SUPPLEMENTAL METHODS AND MATERIALS

Generation of mice carrying the *MGS34F* or *IES34F* alleles

We have used two strategies to create conditional "knock-in" alleles of *U2af1*(S34F) in the mouse germ line. The two approaches involve targeting the normal locus with vector DNA that carries the S34F mutation in the context of a MiniGene (MGS34F) or an Inverted Exon (IES34F). Both alleles can be rearranged by the Cre recombinase to allow production of U2af1(S34F) protein from the native *U2af1* locus. As shown in Fig. 1A, the targeting vector for *MGS34F* contains a WT minigene (exons 2 – 8, NM_024187), a transcriptional "STOP" signal (3 repeats of SV40 late poly A sequence), an Frt-Neo-Frt cassette, and a mutant Exon 2 carrying the S34F missense mutation (TCT > TTT). One LoxP sequence precedes the WT minigene, 470 bp upstream; another LoxP sequence precedes the mutant Exon 2, 475 bp upstream. As shown in Fig. S2A, the targeting vector for *IES34F* contains an Frt-Neo-Frt cassette and an inverted mutant Exon 2 carrying the S34F missense mutation, both of which are located in Intron 2 of *U2af1*. A pair of LoxP sequences and a pair of LoxP 2272 sequences were inserted flanking and in between the WT and mutant versions of exon 2 so that mutant exon 2 replaces WT exon 2 after Cre-mediated recombination. Both vectors contain 5' and 3' homology arms (roughly 5 kb for each arm), and a diphtheria toxin A (DTA) cassette was added at the 3' end of the 3' homology arm to allow negative selection. The homology arms were made through recombination-mediated genetic engineering (or recombineering), using a PI-based artificial chromosome (PAC) clone RP21-477O4 (Strain: 129S6/SvEvTac) as the target DNA. All other elements mentioned above were cloned through PCR-based methods. The targeting vectors were constructed by Biocytogen (Worcester, MA), and their identities were verified by restriction enzyme digestion and Sanger sequencing.

Mouse embryonic stem (ES) cell lines and mice carrying either of the targeted alleles were established at the Embryonic Stem Cell and Transgenic Mouse Core facility (NHGRI). The mouse ES cell line, named HG3, was a B6/129 F1 hybrid, derived from a cross between a B6 male homozygous for a UBC-GFP transgene (C57BL/6-Tg(UBC-

GFP)30Scha/J; Jackson Laboratory (JAX) stock number: 004353) and a WT 129 female mouse (129S6/SvEvTac. Taconic model number: 129SVE-F). 100 μg of the targeting vector was linearized by digestion with Asc I, the recognition site for which is located immediately upstream of the 5' homology arm on the targeting vectors. Linearized DNA was electroporated into HG3 cells (passage 7). Following G418 selection, the surviving cell clones were expanded for Southern blot analysis to identify those clones that had undergone correct homologous recombination (Figs S1A and S2B) and with normal karyotypes. Positive ES cell clones were then microinjected into C57BL/6J blastocysts. Resulting chimera mice with a high percentage of GFP positive cells were used to mate with B6 WT mice (C57BL/6J; JAX stock number: 000664) to test for germline transmission. The genotypes of F1 mice were determined by PCR amplification, using the primers U2af1-A1LoxP-F and U2af1-A1LoxP-R (Table S2), and the identities of the PCR amplicons were verified by Sanger sequencing. Tail DNA samples carrying the targeted allele were further confirmed by Southern blot analysis (Figs S1A and S2B). F1 mice carrying the targeted allele were crossed with mice expressing *Flp* (B6.Cg-Tg(ACTFLPe)9205Dym/J. JAX stock number: 005703) to remove the Neo cassette. Removal of the *Neo* cassette in the progeny mice was further confirmed by PCR-based genotyping (Table S2). Flanking sequences surrounding the remaining Frt site in the targeted alleles, as well as the boundaries of the homology arms, were verified by Sanger sequencing. Mice (*MGS34F/WT; Flp* or *IES34F/WT; Flp*) were crossed with B6 WT mice, and/or B6 mice carrying *UBC-CreERT2* (JAX stock number: 008085), *Mx1-Cre* (JAX stock number: 005673), and/or $Runx1^{F}$ (1) alleles as specified in the Results section for desired genotypes.

