Supplemental Methods

Animals
Wildtype C57BL/6j CD45.2 mice were purchased from Janvier and kept in a SPF facility at the Karolinska Institute until they were up to 24 months of age. All mouse experiments were performed according to the guidelines and obtained permissions from the ethics committees at Stockholm Norra Djurförsökssetisks Nämnd (N69/16).

Human specimens
Bone marrow and buccal swabs from healthy elderly individuals (70-75 years old) were collected with informed consent and approved by the Stockholm regional ethical review board (EPN 2018/901-31).

DNA extraction from bone marrow cells and tail
BM cells were isolated by crushing the femurs, tibiae, cristae, forearms and sternum in PBS without Ca$^{2+}$ and Mg$^{2+}$ supplemented with 5% fetal bovine serum (HyClone) and 2 mM EDTA (Sigma) with a mortar and pestle, and filtered through a 40 µm cell strainer. The tail tip was collected from each mouse as germline control DNA. All DNA isolation was performed using a DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer’s instructions.

In vitro clonogenic expansion of hematopoietic stem cells
Mouse HSCs defined as Lineage$^{neg}$Sca-1$^+$c-Kit$^+$CD150$^+$CD48$^-$ were identified by flow cytometry after staining BM MNCs from aged mice with the following antibodies: c-Kit APCeF780 (2B8), Sca-1 FITC (D7), CD48 BUV737 (HM48-1), CD150 PECy7 (TC15-12F12.2), CD4 PECy5 (RM4-5), CD5 PECy5 (53-7.3), CD8a PECy5 (53-6.7), B220 PECy5 (RA3-6B2), Ter119 PECy5 (TER-119) and Gr-1 PECy5 (RB6-8C5). All FACS experiments included fluorescence-minus-one (FMO) and single-stained controls, and viable cells were identified by DAPI exclusion. Cells were sorted on a FACS ARIA Fusion cell sorter (BD). For single cell sorting, cells were deposited by an ACDU single cell depositor unit directly into
individual wells of a 96-well U-bottom plate in 100 µl of X-vivo medium (Lonza) supplemented mouse SCF (10 ng/mL, R&D Systems), mouse GM-CSF (10 ng/mL, Peprotech), human G-CSF (10 ng/mL, Amgen), human TPO (10 ng/mL, Peprotech), mouse IL-3 (5 ng/mL, Peprotech), human Flt3 ligand (10 ng/mL) and human EPO (1 IU/mL, Roche). Cells were cultured in an incubator at 37°C, 5% CO$_2$ and 90% humidity for 14 days, with a weekly half-medium change. Individual clones evaluated to contain sufficient cells for DNA extraction were isolated for DNA extraction using a QIAamp DNA Micro Kit (Qiagen).

Single human HSCs, defined as Lineage$^\text{neg}$CD34$^+$CD38$^-$CD90$^+$CD45RA$^-$, were identified and sorted as described above after staining viably frozen BM MNCs from aged, healthy individuals with the following antibodies: CD34 BUV395 (8G12), CD45RA BB515 (HI100), CD38 PE-Dazzle-594 (HIT2), CD90 PE (5E10), CD123 PECy7 (6H6), CD2 PECy5 (RPA-2.10), CD3 PECy5 (HIT3a), CD4 PECy5 (RPA-T4), CD7 PECy5 (M-T701), CD8a PECy5 (RPA-T8), CD10 PECy5 (HI10a), CD11b PECy5 (ICRF44), CD14 PECy5 (61D3), CD19 PECy5 (HIB19), CD20 PECy5 (2H7), CD56 PECy5 (B159) and CD235a,b PECy5 (HIR2). Single HSCs were cultured as described for mouse HSCs in StemSpan SFEM medium (StemCell Technologies) supplemented with 10% BIT (StemCell Technologies), 10$^4$ M 2-βMercaptoethanol (Sigma), 1% Penicillin-Streptomycin (PAA Labs), 10 ng/mL human Tpo (Peprotech), 10 ng/mL human SCF (Stemgen), 10 ng/mL human FLT3 ligand (Immunex), 5 ng/mL human IL-3 (Peprotech), 10 ng/mL human GM-CSF (Berlex), 10 ng/mL human G-CSF (Amgen) and 1 IU/mL human EPO (Roche). Individual clones were expanded for up to 29 days and DNA was isolated for further analysis.

