SUPPLEMENTAL DATA

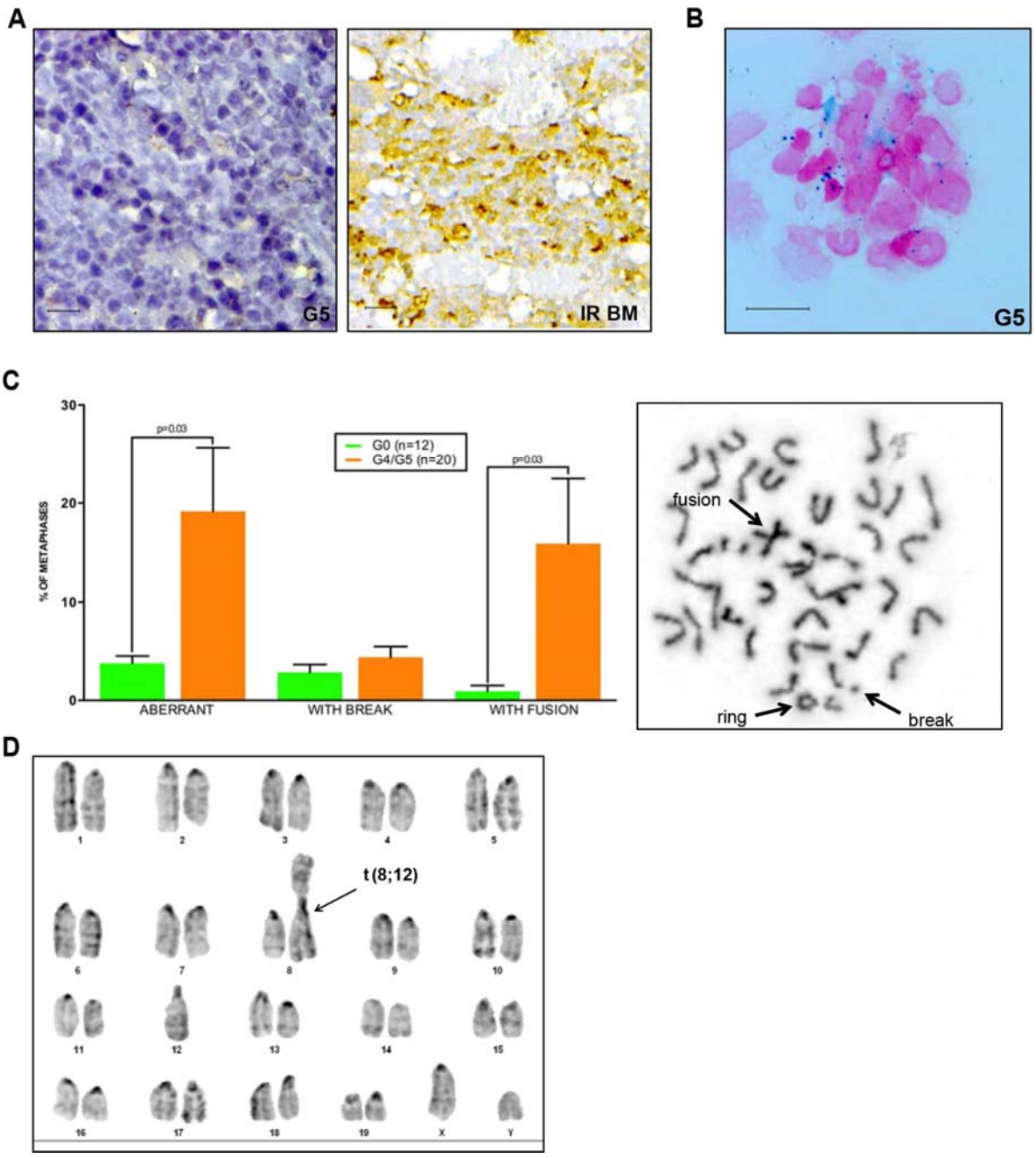


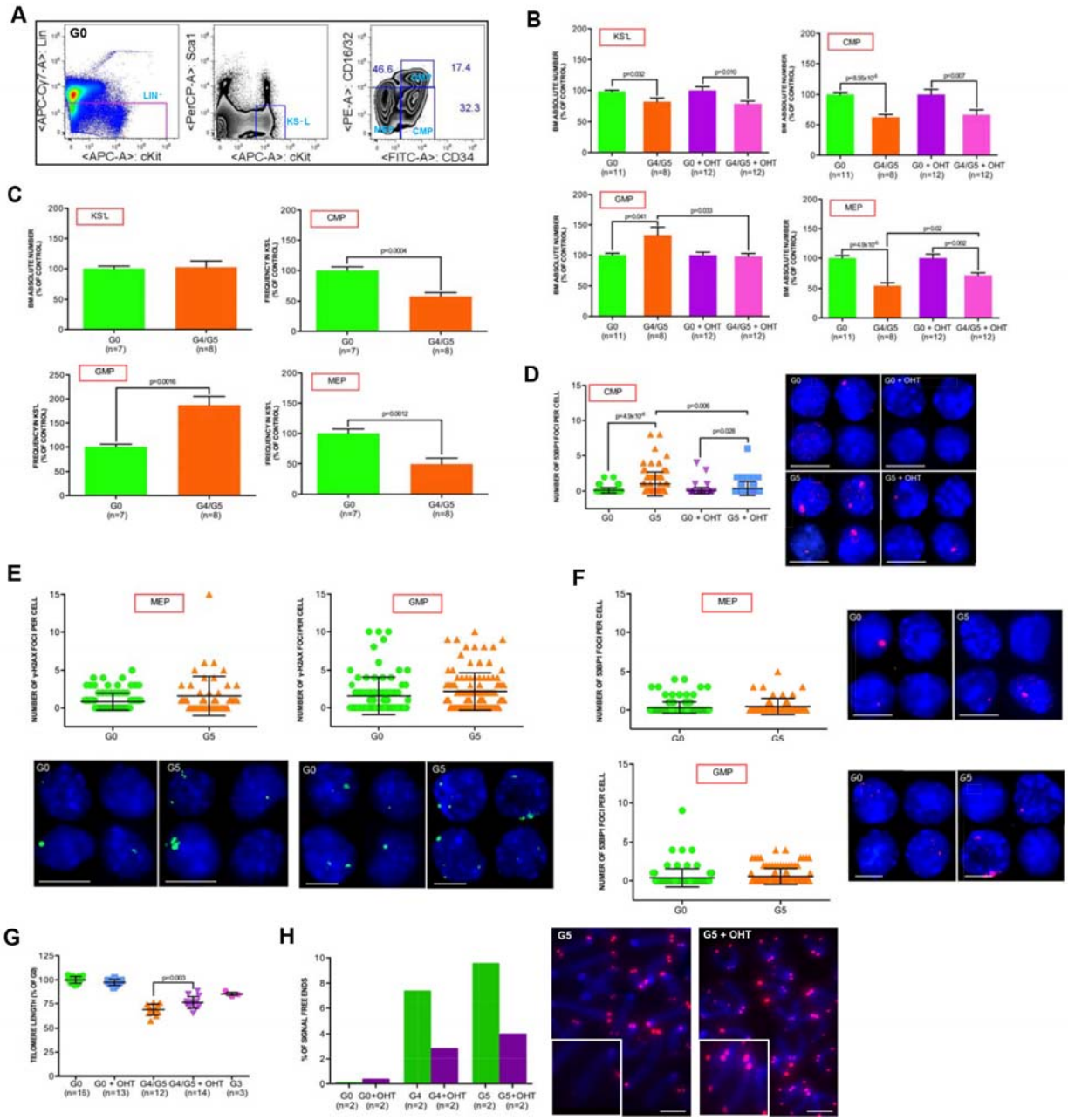
Figure S1, related Figure 1. The hematopoietic compartment of telomere dysfunctional mice recapitulates hallmark features of human myelodysplastic syndrome.