Mouse procedures

Peripheral blood sampling and CBC analysis: Blood was collected by either submandibular or retro-orbital bleeding. The former procedure was done as described (2), except that a 20 G needle instead of a lancet was used to enter the facial vein. The latter procedure involves collecting blood through heparinized microhematocrit capillary tubes (Thermo Fisher Scientific). Freshly collected blood was mixed immediately with

EDTA in an EDTA-coated microfuge tube (Sarstedt) before performing complete blood counts (CBC) and other tests. Automatic CBC was conducted at the Department of Laboratory Medicine (NIH Clinical Center) as a paid service. CBC findings were further confirmed by reviewing the corresponding blood smears.

Bone marrow and spleen cell transplantation: Bone marrow was flushed from freshly dissected long bones (femur and tibia) with Hank's Balanced Salt Solution (HBSS) supplemented with 2% FBS, using a 3 ml syringe attached to a 25 G needle. Cell suspensions were gently passed through the 25G needle 2 - 3 times and then passed through a 70 μm cell strainer. For single cell suspensions of spleen, fresh tissue was placed in a 70 μm cell strainer and pushed through a nylon membrane by a rubber stopper from a 3 ml syringe. The cell strainer was rinsed with HBSS with 2% FBS to collect cell suspensions. Cell viability was assessed by trypan blue staining. Nucleated cells were counted with a hemocytometer by mixing cell suspensions with a 3% acetic acid solution containing methylene blue (Stemcell Technologies). Young B6/129 F1 hybrid mice (6 - 16 weeks, either sex), with or without the CD45.1 allele (JAX stock number: 002014), were used as transplant recipients. Irradiation was performed with a Cesium 137 irradiator, about four hours before bone marrow or spleen cell transplantation via the tail vein. The lethal irradiation dose was 1000 cGy; the sub-lethal irradiation dose was 650 cGy. For non-competitive transplants (Fig. S5), 2 or 2.5 million nucleated bone marrow cells (in 200 μl HBSS supplemented with 2% FBS and 20 U/ml heparin) were transplanted to lethally irradiated recipient mice. For competitive transplants (Fig. 3), equal numbers of donor and competitor cells were mixed; 1 million nucleated cells (in 200 μl HBSS supplemented with 2% FBS and 20 U/ml heparin) were transplanted to lethally irradiated recipient mice. For transplantation of bone marrow or spleen cells in Figs. 4 and S9, 1 million nucleated cells were used for each sub-lethally irradiated recipient. After lethal irradiation and transplantation, mice were housed in sterile cages and fed with sterilized mouse chow and acidified water *ad libitum* for up to four weeks before being returned to normal housing and feeding.

Drug treatment: Poly (IC) (high molecular weight, Invivogen) was reconstituted in saline to 1 mg/ml according to vendor's instruction. Poly (IC) (three doses of 250 μg per

mouse) was given every other day via intraperitoneal injection. ENU treatment was as previously described (3). Briefly, ENU (Sigma-Aldrich) was dissolved in pure ethanol at 100 mg/ml, further diluted in phosphate/citrate buffer to a final concentration of 5 mg/ml and given intraperitoneally at 100 mg/kg, one week after the last dose of poly (IC). BrdU (1 mg/mouse at 10 mg/ml, BD Biosciences) was given intraperitoneally 2 hr before euthanasia.

Moribund criteria: The health of the mice were monitored by technicians who were blinded to the genotypes and treatments of the mice. Mice showing signs of pale paws, hunched posture, lethargy, dyspnea, rough hair coat, and severe weight loss were euthanized by an overdose of carbon dioxide, followed by cervical dislocation.