**Whole genome sequencing analysis**

DNA extracted from single colonies were subjected to whole genome sequencing. Library preparation was performed using KAPA Hyper Prep Kit (KAPA Biosystems), followed by sequencing using NovaSeq 6000 (Illumina) at the National Genomics Infrastructure in Stockholm. Mouse and human colonies were sequenced with an average depth of 22.7x (22.4 – 25.0) and 22.6x (20.0-24.9), respectively. Paired tail or buccal swab DNA was used for excluding SNPs for mice and humans, respectively, as well as sequencing errors. Sequencing reads were aligned to the mouse (GRCm38) or human (GRCh37) genome reference using Burrows-Wheeler Aligner version 0.7.8 with default parameter settings. PCR duplicates were marked with biobambam version 0.0.191. Mutations calling was performed using Genomon2 pipeline (https://github.com/Genomon-Project/GenomonPipeline) with the following
Chin et al.

parameters:
(i) Mapping Quality score ≥ 20
(ii) Base Quality score ≥ 15
(iii) Number of total reads ≥ 8
(iv) Number of variant reads ≥ 4
(v) Variant allele frequency in tumor ≥ 0.25
(vi) Variant allele frequency in normal < 0.1
(vii) Strand ratio in tumor ≠ 0 or 1
(viii) P value by Fisher < 0.01
(ix) VAF by base counts / VAF by reads count ≥0.5 and ≤2
(x) P value by Fisher < 0.01
(xi) P value by EBFfilter¹ < 0.001

Colonies with variant allele fractions distributed around 50%, confirming that they were derived from single HSCs, were used for the analysis (Supplemental Figure 1A). To eliminate mutations potentially acquired as a result of cell culture, only mutations with variant allele frequency above 25% were used for the analysis. Called mutations were annotated by ANNOVAR.² In order to estimate the mutation rate per Mb, genome sizes were calculated with R package “BSgenome.Hsapiens.UCSC.hg19” (Chr1-22, X, Y, MT) and “BSgenome.Hsapiens.UCSC.mm10” (Chr1-19, X, Y, MT).

Mutational signature analysis of single nucleotide mutations detected by whole genome sequencing was performed using R package MutationalPatterns.³ Optimal nonnegative linear combination of mutation signatures to reconstruct the mutation matrix was evaluated by fitting to COSMIC mutational Signatures v2 (https://cancer.sanger.ac.uk/cancergenome/assets/signatures_probabilities.txt). Trinucleotide patterns in human and mouse genomes showed high similarity (cosine similarity: 0.9965, Supplemental Figure 1B), and we compared the mutational patterns between human and mouse colonies without any correction. Similarity of 96 trinucleotide signatures between human and mouse samples was evaluated using cosine similarity and was clustered by “ward.D2” algorithm using Euclidean distances. Bootstrap resampling was performed for fitting to the COSMIC signatures with 1000 time iterations to get the stable results of contributions of each signature. The mean values of 1000 iterations for all the human and mouse colonies are hierarchically clustered with the “ward.D2” algorithm using Euclidean
distances calculated from the frequencies of fitted signatures both for colonies and signatures and visualized by heatmap.2 function (R package gplots). Reconstructed 96 trinucleotide signatures from allocated COSMIC signatures showed high similarities with original 96 trinucleotide signatures in each colony both in human (mean cosine similarity (min-max): 0.966 (0.950-0.996)) and mouse (0.858 (0.804-0.902)).

**Prediction of random mutations in CHIP-associated genes**
The number of mutations in CHIP-associated driver genes targeted to the HSC pool of aged mice and human individuals was estimated using the following information under the assumption that somatic mutations occur randomly throughout the genome.
- Number of HSCs in adult mice: 5,000-10,000 cells
- Number of HSCs in adult humans: 44,000-215,000 cells
- Bait size used for this ECTS: 18597 bp
- Individual gene size:
  - *Sf3b1* 480 bp
  - *Trp53* 1800 bp
  - *Srsf2* 359 bp
  - *Dnmt3a* 4080 bp
  - *Jak2* 120 bp
  - *Asxl1* 5399 bp
  - *Tet2* 6359 bp
- Human Genome size (Chr1-22, ChrX, ChrY and ChrM): 3095693983 bp (calculated using R package “BSgenome.Hsapiens.UCSC.hg19”)
- Mouse Genome size (Chr1-19, ChrX, ChrY and ChrM): 2725537669 bp (calculated using R package “BSgenome.Mmusculus.UCSC.mm10”)
- The average number of mutations per HSC colony from aged mice (Figure 1B): 193.6667 mutations
- The average number of mutations per HSC colony from aged human donors (Figure 1B): 969.3333 mutations
- The probabilities that one HSC has one mutation on bait were calculated as follows:
  Probability = (Average mutation number per HSC colony)/(genome size) x (bait size for target regions)
- The number of mutant cells in total BM was assumed to follow a binominal distribution with parameters of (number of HSCs) and (Probability).
Using these parameters, we calculated the smallest number of mutant HSCs (N) within one aged mouse for each gene and within the total bait size, where the probabilities of having N or less becomes ≥95% independently both for the upper and lower number of HSCs (mouse: 5,000 and 10,000 cells, human: 44,000 and 215,000).