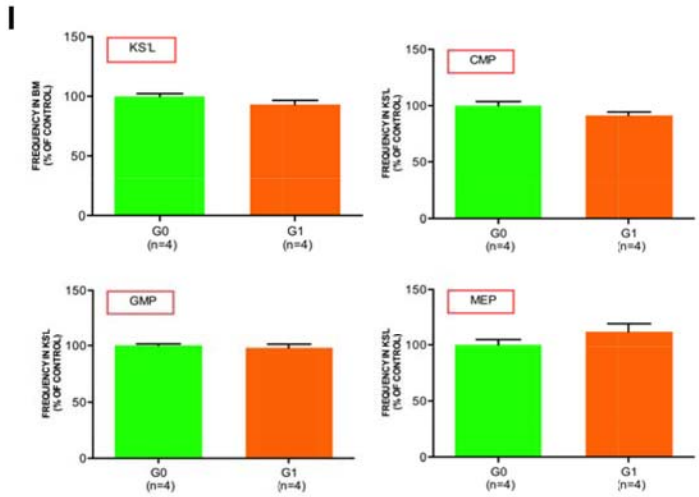
(A) Active caspase 3 immuno-stained section of a representative G5 BM biopsy (left panel) or irradiated (IR) wild type BM biopsy used as positive control (right panel) (scale bar, 15 mm).

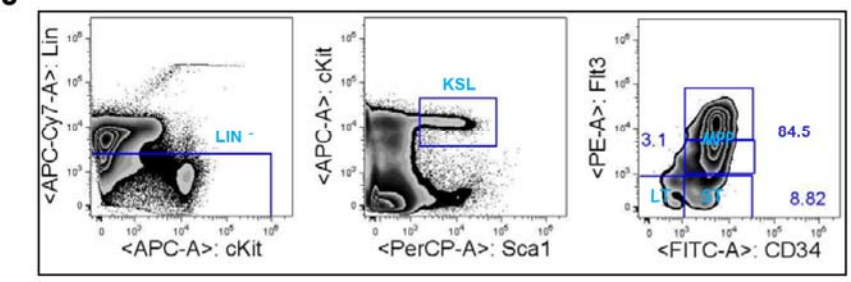
(B) Representative G5 BM cytospin stained with Prussian blue. Iron deposits do not surround the perinuclear membrane indicating the absence of ring sideroblasts (scale bar, 15 □m).

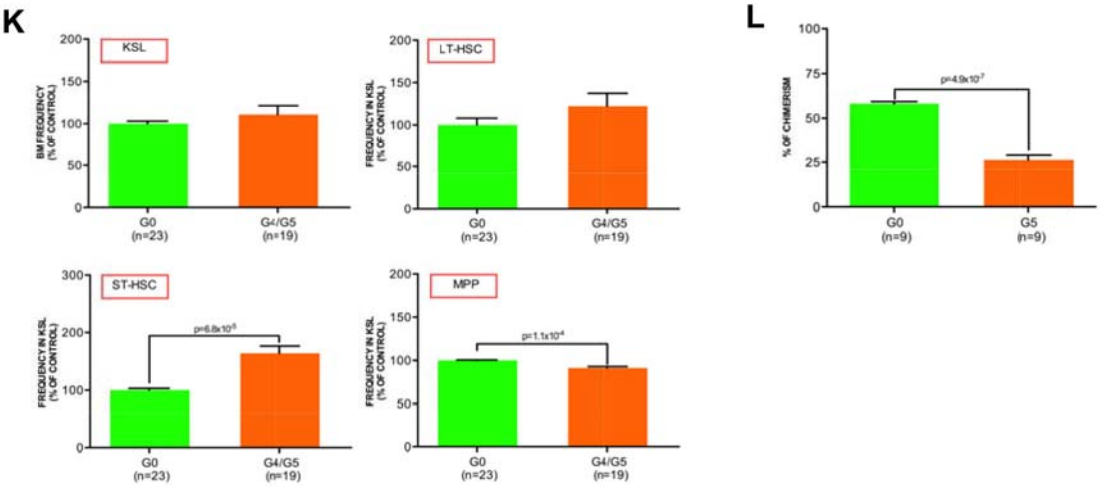
(C) Cytogenetic analysis of G4/G5 $TERT^{ER/ER}$ BM cells showed chromosomal breaks and fusions (on the left); representative G5 metaphase spread (on the right).

(D) Karyotype of a representative G5 mouse with a clonal robertsonian translocation t(8;12).