Generation of mouse embryo fibroblasts (MEFs): MEFs were generated from E13.5 embryos, the products of timed mating between *MGS34F/WT* or *IES34F/WT* male mice and UBC-CreERT2 hemizygous female mice. Freshly dissected mouse embryos (without the placenta, head, and internal organs) were minced by a sterile single blade and then transferred to a 15 ml conical tube containing 2 ml 0.125% trypsin (0.25% trypsin diluted 1:1 with DPBS). Cells in tissue culture medium were incubated at 37 degrees Celsius for 15 min with gentle shaking. At the end of incubation, 12 ml growth media (DMEM supplemented with 10% FBS and penicillin/streptomycin) were added to the tube to inactivate trypsin. The entire culture was then transferred to a 100 mm dish and incubated at 37 degrees Celsius with 5% CO_2. The growth medium was changed the next day. The head of the embryo was used for genotyping. MEFs were passaged before they reached 100% confluence, and cells passaged fewer than five times were used for all assays.

Histology

The nuclear features of bone marrow and spleen cells were evaluated using cytospin preparations. 1 - 5 x 10^5 freshly prepared bone marrow or spleen cells (in 100 µl - 200 µl DPBS with 5% FBS) were applied to a slide chamber and spun onto glass microscope

slides (pre-wetted with DPBS with 5% FBS) using a Shandon Cytospin 2 centrifuge at 500 g for 5 min with low acceleration. Slides were air dried before Wright-Giemsa staining, which was performed either by the Department of Laboratory Medicine (NIH Clinical Center) as a complimentary service or by the researchers using the PROTOCOL Hema 3 stain set. Stained slides were then sealed with cover glass and Cytoseal 60 (Electron Microscopy Sciences).

Nuclear and cytoplasmic features of peripheral blood, bone marrow, and spleen cells were also evaluated using smear or "touch" preparations. Blood smears were prepared using an Autoprep (Dynamic Diagnostics). Bone marrow smears were prepared by first cutting freshly dissected long bones longitudinally along the bone cavity to expose the bone marrow. A 5-O camel's hair brush was wetted with hypotonic FBS (2:1 (v/v) FBS:distilled water) and used to make a "slurry" of the bone marrow. The suspended bone marrow cells on the brush were then "painted" onto glass microscope slides. Spleen "touch" preparations were made by touching the freshly cut surface of mouse spleens onto glass slides, which were air dried and stained with Wright-Giemsa as described above.

Evidence of megakaryocyte clustering, excessive extramedullary hematopoiesis, and blast tissue infiltration were examined using tissue sections. Fresh tissues (bone, spleen, liver, lung, heart, etc.) were dissected and fixed in 10% neutral buffered formalin (VWR) for at least 24 hr. Tissue processing (bone decalcification, tissue dehydration, paraffin-embedding and sectioning at 5 μm) and H&E staining were performed as a paid service at Histoserv Inc (Germantown, MD) or at the Laboratory of Comparative Pathology (WCM). Immunohistochemistry was performed on bone sections using a rabbit anti-CD61 polyclonal antibody (Novus #NBP1-83453. 1:500 dilution) and developed with DAB (3,3'-diaminobenzidine) as a paid service at the Laboratory of Comparative Pathology (WCM).

Histological slides were examined by a board-certified anatomic and clinical pathologist and hematopathologist (BD), who was blinded to mouse genotypes and treatment at the

time of review. Representative images were acquired using an Axio Observer A1 microscope (Carl Zeiss) at 400x or 600x magnification.

