vapor phase liquid nitrogen tanks. Clones chosen for mouse production were expanded successively from the 96-well format, to a 24-well format, then to 6 cm plates and finally to 10 cm plates. Clones were at passage 16-18 at the end of expansion. Preparation for microinjection added 2 additional passages, with most clones reaching passage 18 by the time of blastocyst injection. The cells of the expanded clone were cryopreserved in 3 equal aliquots, each consisting of one-third of the cells and 0.5 mL of freezing medium.

QPCR assay for *neo* and *Sry*.

We performed quantitative real-time PCR using the Applied Biosystems 7700, AmpliTaq Gold DNA Polymerase and TaqMan probes. The quantity of PCR product was determined using the Comparative C_T method. PCR was performed as a multiplex of either the *neo* gene or the *Sry* gene, with *Csk* (c-src tyrosine kinase) as an endogenous reference. The threshold cycle (C_T) indicates the fractional cycle number at which the

amount of amplified target reaches a fixed threshold, well above the background signal. The value ΔC_T was defined as the difference in threshold cycles for target (*neo* or *Sry*) and the reference (*Csk*, c-src tyrosine kinase). The value $\Delta\Delta C_T = \Delta C_{T,q} - \Delta C_{T,cb}$ was the difference in threshold cycles of samples relative to a calibrator. Using these values, the amount of target, normalized to an endogenous reference and relative to a calibrator, was calculated as $2^{-\Delta\Delta CT}$. Based on this value, we determined whether the sample was wild-type ($2^{-\Delta\Delta CT} = 0.0$ for *neo*; *Sry* = 0.0 means the Y chromosome was lost), heterozygous ($2^{-\Delta\Delta CT} = 1.0$ for both *neo* and *Sry*), or homozygous ($2^{-\Delta\Delta CT} = 2.0$). As ES cell cultures represent a population of cells, some of which have lost either *neo* or *Sry*, the proportion of the population that contains *neo* or *Sry* was reflected in the results range from 0.0 to 2.0 or higher. *Neo* QPCR was also used as a routine, non-gene specific method for genotyping animals as wildtype ($2^{-\Delta\Delta CT} = 0.0$) heterozygous ($2^{-\Delta\Delta CT} = 1.0$) or homozygous ($2^{-\Delta\Delta CT} = 2.0$).

Blastocyst microinjection and chimera breeding.

All animal handling methods were approved by the internal animal care and use committee. The Lexicon Pharmaceuticals animal facilities are AALAC certified. Blastocyst hosts for microinjection were collected from 8- to16-week old pregnant females of the albino strain C57BL/6J-*Tyr*$^{c\text{-}Brd}$. Females were examined for vaginal plugs the day after mating; those that exhibited a vaginal plug were housed separately. Embryos were harvested 3.5 dpc by dissection and flushing of the uterine horns.

Approximately 15 129S5/SvEvBrd ES cells or 15-20 C57BL/6N ES cells were injected into each blastocyst. Generally, 30-35 embryo hosts were implanted per injection, and transferred equally into 3 pseudopregnant recipient females (ICR strain, mated to vasectomized males). Isofluorane was used for anesthesia. Chimeric animals were identified (black-eyed pups) within 7 days of birth and the extent of chimerism, measured as a subjective visual estimate of non-albino coat color contribution, was determined between 21 and 28 days of age. To achieve germline transmission, male chimeras were mated to C57BL/6J-*Tyr*$^{c\text{-}Brd}$ females.

RT-PCR analysis.

RNA was extracted from spleen and kidney using a bead homogenizer and RNAzol (Ambion, Austin, TX) according to manufacturer's instructions. Reverse transcription was performed with SuperScript II (Invitrogen, Carlsbad, CA) and random hexamer primers, according to the manufacturer's instructions. PCR amplification was performed for 35 cycles with oligonucleotide primers (5'-CGACATCACCTGCTAGGCTGGCTGA-3' and 5'-AGAACAATAATCAGGAGTGCTTGCC-3') complementary to exons 1 and 8 of the mouse *Slc25a40* gene.

Supplementary Results

Pilot study and QC of C57BL/6N ES cell lines.