Non-competitive bone marrow transplantation
Following transplantation of 20 million bone marrow (BM) mononuclear cells (MNCs) from individual 24-month old C57BL/6j CD45.2 wildtype mice into lethally irradiated (split X-ray dose, total 9.6Gy) B6.SJL-Ptprca Pep3b/BoyJ CD45.1 recipient mice (2 recipients/donor), peripheral blood donor contribution was accessed at 12 weeks post-transplantation using the following antibodies: NK1.1 PB (PK136), CD11b APC (M1/70), Gr1 PO (RB6-8C5), CD19 PE/Cy7 (1D3), CD4 APC-eFluor 780 (RM4-5), CD8a APC-eFluor 780 (53-6.7), CD45.1 PE (A20) and CD45.2 AF700 (104). Transplanted mice were kept for an additional 10 months (range 2-11 months). Upon termination, BM recipient- (CD45.1) and donor-derived (CD45.2) cells were isolated by FACS sorting after staining with CD45.1 PE (A20) and CD45.2 AF700 (104). DNA was isolated from purified 1 million CD45.2 BM donor cells using a DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer’s instructions.

Droplet digital PCR
A 20 µL PCR reaction containing 1x ddPCR Supermix for probes without UTP (Bio-Rad), 1x mutation detection primer-probe assay (designed and ordered from Bio-Rad), 60 ng (equivalent to 10,000 cells) template DNA (isolated from >1 million BM cells) from each individual mouse or gBlock control DNA was prepared and mixed with 70 µL of Droplet Generation Oil for Probes (Bio-Rad). Droplet generation was carried out according to manufacturer’s instructions on a QX200 droplet generator (Bio-Rad). Forty microliters of emulsified PCR mixture were transferred into a 96-well PCR plate and subjected to the following cycling conditions: 10 minutes at 95°C; 40 cycles each consisting of a 30 second incubation at 94°C and a 1 minute incubation at 55 to 60°C (selected based on gradient PCR analysis for the optimal annealing temperature for each individual primer-probe assay); and a final 10 minutes incubation at 98°C. The temperature ramp increment was 2.5°C/sec for all steps. Plates were read on a QX200 droplet reader (Bio-Rad) and the results were analysed using QuantaSoft v1.5.38.1118 software (Bio-Rad).

Mutation detection assay specific for each mutation was designed using Bio-Rad’s Droplet Digital PCR online design tool. A gene fragment block (gBlock, Integrated DNA
Technologies) for each mutation detection assay was generated and mixed with C57BL/6 genomic DNA to serve as a mutated DNA positive control. Genomic DNA from the bone marrow of 5 healthy young (six-weeks old) mice was included as negative controls. At least one young DNA negative control, one gBlock positive control and one no-template water control were included in each ddPCR run.

QuantaSoft software was used to calculate the VAF as fractional abundance of the mutated allele based on Poisson distribution as described in Bio-Rad’s “Droplet Digital PCR Application Guide”. Any wells with less than 10,000 accepted events were excluded from the analysis. The 2D fluorescence amplitude for each well was also visually inspected, and samples with abnormal signal intensity was also excluded. For screening for presence of hotspot mutations, a conservative limit of detection (LOD) at VAF = 0.05% (1 mutant in 2000 events), a minimum of 6000 copies of haploid DNA genome as recommended in the Bio-Rad’s “Droplet Digital PCR Application Guide” was used. A sample was defined as positive for the analysed mutation if the VAF was found to be higher than LOD, and there were 5 or more mutant only (ch1+ch2-) events. A sample was defined to be negative for the analysed mutation if the VAF was lower than the LOD, or had no detectable mutant only event. A sample was defined as inconclusive if it had a VAF higher than the assay LOD, but less than 5 mutant only (Ch1+Ch2-) events. For validation of mutations detected by error corrected targeted DNA sequencing, a lower threshold was set for confirming presence of a mutation requiring one or more mutant only event (Ch1+Ch2-).