A
G0
<APC-Cy7-A>: Lin
<APC-A>: cKit
LIN-
<PerCP-A>: Sca1
KS-L
<PE-A>: CD16/32
<FITC-A>: CD34
46.6
17.4
32.3
MEP
CMP
B
KS'L
CMP
GMP
MEP
BM ABSOLUTE NUMBER (% OF CONTROL)
G0 (n=11)
G4/G5 (n=8)
G0 + OHT (n=12)
G4/G5 + OHT (n=12)
C
KS'L
CMP
GMP
MEP
FREQUENCY IN KS'L (% OF CONTROL)
G0 (n=7)
G4/G5 (n=8)
D
CMP
NUMBER OF 53BP1 FOCI PER CELL
G0
G5
G0 + OHT
G5 + OHT
E
MEP
GMP
NUMBER OF γ-H2AX FOCI PER CELL
F
MEP
GMP
NUMBER OF 53BP1 FOCI PER CELL
G
TELOMERE LENGTH (% OF G0)
G0 (n=15)
G0 + OHT (n=13)
G4/G5 (n=12)
G4/G5 + OHT (n=14)
G3 (n=3)
H
% OF SIGNAL FREE ENDS
G0 (n=2)
G0+OHT (n=2)
G4 (n=2)
G4+OHT (n=2)
G5 (n=2)
G5+OHT (n=2)
G5
G5 + OHT


I
KSL
CMP
GMP
MEP
FREQUENCY IN BM (% OF CONTROL)
FREQUENCY IN KSL (% OF CONTROL)
G0 (n=4)
G1 (n=4)


J
<APC-Cy7-A>: Lin
LIN -
<APC-A>: cKit
<APC-A>: cKit
KSL
<PerCP-A>: Sca1
<PE-A>: Flt3
3.1
MPP
84.5
8.82
<FITC-A>: CD34


K
KSL
LT-HSC
ST-HSC
MPP
BM FREQUENCY (% OF CONTROL)
FREQUENCY IN KSL (% OF CONTROL)
G0 (n=23)
G4/G5 (n=19)

L
% OF CHIMERISM
G0 (n=9)
G5 (n=9)

Figure S2, related to Figure 2. Skewed myeloid-erythroid differentiation of CMP is reversed by telomerase reactivation.

(A) Lineage negative (Lin^-), KS^-L, CMP, GMP and MEP representative profiles.

(B) KS^-L, CMP, GMP and MEP absolute numbers in the BM of indicated genotypes and treatments (mean and s.e.m. of age-matched 3 month-old mice from 3 independent experiments of telomerase reactivation *in vivo*; data are expressed as percentage of corresponding controls).

(C) KS^-L absolute number, as well as CMP, GMP and MEP frequencies in the KS^-L population of age-matched 7 month-old mice of indicated genotypes (data are expressed as percentage of the G0 control; error bars denote s.e.m).

(D) Anti-53BP1 immunofluorescence in CMP sorted from mice of indicated genotypes and treatments (α-53BP1: red; DAPI: blue; scale bar, 20 □m; n=4; right panel); numbers of 53BP1 foci per cell (left panel) (error bars denote s.d).

(E) Anti-γH2AX immunofluorescence in MEP (on the left) and GMP (on the right) sorted from mice of indicated genotypes (α-γH2AX: green; DAPI: blue; scale bar, 10 □m; n=4, bottom panel); numbers of γH2AX foci per cell (upper panel) (error bars denote s.d.).

(F) Anti-53BP1 immunofluorescence in MEP and GMP sorted from mice of indicated genotypes (α-53BP1: red; DAPI: blue; scale bar, 10 □m; n=4, right panel); numbers of 53BP1 foci per cell (left panel) (error bars denote s.d.).

(G) Mean value of telomere length in primary BM cells of indicated genotypes and treatments, as determined by flow-FISH analysis (mean and s.e.m of mice from 3 independent experiments of telomerase reactivation *in vivo*; data are expressed as percentage of the G0 control).

(H) Representative signal free ends in primary BM metaphases of indicated genotypes, 20 metaphases/sample; BM cells from two G0, one G4 and one G5 mice were analyzed after vehicle or OHT treatment (left panel); representative G5/vehicle or G5/OHT-treated BM metaphases (right panel) (scale bar, 1 □m).

(I) KS^-L frequency in the BM, as well as the CMP, GMP and MEP frequencies in the KS^-L population of G0 or G1 mice (mean and s.e.m. of age-matched 3 month-old mice; data are expressed as percentage of corresponding controls).

(J) Lineage negative (Lin^-), KSL, LT-HSC, ST-HSC and MPP representative profiles.

(K) KSL frequency in the BM, as well as the LT-HSC, ST-HSC and MPP frequencies in the KSL population of indicated genotypes (mean and s.e.m. of age-matched 3 month-old mice from 6 independent experiments; data are expressed as percentage of corresponding controls).

(L) Repopulation capacity of G0 or G5 LT-HSC was determined in experiments of competitive transplantation using CD45.1- and CD45.2-specific antibodies (mean and s.e.m. of n=9 recipient mice/ group). Shown is the percentage contribution of donor cells two□months after transplantation. Similar results were obtained 4 months after transplantation (data not shown).

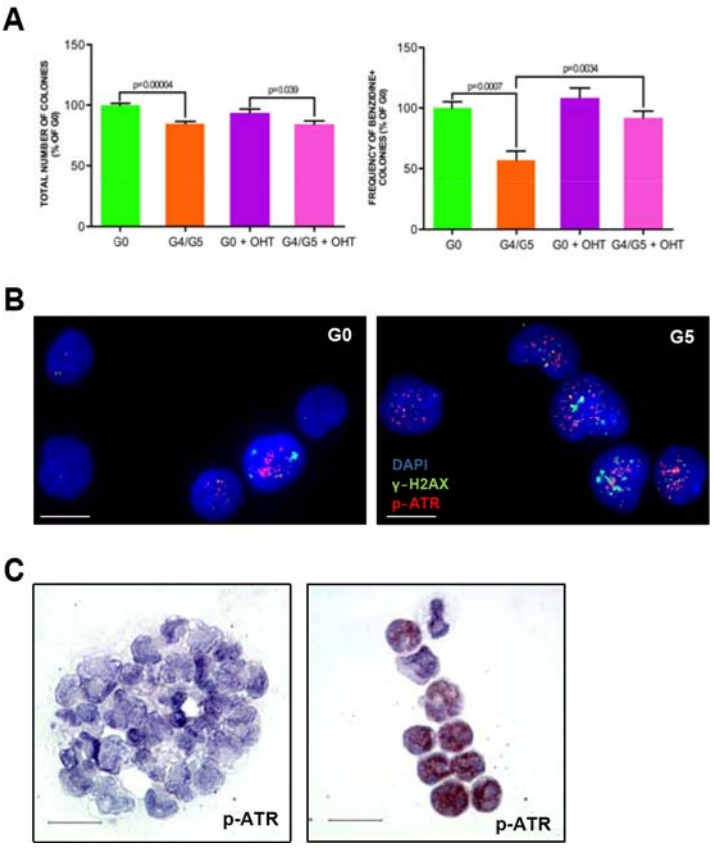

A
TOTAL NUMBER OF COLONIES (% OF G0)
150
100
50
0
p=0.00004
p=0.039
G0
G4/G5
G0 + OHT
G4/G5 + OHT
FREQUENCY OF BENZOINE+ COLONIES (% OF G0)
150
100
50
0
p=0.0007
p=0.0034
G0
G4/G5
G0 + OHT
G4/G5 + OHT
B
G0
G5
DAPI
γ-H2AX
p-ATR
C
p-ATR
p-ATR

Figure S3, related to Figure 3. Defective CMP differentiation is due to cell intrinsic DNA damage signaling activation.