Fifty lines of ES cells were established and confirmed by karyotyping as XY. The strain of origin of the ES cells was confirmed as C57BL/6N. The cells were found to be free of mouse pathogens. Seven lines with euploid karyotype were tested for competency in germ-line transmission; one line (Lex3.13) produced higher percentage chimeras than the others and no female chimeras after a single round of blastocyst injection. Germline transmission of coat color was achieved with these initial chimeras. This line also transmitted coat color and the desired associated mutation in a test of its use for gene-targeting experiments. Line Lex3.13 was chosen for expanded testing in a pilot experiment to examine performance in germline transmission after gene trapping. For this preliminary assessment, the ES cells of line Lex3.13 were infected with a standard, promoterless gene trapping vector (accession EU676804) and were isolated under selection with G418. A total of 1,920 G418 resistant clones were picked, grown in triplicate, and processed for cryopreservation and sequence tagging. To assess the germline competence of the mutated ES cell clones, 103 novel gene trap mutations were selected for further testing. Sixty-three of these 103 clones (62%) produced at least one surviving male chimera of 20% or greater black coat-color. Chimeras from 56 projects were mated, with 31 projects (55%) achieving germline transmission of coat color.

When compared to the laboratory standard 129S5/SvEvBrd cells cultured with the same media and reagents, we have observed a tendency for the OmniBankII C57BL/6N lines to resist trypsinization and trituration; the resulting cell clumps, if not broken up thoroughly, are prone to widespread differentiation. We have added additional minutes of trypsinization at 37° (up to 20 minutes) and up to 60 vigorous triturations during splitting

to ensure that cell clumps are thoroughly disrupted. No mycoplasma contamination has been detected during the production of the OmniBankII library. To date, over 400 individual clones representing the entire range of production have been thawed from the library and processed for quality control assays. None have exhibited either mycoplasma or other microbial contamination.

Insertion site signature

In addition to the common insertion site signature classes described in the main text, we identified a rare class of insertion where the duplicated junction sequence was non-identical. A survey of these sequences revealed that the upstream insertion junctions consistently displayed two (or rarely one) A nucleotides immediately adjacent to the terminal proviral sequence of TTTCA that was not representative of the host chromosome. These A nucleotides are likely retrovirus-derived and represent the two unpaired 5' nucleotides usually excised during the final step of integration, the 5' end joining reaction (Coffin et al. 1997). To determine if the insertion site target sequence might play a role in the retention of these unpaired nucleotides, we queried the base frequencies in the mouse genome surrounding these insertion sites to see if they differed from the overall MuLV insertion site patterns described above. This analysis revealed explicit base preferences at two positions (Supplemental Figure 7D). We observed a near exclusivity for an A nucleotide at position 3 (two nucleotides downstream from the 5' cleavage site), and exclusion of an A at position 4. Therefore, retention of additional virus-derived bases at the insertion junction appears to be driven by host sequence configuration.

Supplemental References

Coffin, J.M., S.H. Hughes, and H.E. Varmus. 1997. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Doyle, A. and J. Griffiths. 1998. *Cell and Tissue Culture: Laboratory procedures in biotechnology*. John Wiley and Sons, New York.

Hogan, B., R. Beddington, F. Costantini, and E. Lacy. 1994. *Manipulating the Mouse Embryo: A laboratory manual.* Cold Spring Harbor Laboratory Press, Plainview, NY.

Matise, M., W. Auerbach, and A. Joyner. 2000. Production of targeted embryonic stem cell clones. . In *Gene Targeting, A Practical Approach* (ed. A. Joyner), pp. 101-132. Oxford University Press, New York.

McMahon, A.P. and A. Bradley. 1990. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62:** 1073-1085.

Nichols, J., E.P. Evans, and A.G. Smith. 1990. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110:** 1341-1348.

Soriano, P., G. Friedrich, and P. Lawinger. 1991. Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells. *J Virol* **65:** 2314-2319.

Uphoff, C.C. and H.G. Drexler. 2004. Detection of Mycoplasma Contaminations. In *Basic Cell Culture Protocols* (eds. C.D. Helgason and C.L. Miller), pp. 13-23. Humana Press, Totowa.