**Error corrected targeted DNA sequencing (ECTS) analysis**

BM DNA isolated from >1 million BM cells from aged and transplanted mice was subjected to ECTS. 200ng (equivalent to 30,000 cells) of fragmented DNA by E220 Focused-ultrasonicator (Covaris) was subjected to library preparation using KAPA Hyper Prep Kit (KAPA Biosystems), xGen Dual Index UMI Adapters (IDT) and the custom designed capture probes (xGen Lockdown Probe, IDT) for the coding regions of Dnmt3a, Tet2, Asxl1, and Trp53, as well as exons 13-14 in Sf3b1, exon 1 of Srsf2 and exon 14 of Jak2.

After hybridization using xGen Hybridization and Wash Kit (IDT), the enriched fragments are amplified using KAPA HiFi HotStart ReadyMix (KAPA Biosystems). The prepared libraries were sequenced on a Novaseq (Illumina) at the National Genomics Infrastructure in Stockholm. Sequencing reads were mapped to GRCm38 using Burrows-Wheeler Aligner. After grouping reads with the same unique molecular identifier (UMI) using Pickard (http://broadinstitute.github.io/picard) and fgbio (https://github.com/fulcrum-genomics/fgbio),
Chin et al.

Consensus reads were generated using fgbio FilterConsensusReads with the options of min-read=5, min-base-quality=40, max-read-error-rate=0.025, max-base-error-rate=0.1, max-no-call-fraction=0.05. Consensus reads were subjected to indel realignment and base quality score recalibration using GATK3 and recalculation of MD/NM tags using SAMtools version 1.9 resulting in an average depth of 7903 (3980-14470) in the final bam files. Mutation calling was performed with EBCall with the parameters as outlined below:

(i) Mapping quality threshold = 30
(ii) Base quality threshold = 15
(iii) Minimum depth in BM = 8
(iv) Minimum number of variant reads in BM = 3
(v) Minimum amount of tumor allele frequency = 0.00005
(vi) EBCall P value = $10^{-2}$

After mutation calling, the following candidates were removed.

(i) Outsides of target regions +/- 10 bp
(ii) Synonymous SNV, intronic, UTR3/5
(iii) Strand ratio =1 or 0
(iv) Number of alternative reads in BM <10
(v) EBCall P value $\geq 10^{-5}$
(vi) VAF in BM $\leq 99\%$ confidence interval of VAF in normal samples by Bootstrap
(vii) VAF by base counts / VAF by reads count $\geq 0.5$ and $\leq 2$

Tail DNA from 27 aged mice were used as negative controls and all detected mutations were validated by ddPCR when design of ddPCR probes were possible. The positive predictive value (PPV) and the sensitivity of the ECTS were evaluated using a target gene panel designed for capturing the corresponding human genes and by mixing DNA from Jurkat cells with DNA from K562 cells (both from ATCC) at different ratios (0.1%, 0.2%, and 1%). Libraries were also prepared from pure Jurkat and K562 DNA in order to identify unique single nucleotide variants (SNVs) that were used for the downstream analysis. Sequencing data was processed as described from mice. Seven Jurkat specific SNVs (6 heterozygous and 1 homozygous SNV) were detected and were used for the evaluation of PPV and sensitivity.

**Statistical analysis**
Welch’s t test and Fisher’s exact test were used to evaluate statistical significance. The significance of detected driver mutations in error-corrected sequencing was evaluated by dN/dS ratios in each targeted gene using dNdScv (https://github.com/im3sanger/dndscv). Information of coding regions of GRCm38 mouse was downloaded from BioMart (https://www.ensembl.org/biomart/martview/). The reference for dNdScv was generated after selecting transcripts with completed coding regions for targeted genes using "buildref" function. The maximum likelihoods of dN/dS were evaluated only for targeted genes. q-values were obtained by Benjamini-Hodgberg’s multiple testing correction. All reported p values were for two-sided tests and the significance level was set at 0.05.