(A) Clonogenic myeloid colony formation in methylcellulose from MNCs of indicated genotypes cultured in the presence of vehicle or OHT: total number of colonies (left panel) and frequency of benzidine positive colonies (right panel) (mean and s.e.m of replicates from 2 independent experiments; each experiment includes MNCs obtained from a pool of 3 mice of indicated genotypes; data are expressed as percentage of G0 control).

(B) Representative anti-p-ATR immunofluorescence in CMP sorted from G0 (on the left) or G5 (on the right) mice; α-γH2AX: green; α-p-ATR: red; DAPI: blue; (scale bar, 20 □m).

(C) Representative anti- p-ATR immunohistochemistry of $CD34^+$ cells isolated from a lower (on the left) or a higher risk MDS patient (on the right) (scale bar, 15 □m).

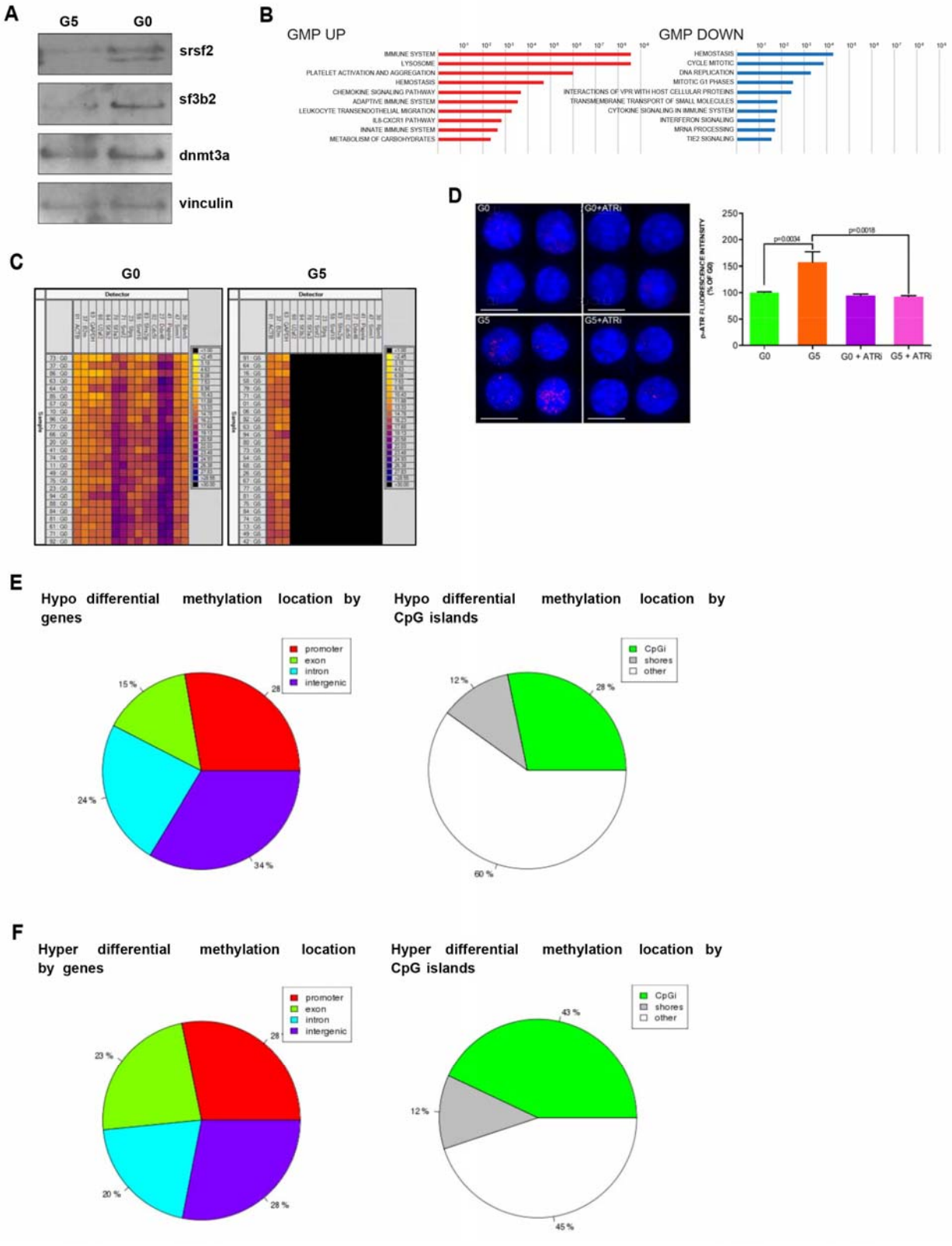

A
G5
G0
srsf2
sf3b2
dnmt3a
vinculin
B
GMP UP
IMMUNE SYSTEM
LYSOSOME
PLATELET ACTIVATION AND AGGREGATION
HEMOSTASIS
CHEMOKINE SIGNALING PATHWAY
ADAPTIVE IMMUNE SYSTEM
LEUKOCYTE TRANSENDOTHELIAL MIGRATION
IL8-CXCR1 PATHWAY
INNATE IMMUNE SYSTEM
METABOLISM OF CARBOHYDRATES
GMP DOWN
HEMOSTASIS
CYCLE MITOTIC
DNA REPLICATION
MITOTIC G1 PHASES
INTERACTIONS OF VPR WITH HOST CELLULAR PROTEINS
TRANSMEMBRANE TRANSPORT OF SMALL MOLECULES
CYTOKINE SIGNALING IN IMMUNE SYSTEM
INTERFERON SIGNALING
MRNA PROCESSING
TIE2 SIGNALING
C
G0
G5
Detector
Samples
D
G0
G0+ATRi
G5
G5+ATRi
p=0.0034
p=0.0018
pATR FLUORESCENCE INTENSITY (% OF G0)
250
200
150
100
50
0
G0
G5
G0 + ATRi
G5 + ATRi
E
Hypo differential methylation location by genes
promoter
exon
intron
intergenic
15 %
28 %
24 %
34 %
Hypo differential methylation location by CpG islands
CpGi
shores
other
12 %
28 %
60 %
F
Hyper differential methylation location by genes
promoter
exon
intron
intergenic
23 %
28 %
20 %
28 %
Hyper differential methylation location by CpG islands
CpGi
shores
other
43 %
12 %
45 %

Figure S4, related to Figure 4. Telomere dysfunction induces aberrant RNA splicing by repressing splicing gene expression in CMP.

