frequency was then calculated as the number of *cII*-mutated plaques (24 °C)/total number of plaques (37 °C).

As a positive control for the *in vivo* λ select-*cII* mutagenesis assay, mice were exposed to 150 mg kg^{-1} N-ethyl-N-nitrosourea (ENU, Sigma) in a single intraperitoneal injection on one occasion. Mice were allowed to recover and were euthanized 3 weeks later.

CFU-S_{12} assays. CFU-S_{12} assays were performed as described previously[5], except that CFU-S colonies were counted after 12 days. In brief, to assess the frequency of CFU-S in mutant mice, total bone marrow was flushed from the femora and tibiae of mutant mice and appropriate controls. Nucleated cells were enumerated using a solution of 3% acetic acid and methylene blue and injected intravenously into 20 recipient mice that had been lethally irradiated. After 12 days the spleens were fixed in Bouin's solution (Sigma), and the number of colonies were counted and expressed relative to the number of total bone marrow cells injected.

To assess the survival of CFU-S_{12} after exposure to acetaldehyde, we treated total bone marrow cells with 4 mM acetaldehyde for 4 h *in vitro* before injecting them into lethally irradiated recipient mice. After 12 days, the number of CFU-S were counted. Survival was expressed relative to the untreated control for each genotype. Each data point represents the mean CFU-S survival in ten recipient mice, expressed relative to untreated samples of the corresponding genotype.

Single HSC transplants. The single-stem-cell transplants were performed as described previously[39,40]. CD45.1 recipient mice (C57BL/6J × 129S6/Sv F1, 8-to-12-weeks old) were fed with water supplemented with the antibiotic enrofloxacin (Baytril, Bayer Corporation) for seven days before irradiation and for the duration of the experiment. The lethal radiation was delivered as a split dose of 1,000 rad (500 rad each, 3 h apart) using a ^{137}Cs GSR C1m source (GSR Gmbh).

The lethally irradiated recipients were injected with single HSCs sorted from *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ and control mice on a C57BL/6Jo1a × 129S4S6/Sv F1 background (CD45.2, 8-to-12 weeks old). Bone marrow cells from these mice were extracted with IMDM medium (GIBCO), filtered through a 70-μm strainer, spun down for 5 min at 1,200 r.p.m. and resuspended in 6 ml IMDM medium at room temperature. These cells were then overlaid onto 6 ml Lympholyte M (Cederlane) in a 15-ml Falcon tube and spun down for 20 min at 1,400*g* at room temperature with the brake off. The interface containing the mononucleated cells was transferred into another 15-ml tube and topped up with ice-cold MACS buffer (PBS pH 7.2, 0.5% BSA, 2 mM EDTA). After 5 min of centrifugation at 1,200 r.p.m., the cell pellets were resuspended in 320 μl of MACS buffer, stained with the lineage-depletion kit (130-090-858, MACS Miltenyi Biotec) following the manufacturer's instructions and passed through LS magnetic columns. Lineage-depleted cells were spun down for 5 min at 1,200 r.p.m. and the pellets were resuspended in 200 μl MACS buffer with the antibodies described previously (see Flow cytometry), except that anti-CD48 was directly conjugated to BV421 (clone HM48-1, BioLegend). The cells were resuspended in 500 μl of MACS buffer and run on a Synergy sorter (Sony Biotechnology Inc.).

Single HSCs, defined as lineage$^-$c-Kit$^+$Sca-1$^+$CD41$^-$CD48$^-$CD150$^+$, were sorted into 100 μl StemSpan SFEM medium (StemCell Technologies) in each well of a round bottom, 96-well plate (Costar) using a Synergy cell sorter (Sony Biotechnology). The plates were spun down for 5 min at 180*g* and the presence of a single cell per well was confirmed visually. The full content of selected wells was loaded into insulin syringes (29G, 0.5-inch needle) containing 2×10^5 carrier cells in 300 μl Hank's balanced salt solution (StemCell technlogies). The contents of the syringe were used to dislodge the single HSC from the bottom of the well, avoiding the creation of bubbles. The entire volume (400 μl) was injected into the tail veins of irradiated recipients and chimaerism was measured every 2 weeks using flow cytometry. Recipients were considered reconstituted by the single stem cell if chimaerism for CD45.2$^+$ WBCs was $\geq$ 0.1%.

Whole-genome sequencing of mouse HSC clones. Single HSCs were allowed to expand *in vivo* for four months to guarantee that the transplanted cell was a stem cell. Recipient mice that were positive for reconstitution were euthanized four months after transplantation, and the blood, bone marrow, spleen and thymus were collected. All tissues were prepared for flow cytometry as described above and stained with CD45.1 (FITC, clone A20, BioLegend) and CD45.2 (APC, clone 104, BioLegend) antibodies. CD45.2$^+$ cells were sorted from each tissue on a Synergy cell sorter (Sony Biotechnology), spun down and frozen at −80 °C. Genomic DNA was then extracted from the CD45.2$^+$ bone-marrow cells and from a tail biopsy that had been collected at 2 weeks of age from the same mouse that provided the single HSC. Genomic DNA was extracted with the Puregene Cell and Tissue kit (Qiagen) following the manufacturer's instructions.