Acknowledgements
The authors thank Kari Högren, Allegra Lord, Edyta Wojtowics, Laura Kettyle and the staff at Komparative Medicine animal facility for experimental assistance. This project was supported by grants to P.S.W. from Knut and Alice Wallenberg Foundation (KAW 2015.0195 and KAW 2016.0105), the MDS Foundation and the Swedish Research Council (2015-03561), and to S.E.W.J. from the Knut and Alice Wallenberg Foundation (KAW 2016.0105), Tobias Foundation (4-1122/2014), Center for Innovative Medicine (CIMED) at Karolinska Institutet (613/06), an international recruitment grant from The Swedish Research Council (538-2013-8995) and The UK Medical Research Council (MC_UU_12009/5).
**Supplementary Tables:**

| Ratio of spike-in | Sensitivity (%) | PPV (%) | # of detected mutations | # of true mutations among detected |
|-------------------|-----------------|---------|-------------------------|-----------------------------------|
| 0.1%              | 14.3            | 100     | 1                       | 1                                 |
| 0.2%              | 28.6            | 100     | 2                       | 2                                 |
| 1%                | 100             | 100     | 7                       | 7                                 |

**Supplemental Table 1. Sensitivity and positive predictive values (PPV) of error corrected targeted DNA sequencing.**

Indicated ratios of DNA isolated from Jurkat cells were mixed with DNA isolated from K562 cells. 7 (6 heterozygous and 1 homozygous) SNPs were used for the calculation.
| Mouse ID | Mutation          | Chr | Start   | End     | Ref | Alt | Ref reads | Mut reads | %VAF |
|----------|-------------------|-----|---------|---------|-----|-----|-----------|-----------|------|
| 1        | Tet2 Q761X        | 3   | 133486391 | 133486391 | G   | A   | 10966     | 35        | 0.32 |
| 1.1      | Tet2 Q761X        | 3   | 133486391 | 133486391 | G   | A   | 11186     | 3         | 0.03 |
| 1.2      | Tet2 Q761X        | 3   | 133486391 | 133486391 | G   | A   | 5629      | 2         | 0.04 |
| 2        | Asxl1 P795fs      | 2   | 153399909 | 153399909 | G   | A   | 8197      | 12        | 0.15 |
| 2.1      | Asxl1 P795fs      | 2   | 153399909 | 153399909 | G   | A   | 11296     | 1         | 0.01 |
| 2.2      | Asxl1 P795fs      | 2   | 153399909 | 153399909 | G   | A   | 12784     | 0         | 0.00 |
| 3        | Trp53_p.R270H     | 11  | 69589608  | 69589608  | G   | A   | 9965      | 58        | 0.58 |
| 3.1      | Trp53_p.R270H     | 11  | 69589608  | 69589608  | G   | A   | 9408      | 6         | 0.06 |
| 3.2      | Trp53_p.R270H     | 11  | 69589608  | 69589608  | G   | A   | 6534      | 0         | 0.00 |
| 3.1      | Trp53_p.Y160S     | 11  | 69588476  | 69588476  | A   | C   | 11744     | 17        | 0.14 |
| 3.2      | Trp53_p.Y160S     | 11  | 69588476  | 69588476  | A   | C   | 11844     | 0         | 0.00 |
| 4        | Trp53_R277K       | 11  | 69589608  | 69589608  | G   | A   | 8323      | 29        | 0.35 |
| 4.1      | Trp53_R277K       | 11  | 69589608  | 69589608  | G   | A   | 7597      | 6         | 0.08 |
| 4.2      | Trp53_R277K       | 11  | 69589608  | 69589608  | G   | A   | 8126      | 19        | 0.23 |
| 4        | Tet2_p.A1395T     | 3   | 133469402 | 133469402 | C   | T   | 8323      | 29        | 0.35 |
| 4.1      | Tet2_p.A1395T     | 3   | 133469402 | 133469402 | C   | T   | 10053     | 0         | 0.00 |
| 5        | Asxl1_p.A873G     | 2   | 153400146 | 153400146 | C   | G   | 5029      | 221       | 4.21 |
| 5.2      | Asxl1_p.A873G     | 2   | 153400146 | 153400146 | C   | G   | 11556     | 430       | 3.59 |
| 6        | Trp53_p.Q101_G102del | 11  | 69587566  | 69587571  | CCAGGG | -  | 7220      | 18        | 0.25 |
| 6.1      | Trp53_p.Q101_G102del | 11  | 69587566  | 69587571  | CCAGGG | -  | 10420     | 0         | 0.00 |
| 7        | Dnmt3a_p.G564E    | 12  | 3899899   | 3899899   | G   | A   | 8974      | 92        | 1.01 |
| 7.1      | Dnmt3a_p.G564E    | 12  | 3899899   | 3899899   | G   | A   | 8100      | 0         | 0.00 |
| 8        | Dnmt3a_p.I820M    | 12  | 3905656   | 3905656   | A   | G   | 3842      | 15        | 0.39 |
| 8.1      | Dnmt3a_p.I820M    | 12  | 3905656   | 3905656   | A   | G   | 10255     | 0         | 0.00 |
| 9        | Tet2_p.G1046D     | 3   | 133485535 | 133485535 | C   | T   | 8459      | 28        | 0.33 |
| 9.2      | Tet2_p.G1046D     | 3   | 133485535 | 133485535 | C   | T   | 6391      | 10        | 0.16 |
| 10       | Tet2_p.H1790Y     | 3   | 133467132 | 133467132 | G   | A   | 9798      | 22        | 0.22 |
| 10.1     | Tet2_p.H1790Y     | 3   | 133467132 | 133467132 | G   | A   | 8829      | 1         | 0.01 |
| 11       | Asxl1_p.L1210F    | 2   | 153401156 | 153401156 | C   | T   | 17803     | 20        | 0.11 |
| 12       | Tet2_p.K540E      |     |          |          |     |     | 8445      | 0         | 0.00 |
Supplemental Table 2. ECTS sequencing results from mice with detected mutations. Annotation of position and base change is indicated for the mouse where ECTS had the highest VAF. ID, mouse identification for aged mice where decimal indicate transplanted mice; Chr, chromosome; Ref, reference; Alt, alternative; Ref reads, number of reference reads; Mut reads, number of mutant reads; VAF, variant allele frequency.