(A) Representative western blot analysis of srsf2, sf3b2, dnmt3a and vinculin proteins in 25,000 CMP isolated from a pool of 4 G0 or G5 mice. To show the concomitant downregulation of multiple splicing factors and dnmt3a, protein detection was performed on the same G0 and G5 CMP sample.

(B) Significantly downregulated and upregulated pathways identified by GSEA in G4/G5 compared to G0 GMP (FDR =0.05).

(C) Fluidigm-based gene expression analysis of single cells (rows) for representative genes in the mRNA processing / spliceosome pathways (columns), which are significantly altered in sorted CMP from the G0 and G5 mice after LT-HSC transplantation in wild type mice (n=3). Genes analyzed were (from left to right): ACTB, β2m, GAPDH (housekeeping genes; internal controls), *U2AF2*, *SF3B2*, *SF3A3*, *SRSF2*, *SFPQ*, *SFRS10*, *SFRS2IP*, *CDC51*, *DDX46*, *PAPOLA*, *SRRM1* and *RBM5*. Donor cells were identified by staining with anti-CD45.2 antibody. Color scale on the right shows correspondence between color code and Ct values.

(D) Representative anti-p-ATR immunofluorescence in CMP sorted from G0 or G5 mice with or without ATR inhibitor treatment (α-p-ATR: red; DAPI: blue; left panel; scale bar, 20 □m); p-ATR fluorescence intensity in CMP sorted from G0 or G5 mice with or without ATR inhibitor treatment (mean and s.e.m of cells from a pool of 2 or 3 mice for each condition; data are expressed as percentage of G0 control; panel on the right).

(E) On the left**:** pie charts illustrating the proportion of differentially hypo-methylated CpGs on promoter regions (red), exons (green), introns (blue) and intragenic regions (magenta); on the

right: pie charts illustrating the proportion of differentially hypo-methylated CpGs on CpG islands (green), CpG island shore (grey), and other regions (white).

(F) On the left**:** pie charts illustrating the proportion of differentially hyper-methylated CpGs on promoter regions (red), exons (green), introns (blue) and intragenic regions (magenta); on the right: pie charts illustrating the proportion of differentially hyper-methylated CpGs on CpG islands (green), CpG island shore (grey), and other regions (white).

Table S1, related to Figure 4. List of probes down-regulated and up-regulated in G4/G5 CMP.

Table S2, related to Figure 4. List of genes validated by single cell Fluidigm Real Time-PCR.

Table S3, related to Figure 4. List of aberrantly spliced exons in G4/G5 as compared to G0 CMP ($p<0.05$).

Table S4, related to Figure 4. Predicted consequences of aberrant splicing on transcripts leading to frameshifts or loss of ATG / STOP codons, thereby impairing protein function.

Table S5, related to Figure 4. List of transcripts (with intact ATG and STOP codons) that are predicted to lose known protein domains due to aberrant splicing.

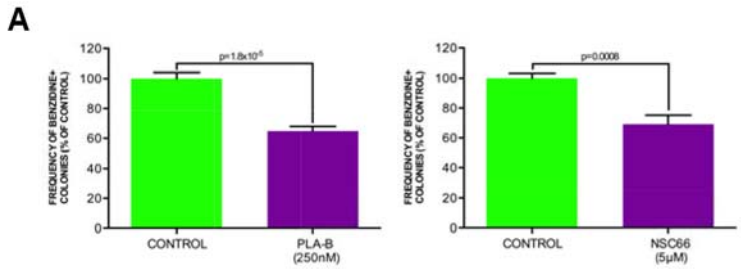


Figure S5, related to Figure 5. *SRSF2* haploinsufficiency induces skewed myeloid differentiation of CMP.

(A) Clonogenic myeloid colony formation in methylcellulose from sorted G0 CMP pre-treated with vehicle, Pladienolide B (PLA-B, 250 nM, left panel) or NSC663284 (NSC66, 5 μM, right panel) for 4 hr. Erythroid cells were scored by benzidine staining and expressed as frequency of the total number of colonies (mean and s.e.m of replicates from 2 independent experiments; each experiment includes CMP sorted from a pool of 3 wild type mice; data are expressed as percentage of vehicle control).

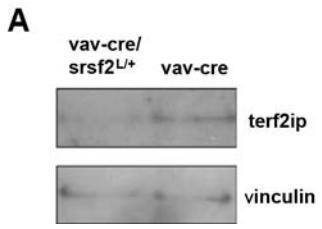


Figure S6, related to Figure 6. Aberrant RNA splicing due to reduced *SRSF2* expression induces telomere dysfunction.

(A) Representative western blot analysis of TERF2IP and vinculin proteins in 25,000 CMP isolated from a pool of 4 *Vav-cre* or *Vav-cre*/ *SRSF2*$^{L/+}$ mice.

Table S6, related to Figure 6. List of aberrantly spliced exons in *Vav-cre*/ *SRSF2*$^{L/+}$ as compared to *Vav-cre* CMP ($p < 0.05$).

Table S7, related to Figure 6. List of aberrantly spliced exons in CD34^{+} CMML cells with *SRSF2*(P95) mutation as compared to CD34^{+} CMML cells without mutation ($p<0.01$).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Generation and analysis of mice

Mice were maintained in specific pathogen-free (SPF) conditions at MD Anderson Cancer Center. All manipulations were performed with IACUC approval. The heterozygous (G0 $TERT^{ER/+}$) and late generation homozygous (G4/G5 $TERT^{ER/ER}$) mice were generated based on standard breeding protocol of successive generations of telomerase-deficient mice (Sahin et al., 2011). All studies were performed on adult (12-16 week old) G0 $TERT^{ER/+}$ and telomere dysfunctional G4/G5 $TERT^{ER/ER}$ mice, unless otherwise noted. OHT time-release pellets (2.5 mg; Innovative Research of America) were inserted subcutaneously to reach steady state blood levels of 1 ng ml^{-1} OHT.