Whole-genome sequencing was performed as described previously[41]. In brief, short-insert 500-bp genomic libraries were constructed according to Illumina library protocols and 100-base paired-end sequencing was performed on HiSeq 2000 or HiSeq X genome analysers to an average of 20× coverage. Short-insert paired-end reads were aligned to the reference mouse genome (NCBIM38) using BWA-MEM (http://bio-bwa.sourceforge.net/). For each HSC clone, the matched tail sample was used as the reference. We sequenced two of the *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSC clones to 40× coverage, together with their matched germ-line references. We noted a big overlap between the variants found at 20× and 40× coverage (data not shown), showing that doubling the coverage did not actually uncover many more mutations. Any additional calls found when the coverage was increased to 40× were predominantly subclonal, with a mean VAF of 0.22. Therefore, we concluded that 20× whole-genome sequencing provided sufficient coverage to allow us to uncover mutations present in the transplanted HSC.

Substitutions, indels and structural changes were called with the CaVEMan, Pindel and BRASS algorithms, which are described in detail elsewhere[42]. In addition to previously reported filtering[41], we asked that all variants were unique to each HSC clone and not found in unrelated HSC or tail samples, or in unrelated mouse strains. We did not apply a VAF filter for clonality because when we examined the VAF distribution of the final datasets, these were centered around 0.5 (Extended Data Fig. 9). For the transcriptional analysis, previously reported HSC RNA-seq data was aligned with Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) against NCBIM38[43], and the overlap between indels and transcribed genes was calculated in R. To calculate the fraction of the genome covered by genes, positions of genes were retrieved from Ensembl. Overlapping regions between two genes were taken into account and counted only once.

Validation of indel and rearrangement calls. Indel calls of less than 50 bp were validated using multiplex PCR and targeted re-sequencing. We used 13 multiplex-primer combinations to capture and simultaneously amplify 172 (50%) of the previously identified indels. Primers were designed using MPprimer[44] to capture regions of between 190 and 250 bp in size (sequences available upon request). The first of round multiplex-PCR amplifications was performed with tailed gene primers and was individually barcoded by a second round of PCR with pre-validated MiSeq-ready primers[45], using a high fidelity polymerase (Q5 Hot Start HF, New England Biolabs). The PCR reaction conditions were as follows: Group 1, 100 ng DNA input in 25 μl PCR reaction, 95 °C for 2 min, six cycles of 98 °C for 20 s, 65 °C for 60 s, 60 °C for 60 s, 55 °C for 60 s, 50 °C for 60 s and 70 °C for 60 s, the reaction was then held at 4 °C until addition of barcoded second-round primers, followed by 19 cycles of 98 °C for 20 s, 62 °C for 15 s and 72 °C for 30 s, then 72 °C 60 s; Group 2, 10 ng DNA input in 5 μl PCR reaction, 95 °C for 2 min, seven cycles of 95 °C for 20 s, 58 °C for 17 min and 70 °C for 60 s, then held at 4 °C until addition of barcoded second-round primers, followed by 23 cycles of 98 °C for 20 s, 62 °C for 15 s and 72 °C for 30 s, then 72 °C 60 s. Each sample was pooled, size selected by SPRI (×0.8) and quantified before being stored at −20 °C until sequencing. Two MiSeq runs (300-bp paired-end) were used for variant confirmation, and reads were mapped with BWA. We analysed positions where the coverage was higher than 100× (159/172). With this approach, we were able to validate 91.2% of the original calls: 14/159 (8.8%) calls had VAF values <0.1 and were deemed false positives. The normal distribution of VAFs around 0.5 is consistent with most indels being clonal in origin.

For validation of rearrangements, we designed nested PCRs surrounding the breakpoints determined by the BRASS algorithm (primer sequences available upon request). PCR reactions were carried out in 20 μl, using 10 ng (HSC clones or tails) or 400 ng (donor bone marrow) of genomic DNA and GoTaq G2 Hot Start Polymerase (M7401, Promega). The first round of PCR amplifications was performed at 95 °C for 2 min, 35 cycles of 95 °C for 30 s, 55 °C for 20 s and 72 °C for 30 s, then 72 °C for 5 min. The reactions were diluted to 1 in 50 and 1 μl of the diluted reaction was used as template for a second round of PCR with nested primers, to increase specificity and sensitivity, performed at 95 °C for 2 min, 35 cycles of 95 °C for 30 s, 60 °C for 20 s and 72 °C for 30 s, then 72 °C for 5 min. The reactions were analysed on 2% agarose gels, bands of the expected sizes were excised and the identities of all products were confirmed by Sanger sequencing.

Using this approach, we found that 16/27 (59%) of rearrangements could be detected in the bone marrow of donor mice at the time the HSCs were transplanted (Extended Data Fig. 10). Any rearrangements present before transplantation must be clonal (that is, would not have arisen after the transplant). The failure to amplify the remaining 11 rearrangements by PCR does not mean that these are sub-clonal (that is, post-transplant) events. PCR amplification will depend on how much the transplanted HSC was contributing to blood production in the donor animal at the time of the transplant, as well as the sensitivity of each PCR. Therefore, we inferred clonality for the remaining calls by looking at loss of copy number (in the case of deletions, see Fig. 5l) and the number of reads involved in the rearrangement at the breakpoint for copy number-neutral changes.

Statistical analysis. Sample number (*n*) indicates the number of independent biological samples in each experiment. Sample numbers and experimental repeats are indicated in figure legends or Methods. Normality of data distribution was tested using the D'Agostino–Pearson omnibus normality test and variance was estimated before deciding on a statistical test. Unless otherwise stated in the