|   |     | 3  | 133487054 | 133487054 | T  | C  | 10535 | 11  | 0.10 |
|---|-----|----|-----------|-----------|----|----|-------|-----|------|
| 13| Tet2_p.K540E | 3  | 133487054 | 133487054 | T  | C  | 10535 | 11  | 0.10 |
| 13.1| Tet2_p.A1400G | 3  | 133469386 | 133469386 | G  | C  | 9787  | 24  | 0.24 |
| Mouse ID | Mutation       | Ch1+Ch2+ | Ch1-Ch2- | Ch1-Ch2+ | Ch1-Ch2- | %VAF | Min | Max |
|--------|---------------|----------|----------|----------|----------|------|-----|-----|
| 1      | Tet2_p.Q761X  | 35       | 19       | 7795     | 2905     | 0.39 | 0.28| 0.49|
| 1.1    | Tet2_p.Q761X  | 2        | 2        | 5599     | 5228     | 0.05 | 0.00| 0.10|
| 1.2    | Tet2_p.Q761X  | 2        | 1        | 6940     | 3450     | 0.03 | 0.00| 0.06|
| 2      | Assx1_p.P795fs| 1        | 2        | 5980     | 4160     | 0.03 | 0.00| 0.07|
| 2.1    | Assx1_p.P795fs| 0       | 0        | 5617     | 6298     | 0.00 | 0.00| 0.00|
| 2.2    | Assx1_p.P795fs| 0       | 0        | 5808     | 5176     | 0.00 | 0.00| 0.00|
| 3      | Trp53_p.R270H | 0        | 0        | 6486     | 5999     | 0.00 | 0.00| 0.00|
| 3.1    | Trp53_p.R270H | 1        | 5        | 1901     | 33649    | 0.31 | 0.06| 0.55|
| 3.2    | Trp53_p.R270H | 0        | 10       | 4989     | 16639    | 0.18 | 0.07| 0.28|
| 3.3    | Trp53_p.Y160S | 0        | 0        | 6436     | 6926     | 0.00 | 0.00| 0.00|
| 3.4    | Trp53_p.Y160S | 1        | 3        | 2893     | 34012    | 0.13 | 0.00| 0.26|
| 3.5    | Trp53_p.Y160S | 0        | 0        | 5047     | 5684     | 0.00 | 0.00| 0.00|
| 4      | Trp53_p.R277K | 0        | 0        | 8995     | 8121     | 0.00 | 0.00| 0.00|
| 4.1    | Trp53_p.R277K | 3        | 5        | 6143     | 7644     | 0.10 | 0.03| 0.17|
| 4.2    | Trp53_p.R277K | 3        | 8        | 7425     | 8171     | 0.11 | 0.04| 0.17|
| 4.3    | Tet2_p.A1395T | 3        | 0        | 7182     | 5663     | 0.03 | 0.00| 0.06|
| 4.4    | Tet2_p.A1395T | 17       | 12       | 5972     | 7801     | 0.37 | 0.23| 0.50|
| 4.5    | Tet2_p.A1395T | 1        | 0        | 5730     | 7986     | 0.01 | -0.01| 0.04|
| 5      | Assx1_p.A873G | 1        | 0        | 8415     | 7116     | 0.01 | -0.01| 0.02|
| 5.1    | Assx1_p.A873G | 117      | 195      | 6306     | 9272     | 3.69 | 3.29| 4.08|
| 5.2    | Assx1_p.A873G | 176      | 194      | 5703     | 7278     | 4.62 | 4.16| 5.07|
| 6      | Trp53_p.Q101_G102del | 0 | 0 | 7126 | 6823 | 0.00 | 0.00 | 0.00 |
| 6.1    | Trp53_p.Q101_G102del | 4 | 9 | 5352 | 7285 | 0.19 | 0.09 | 0.29 |
| 6.2    | Trp53_p.Q101_G102del | 0 | 0 | 5714 | 6887 | 0.00 | 0.00 | 0.00 |
| 7      | Dnmt3a_p.G564E | 0 | 0 | 7601 | 5892 | 0.00 | 0.00 | 0.00 |
| 7.1    | Dnmt3a_p.G564E | 32 | 49 | 5625 | 5126 | 1.01 | 0.79 | 1.22 |
| 7.2    | Dnmt3a_p.G564E | 0 | 0 | 5975 | 7022 | 0.00 | 0.00 | 0.00 |
| 8      | Dnmt3a_p.I820M | 2 | 0 | 7881 | 6253 | 0.02 | -0.01 | 0.04 |
| 8.1    | Dnmt3a_p.I820M | 16 | 16 | 6423 | 8044 | 0.37 | 0.25 | 0.50 |
| 8.2    | Dnmt3a_p.I820M | 2 | 0 | 6164 | 7578 | 0.02 | -0.01 | 0.06 |
| 9      | Tet2_p.G1046D | 0 | 0 | 7437 | 5457 | 0.00 | 0.00 | 0.00 |
| 9.1    | Tet2_p.G1046D | 0 | 7 | 2834 | 26080 | 0.23 | 0.06 | 0.41 |
| 9.2    | Tet2_p.G1046D | 0 | 0 | 2986 | 31974 | 0.00 | 0.00 | 0.00 |
| 10     | Tet2_p.H1790Y | 0 | 0 | 7275 | 6459 | 0.00 | 0.00 | 0.00 |
| 10.1   | Tet2_p.H1790Y | 0 | 7 | 3299 | 34855 | 0.20 | 0.05 | 0.35 |
| 10.2   | Tet2_p.H1790Y | 0 | 0 | 1440 | 36120 | 0.00 | 0.00 | 0.00 |
| 11     | Assx1_p.L1210F | 0 | 0 | 6809 | 5692 | 0.00 | 0.00 | 0.00 |
| 11.1   | Assx1_p.L1210F | 0 | 7 | 3388 | 36166 | 0.20 | 0.05 | 0.34 |
Supplemental Table 3. ddPCR analysis of mutations identified by ECTS.
ID, mouse identification for aged mice where decimal indicate transplanted mice; Ch1, mutant allele; Ch2, wildtype allele; VAF, variant allele frequency; CI, confidence interval; min, minimum VAF of 95% CI; max, maximum VAF of 95% CI.