Due to the variability in telomere length in the G4/G5 $TERT^{ER/ER}$ cohort, about 20 mice per group were used in the telomerase reactivation experiment for hematological characterization. Age- and gender-matched mice were randomly chosen for either placebo or OHT treatment. The *in vivo* experiments were not performed blinded. For mechanistic studies, G4/G5 $TERT^{ER/ER}$ mice with obvious defective hematopoiesis from CBC analysis were used.

The ATR inhibitor VE-821 (Selleck Chemicals) was dissolved in 0.5% methylcellulose and administered to the G0 or G5 mice by oral gavage at a concentration of 60 mg/kg. The drug was administered to the mice at a volume of 10 μl per 1 g body weight. Mice were sacrificed and analyzed 24 hours later.

The conditional deletion of *SRSF2* in the hematopoietic compartment was accomplished by crossing *Vav-cre* mice (Jackson laboratories) with the $SRSF2^{L/L}$ mice (Jackson laboratories) to generate heterozygous *Vav-cre*/ $SRSF2^{L/+}$ mice.

Peripheral blood samples were collected in EDTA-coated tubes and complete blood count (CBC) was performed with an automated Hemavet hematology analyzer through the Department of Veterinary Medicine and Surgery Histopathology Core at MDACC. Animals were autopsied, and the BM and spleen tissues were examined regardless of their pathological status. Tissue samples were fixed in 10% neutral-buffered formalin (Sigma) overnight, and washed once with 1× PBS and then transferred into 70% ethanol and stored at 4□°C. Tissues were processed by ethanol dehydration and embedded in paraffin according to standard protocols. Sections (3□μm) were prepared for antibody detection and haematoxylin and eosin staining.

Morphological evaluation and cytochemical stains

Cytospins of BM cells were prepared following standard procedure, fixed in methanol for 7 min and stained using the Giemsa-Wright method or Prussian blue. Cytochemical assays involved tests for myeloperoxidase (Sigma) and nonspecific esterase (α-naphthyl butyrate, Sigma) activity.

Histological analysis

Paraffin embedded BM and splenic sections were used for chromogenic immunohistochemistry, which was performed according to standard procedures. Antibodies used include myeloperoxidase (MPO, catalog number RB373R7, Thermo Scientific Lab Vision), CD11b (clone number EPR1344, Abcam), Cleaved Caspase-3 (catalog number CP229, Biocare Medical) and Ter119 (clone number TER-119, BD Pharmigen).

Cytogenetic analysis

Bone marrow single-cell suspensions isolated from 1 tibia were cultured overnight and treated with KaryoMAX Colcemid solution (Invitrogen) for 2□h before collection. Telomere fluorescence *in situ* hybridization (FISH) was performed on metaphase nuclei, as described previously(Maser et al., 2007). At least 20 metaphases from harvested cell cultures were analyzed for telomere integrity by telomere-specific peptide nucleic acid (PNA)-FISH. Telomere signal was normalized using a Pacific Blue centromeric PNA probe. Telomere length of the BM cells was also confirmed by using the DAKO Telomere PNA Kit/FITC for Flow Cytometry (DAKO). Relative telomere length (RTL) was determined by comparing isolated test cells with a control cell line (1301; subline of the Epstein–Barr virus (EBV) genome negative T-cell leukemia line CCRF-CEM, Sigma). Cells were STR profiled and tested mycoplasma negative.

Flow cytometry analysis

Single-cell suspensions were prepared from spleen and bone marrow (from femoral and tibial bones) by passing cells through pre-separation filters (Miltenyi). Cell numbers were subsequently counted. For FACS sorting and analysis we used described staining protocols and published stem and progenitor cell definitions (Amrani et al., 2011; Flach et al., 2014). Cells were acquired using LSR Fortessa (BD Bioscience) or sorted by Influx Cell Sorter (BD Bioscience) and analyzed using FlowJo software (Tree Star). Cell doublets were excluded from all analyses and dead cells were excluded by the use of DAPI. For splenic preparations, RBCs were lysed before antibody staining. Lin^- cells were enriched using a lineage cell depletion kit (Miltenyi) and an anti-CD127 (clone number A7R34) antibody conjugated to biotin. The following antibodies (all from eBioscience, BD Pharmingen or Biolegend) conjugated to FITC, APC, PE, PercpCy5.5, BV421, BV605 or APC-Cy7 were used for the flow cytometry analysis: -

TER119 (clone number TER-119), -CD71 (clone number R17217) -CD16/CD32 (clone number 93), -c-Kit (clone number 2B8), -Sca-1 (clone number D7), -CD34 (clone number RAM34),-Streptavidin, -Flk2 (clone number A2F10).

Absolute numbers were calculated based on BM cellularity of 4 hinds limbs, frequency of specific progenitor populations, and weight of individual animals to normalize for differences in body size, and presented as % of corresponding controls, as previously reported (Rossi et al., 2007).

In experiment of transplantation, donor and recipient cells were identified by staining with anti-CD45.2 (clone number 104) or CD45.1 (clone number A20) antibodies, respectively.

Indirect immunofluorescence microscopy and Tif assay

Sorted progenitor cells were resuspended in PBS, spotted on immunofluorescence slides (ThermoScientific), fixed for 30 min in 4% paraformaldehyde (Sigma-Aldrich), permeabilized in 0.2% Triton X-100 for 5 min and blocked in 5% Bovine Serum Albumin for 1 hour. Then, cells were stained with anti γH2AX (1:200; clone number JBW301, Millipore), 53BP1 (1:200, catalog number IHC00001, Bethyl Laborathories) or phospho-ATR (1:50; catalog number 2853, Cell Signaling) primary antibodies. Alexa Fluor-555 and Alexa Fluor-488 (Life Technologies) were used as secondary antibodies. Nuclei were stained with DAPI (Sigma-Aldrich). For every staining, the analysis was performed blinded with 4 mice of different genotypes or treatments.

For Tif assay, cells were co-stained with γH2AX and the telomere specific PNA probe using the Telomere PNA FISH Kit/Cy3 (DAKO, Glostrup, Denmark), according to the manufacturer's instructions. Coverslips were mounted with Prolong Gold Antifade reagent (Life Technologies).

Images were collected in Widefield Microscopy (Nikon Instruments Inc.) and analyzed using ImageJ software (http://rsbweb.nih.gov/ij/).

Treatments and colony forming assay

Mononuclear cells (MNCs) (20 x 10^4/replicate) or sorted CMP (500 cells/replicate) were seeded into cytokine supplemented methylcellulose medium (Methocult, M3434, Stem Cell Technologies). In separate experiments sorted CMP were pre-treated with a specific inhibitor of ATR (EMD Millipore 118510, 1 μM) or ATM (EMD Millipore 118500, 2 μM) for 1 hour, cisplatin (Sigma, 5 μM), pladienolide B (Bioaustralis Fine Chemicals, 250nM) or NSC663284 (Sigma, 5 μM) for 4 hours, washed and seeded in methylcellulose. In experiments of irradiation wild type mice were irradiated with γ radiation (3Gy) and CMP were sorted after 15 hours or 4 months in long-term experiments. In experiments of telomerase reactivation, MNCs or CMP were cultured in the presence of vehicle or OHT (Sigma, 100 nM). Colonies were counted after 7-10 days. Erythroid cells were scored by benzidine staining.

LTC-IC assay

Stromal layers were prepared from bone marrows isolated from wild type mice, cultured in 12.5% horse-serum IMDM supplemented with 12.5% FBS, penicillin, streptomycin, 1 μM hydrocortisone and 50 μM 2-mercaptoethanol at 37 °C. After 3 weeks, confluent stromal layers were trypsinized, irradiated (15 Gy) and subcultured in 6 multiwell plates. Cultures were then seeded with bone marrow MNCs (2.5 x 10^5 per well) in the presence of vehicle or OHT (100 nM). Cells derived from 5-week LTC-IC were seeded in cytokine supplemented

methylcellulose medium (Methocult, M3434, Stem Cell Technologies). Colonies were scored after 10 days.

Immunocytochemical analysis

Immunocytochemical analysis was performed in cytospin preparations of $CD34^+$ cells according to standard procedures. Cells were stained with anti-pATR antibody (1:50, catalog number 2853, Cell Signaling). Correlation between ATR phosphorylation and IPSS risk was evaluated using Fisher's exact test.

Quantitative real-time PCR (qPCR) in human MDS $CD34^+$ cells

After purification, $CD34^+$ cells were lysed with Trizol, according to the manufacturer's specifications for RNA extraction. Purified RNA was dissolved in sterile distilled water. cDNA was synthesized using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Carlsbad, CA), according to the manufacturer's protocol and using 200 ng of RNA per reaction. Q-PCR reactions were performed using TaqMan Expression Assays for *Srsf2* and *gapdh* genes (Applied Biosystems) and TaqMan Universal PCR Mastermix (Applied Biosystems) in a 7500 Real-Time PCR System (Applied Biosystems). Each condition was run in duplicate. Expression levels of *Srsf2* were normalized to those of *Gapdh*.

Single-cell gene expression profiling

Single CMP cells were sorted directly into 96-well plates in the Cells Direct Reaction Mix (Invitrogen). Individual cell lysis, cDNA synthesis, and amplification were performed according to Fluidigm Advanced Development Protocol, and single-cell microfluidic real-time PCR using

Dynamic Array IFCs (Biomark Fluidigm) was performed. Pre-amplified products (22 cycles) were diluted fivefold before analysis with Universal PCR Master Mix and pre-designed primers (DELTAgene Assays, Fluidigm) in 96.96 Dynamic Arrays on a BioMark System (Fluidigm). Primer sequences used in this study are available upon request. Ct values were calculated from the system's software (BioMark Real-time PCR Analysis; Fluidigm) and filtered according to a set of quality control rules. Gene filter: (1) for each gene, including controls, data with CtCall = FAILED and CtQuality < threshold were removed. (2) For each gene, including controls, genes with a difference of duplicate CtValues ≥ 2.0 were considered inconsistent and removed. Sample filter: (1) if the control genes (*Gapdh*, *B2m* and *Actb*) were not expressed or removed according to gene filters, the whole sample was removed. (2) If the mean of the Ct values of all genes including controls in a row was ≥ 27.0 the whole sample row was removed. To estimate for relative fold changes between control vs. G5 CMP, an arbitrary value of 30 is assigned if the intensity of the probe is read as "999" from the instrument.

Microarray and pathway analysis

Bone marrow CMP and GMP cells were sorted from 2 paired pools of G0 $TERT^{ER/+}$ or G4/G5 $TERT^{ER/ER}$ mice (5,000-20,000 cells per sample) using the Influx Cell Sorter. Every paired pool includes CMP and GMP sorted from 4 age and gender matched G0 or G4/G5 mice. RNA from the respective sorted cells was extracted using Trizol (Ambion) and profiled on 2100 Bioanalyzer (Agilent). Gene expression profiling was performed at the Sequencing and Non-coding RNA Program at MD Anderson Cancer Center. Briefly, the GeneChip® 3' IVT Express Kit (Affymetrix) was used to generate biotin-labeled cRNA, which were purified and fragmented, before target hybridization on the GeneChip® Mouse Genome 430 2.0 Array

(Affymetrix) according to the manufacturer's instructions. Affymetrix raw data (CEL files) were normalized using Affymetrix Microarray Suite (MAS) version 5.0. Due to the low number of replicates, differentially expressed genes were identified using an arbitrary fold change of 1.3 in paired pools. To identify pathways that were enriched in the sorted G4/G5 $TERT^{ER/ER}$ CMP and GMP cells, Gene Set Enrichment Analysis (GSEA; http://www.broadinstitute.org/gsea/msigdb/annotate.jsp) was performed on genes which were consistently upregulated or downregulated in both G4/G5 CMP samples when compared to the respective G0 controls in a pairwise manner.