| ID  | Mutation   | Ch1 | Ch2 | VAF | CI  | min | max |
|-----|------------|-----|-----|-----|-----|-----|-----|
| 12  | Tet2_p.K540E | 0   | 0   | 7586 | 5394 | 0.00 | 0.00 |
| 12.1| Tet2_p.K540E | 0   | 5   | 2363 | 36076 | 0.20 | 0.03 |
| 13  | Tet2_p.A1400G| 0   | 0   | 7640 | 5352 | 0.00 | 0.00 |
| 13.1| Tet2_p.A1400G| 3   | 16  | 5315 | 9096 | 0.29 | 0.16 |
| Gene   | Maximum-likelihood of the dN/dS ratios | p value       | q value       |
|--------|---------------------------------------|---------------|---------------|
|        | Missense  | Nonsense | Splice site | Indel        |               |               |
| Trp53  | 22.53     | 0.00     | 0.00      | 9.46         | 1.49E-03      | 0.0104*       |
| Tet2   | 8.00      | 14.47    | 14.47     | 0.00         | 0.0037        | 0.0130*       |
| Asxl1  | 4.12      | 0.00     | 0.00      | 2.44         | 0.2979        | 0.6027        |
| Dnmt3a | 7.10      | 0.00     | 0.00      | 0.00         | 0.3444        | 0.6027        |