RNA-Seq sequencing and analysis

Total RNA from sorted CMP isolated from 3 independent G0 $TERT^{ER/+}$ and 4 G4/G5 $TERT^{ER/ER}$, 3 *Vav-cre* and 6 *Vav-cre*/ $SRSF2^{L/+}$ mice, as well as CMML patient-derived $CD34^{+}$ cells (n=15) was isolated by Trizol (Ambion) and profiled on 2100 Bioanalyzer (Agilent). About 1 ng total RNA was amplified using the Ovation® RNA-Seq System V2 kit (Nugen), and libraries constructed using Ovation® Ultralow Library Systems and Ovation® Ultralow DR Multiplex System 1–8 (Nugen) according to the manufacturer's instructions. Transcriptomic sequencing (RNA-Seq) was performed on the Illumina HiSeq platform using the standard paired-end protocol. In total 60-160 million 76 base pair (bp) reads were generated per sample. An initial sequence-level quality assessment was performed using FastQC (version 0.10.1, Simon Andrews). The RNA-seq reads were then mapped to the mouse (NCBI Build 37.2) or human (GRCh37) reference genome using Tophat2, allowing a maximum of two mismatches per 75 bp sequencing end. The NCBI RefSeq gene model and HTSeq software (version 0.5.4p2, Simon Anders) were used to quantify the gene-level expression, exon-specific expression and intron

retention levels. The differential analyses for gene/isoform expression and intron retention were analyzed with DESeq2 (Anders and Huber, 2010), while exon usage was analyzed with DEXSeq (Anders et al., 2012). For an aberrantly spliced exon to be considered statistically significant relative to the controls, the criteria of $P < 0.05$ (in the G0 versus G4/G5 or *Vav-cre* versus *Vav-cre/Srsf2*$^{L/+}$ CMP analyses) or $P < 0.01$ (in the mutant *Srsf2* versus not mutant *Srsf2* CD34^{+} cell analysis) was employed. Pathway enrichment analysis was performed with Pathway Studio (version 10.2.2.9, Elsevier Inc.). Perl scripts were written to predict the effects of loss of exon on the protein sequences encoded. Domain prediction was performed using Pfam, and E-val <= 0.01 was considered significant.

Reduced representation bisulfite sequencing (RRBS)

Genomic DNA was extracted from sorted CMP populations isolated from 3 pools of G0 or 2 pools of G4/G5 mice using UltraPure™ Phenol:Chloroform:Isoamyl Alcohol according to manufacturer's instructions (Life Technologies). 14,000 to 30,000 cells were available for each sample, resulting in a minimum of 45ng of DNA. Genome-wide DNA methylation profiling was performed by RRBS. Library preparation and sequencing were performed at the UT MD Anderson Cancer Center's DNA Methylation Analysis Core and Sequencing and Microarray Facility, according to published protocols (Gu et al., 2011). RRBS sequencing data were aligned and methylation was called using Bismark v0.7.11(Krueger and Andrews, 2011). In brief, bisulphite-treated DNA was aligned to UCSC Genome Browser mm10 reference genome using Bowtie. In total 29-38 million reads were generated per sample with alignment rates around 63%. Next, MethylKit(Akalin et al., 2012) implemented with Fisher's exact test was used to compare the cytosine methylation profiles of G0 and G5 CMP. Gene promoter regions were

calculated based on RefSeq gene annotations with regions starting 1 kb upstream of the annotated transcription start site (TSS) and extending 500 base pairs downstream of TSS. Exons, introns, and CpG islands coordinates were collected from the UCSC Genome Browser mm10 version.

Western blotting

Western blotting in small amount of cells was performed as previously described (Nakada et al., 2010). Briefly, the same number of CMP (25,000 cells) purified from a pool of 4 G0 and G4/G5 or *Vav-cre* and *Vav-cre*/ *SRSF2*$^{L/+}$ mice was sorted into Trichloroacetic acid (TCA) and adjusted to a final concentration of 10% TCA. Extracts were incubated on ice for 30 minutes and spun down for 10 minutes at 16.1 rcf at 4°C. The supernatant was removed and the pellets were washed with acetone twice then dried. The protein pellets were solubilized with Solubilization buffer (9 M Urea, 2% Triton X-100, 1% DTT) before adding LDS loading buffer. Proteins were separated on a Bis-Tris polyacrylamide gel (Invitrogen) and transferred to a PVDF membrane Antibodies were anti-sfrs2 (clone number 1SC-4F11, Millipore), anti-sf3b2 (clone number 5D2, Sigma), anti-dnmt3a (clone number H-295, Santa Cruz), anti terf2ip (clone number D9H4, Cell Signaling) and anti-vinculin (clone number hVIN-1, Sigma).

Bone Marrow Transplantations

Experiments of BM transplantation were performed as previously described (Sahin et al., 2011). Briefly, CD45.1+ recipient mice (Jackson Laboratories) were irradiated with a total of 10.5 Gy γ-radiation (5 Gy and 5.5 Gy 3 hours apart) on the day of transplantation. BM transplants were performed using purified LT-HSC (1,500 cells / mouse) isolated from a pool of 3 G0 or

G5 $TERT^{ER/ER}$ (all CD45.2+) donor mice. Competitive blood repopulation was performed by mixing the specified CD45.2 cells together with nucleated bone marrow cells prepared from congenic CD45.1 mice. Specifically, each recipient received 1,500 purified LT-HSC derived from a pool of 3 G0 or G5 $TERT^{ER/ER}$ mice together with 3 x 10^5 CD45.1 competitor bone marrow cells. Donor derived peripheral blood reconstitution (i.e. chimaerism) was assessed after 2 or 4 months following transplantation by FACS analysis of nucleated peripheral blood cells stained with anti-CD45.1 and anti-CD45.2-specific antibodies. Blood chimaerism for each recipient was calculated as the percentage of all CD45+ cells that were CD45.2+.

Statistical analysis

All the cytogenetics, indirect immunofluorescence microscopy, immunohistochemistry, and Fluidigm PCR analyses were performed blinded without the investigator knowing the sample annotation and outcome.

All the data were analyzed by a two-tailed Student's *t*-test ($p<0.05$ is considered to be statistically significant). For all experiments with error bars, standard error mean was calculated to indicate the variation within each experiment and data, and values represent mean ± s.e.m or mean ± s.d., as indicated in the figure legends. Variance within each group of data was not evaluated and final data was analyzed by pooling biological replicates from different experiments.

SUPPLEMENTAL REFERENCES

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology *13*, R87.

Gu, H., Smith, Z.D., Bock, C., P., B., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nature Protocol *6*, 468-481.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Maser, R.S., Wong, K.K., Sahin, E., Xia, H.L., Naylor, M., Hedberg, H.M., Artandi, S.E., and DePinho, R.A. (2007). DNA-dependent protein kinase catalytic subunit is not required for dysfunctional telomere fusion and checkpoint response in the telomerase-deficient mouse. Molecular and Cellular Biology *27*, 2253-2265.