**Supplemental Table 4. Statistical evaluation of positive selection of non-synonymous mutations over synonymous mutations among mutations detected by ECTS.**

Significance of dN/dS ratio was evaluated by dNdScv package. Only genes with mutations are listed. * indicates significance (see supplementary methods).
Supplementary Figures:

Supplemental Figure 1. Mutations identified by whole genome sequencing of single HSC-derived clones.

A. Variant allele frequency distribution for mutations detected in investigated single human (left) and mouse (right) HSC-derived clones (n=6 each for human and mice).

B. Comparison of trinucleotide frequency in the human and mouse genome.

C. Heatmap of cosine similarity score of the mutational trinucleotide spectra of individual

Supplemental Figure 1. Mutations identified by whole genome sequencing of single HSC-derived clones.

A. Variant allele frequency distribution for mutations detected in investigated single human (left) and mouse (right) HSC-derived clones (n=6 each for human and mice).

B. Comparison of trinucleotide frequency in the human and mouse genome.

C. Heatmap of cosine similarity score of the mutational trinucleotide spectra of individual
mouse and human HSCs. Mouse and human HSC numbers (#) refers to the corresponding numbers shown in Figure 1E. Values indicate the cosine similarities.

D. Predicted number of acquired mutations per aged normal individual within ECTS-screened CH-associated genes/regions in the HSC pool of aged mice (24 months of age) and human healthy individuals (>70 years of age) based on the HSC mutation rate, gene size and estimated HSC pool (see supplemental methods). Each line indicates the lower and upper limit of predicted mutations for each gene (black) or within all captured DNA regions (red) (see Supplemental Methods).
Supplemental Figure 2. Screening strategies for CH hotspot mutations in aged mice.

A. Mean (SEM) percent donor CD45.2 contribution to mature myeloid, B and T cells in peripheral blood of recipient mice (n=88) 12 weeks following transplantation with 20 million BM MNCs from aged donor mice (n=48).

B. Representative ddPCR plots from one experiment for the Jak2 V617F mutation analyzed in DNA from one gBlock positive control, young B6 BM MNCs negative controls (n=6 mice) and aged BM MNCs (n=54 mice). DNA from each mouse BM MNCs was investigated individually by ddPCR and representative images were generated by pooling the ddPCR events for each mouse into one plot for each group (young or aged mice).

C. Absence of CH hotspot mutations in BM MNCs isolated from aged (24 months) mice (n=97) and transplanted recipients (n=88, representing 48 donors) as screened for by ddPCR. gBlock (n=2) and young wildtype C57BL/6 mice (n=5) were included as positive and negative controls, respectively. Minimum, median and maximum % VAF are shown and dotted line represents the ddPCR detection limit of 0.05% VAF.
Supplemental Figure 3. Validation of ECTS for mouse CH genes.

A. Median sequencing depth based on detection of unique molecular identifiers (UMI) for all captured DNA regions by error-corrected targeted DNA sequencing (ECTS) in DNA isolated from BM of aged mice and post-transplantation (n=97 and 88, respectively). DNA isolated from tailtips were included as germline controls (n=27). Boxes indicate the interquartile range (IQR), the dashed lines indicate 1.5xIQR and each circle indicate the individual samples.

B. Median UMI depth for all genes captured by ECTS in all DNA samples included in the study (n=213). Boxes indicate IQR, the dashed lines indicate 1.5xIQR and the dots indicate outliers.
Supplemental References

1. Shiraishi Y, Sato Y, Chiba K, et al. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res*. 2013;41(7):e89.

2. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.

3. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med*. 2018;10(1):33.

4. Oguro H, Ding L, Morrison SJ. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*. 2013;13(1):102-116.

5. Lee-Six H, Obro NF, Shepherd MS, et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature*. 2018;561(7724):473-478.

6. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498.

7. Martincorena I, Raine KM, Gerstung M, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. 2017;171(5):1029-1041 e1021.