Assays and reagents

Southern blot: Southern blot was performed as previously described (4). Briefly, high molecular weight genomic DNA was extracted from tissues, cultured cells, or mouse tails using a lysis buffer (10 mM Tris, pH 8.0, 100 mM NaCl, 10 mM EDTA, 0.5% SDS, 500 μg/ml protease K), followed by phenol/chloroform extraction, isopropanol precipitation, washing with 70% ethanol, and resuspension in TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA). Up to 25 μg DNA was digested with the appropriate restriction enzymes (see the diagrams in Figs 1A and S2A), separated on a 1% agarose gel, and transferred onto positive-charged nylon membranes (Ambion) using an alkaline buffer (0.4 M NaOH, 1 M NaCl). The membranes were blocked with ExpressHyb Hybridization Solution (Clonetch). The 5', 3', and internal probe DNA fragments were ^{32}P-labeled by the Ready-To-Go DNA Labelling Beads (-dCTP) (GE Healthcare) and dCTP [α-^{32}P] (Perkin Elmer), and incubated with the nylon membranes at 65 degrees Celsius. Radioactive signals were detected with a Phosphorimager (Fuji).

RNA and RT-qPCR: RNA was extracted using Trizol (Molecular Research Center or Invitrogen) combined with a column “cleanup” step using RNeasy (Qiagen). cDNA was synthesized from RNA (up to 2 μg) using a High-Capacity cDNA Reverse Transcription (RT) kit (Applied Biosystems). Isoform-specific primers (Table S2) were used to quantify the changes in alternative splicing or polyadenylation. The specificity and efficiency (> 90%) of each primer set were determined as described before (4). Quantitative PCR (qPCR) was performed using either the 7900HT real-time PCR system (Applied Biosystems) with the USB VeriQuest SYBR Green qPCR Master Mix (Affymetrix) or the QX200 Droplet Digital PCR system (ddPCR, Bio-Rad) with the EvaGreen Supermix (Bio-Rad). The RQ Manager software (v1.2, Applied Biosystems) was used for quantifying qPCR results from 7900HT using a relative quantification method (ΔΔCt). For ddPCR analysis, the concentration of each isoform was calculated using the

QuantaSoft Software (Bio-Rad) and results were presented as percent inclusion of the indicated exon or portion of an exon. The allele-sensitive S34F/WT SNP Taqman assay was custom synthesized (Applied Biosystems) and characterized in Fig. S3C - S3E. The plasmids for calibrating the assay carry modified versions of the *U2af1*(WT) or *U2af1*(S34) cDNA, as described previously (4).

Genotyping PCR: Routine genotyping to detect the *MGS34F* and *IES34F* alleles was done by conventional semi-quantitative PCR using the Hotstar Taq master mix (Qiagen), the primers U2af1-A1LoxP-F and U2af1-A1LoxP-R (750 nM final concentration for each primer. Table S2), and tissue lysate (1 µl) in a 10 µl reaction. Tissue lysate was prepared by heating a piece of mouse tail (or the head of an embryo) in 500 µl lysis buffer (50 mM NaOH, 2 mM EDTA) for 30 min at 95 degrees Celsius, followed by neutralization with 10 µl Tris (1 M, pH 8.0). PCR condition: 95 degrees Celsius for 15 min, followed by 35 cycles at 94 degrees Celsius for 20 sec, 62 degrees Celsius for 30 sec, and 72 degrees Celsius for 30 sec. All other alleles were genotyped according to the vendors' protocols or as previously described (1).

Fragment analysis: Fragment analysis was used as an alternative method to Southern blotting to measure the extent of Cre-mediated recombination (Figs S3A and S3B). PCR was performed using the Hotstar Taq master mix (Qiagen), the primers U2af1-78, U2af1-79, and M13F-FAM (100 nM final concentration for each primer. Table S2), and 1 µl genomic DNA extracted by a DNeasy kit (Qiagen) in a 10 µl reaction. PCR conditions: 95 degrees Celsius for 15 min followed by 30 cycles of 94 degrees Celsius for 20 sec, 56 degrees Celsius for 60 sec, and 72 degrees Celsius for 60 sec. The final extension step was done at 72 degrees Celsius for 5 min. PCR products were loaded onto a 3730 DNA analyzer for high resolution capillary electrophoresis as described previously (5), performed at the Genomics Core (NHGRI). Recombination efficiency (RE) was calculated by the formula: $FI^{S34F}/(FI^{MGS34F} + FI^{S34F})$. FI, fluorescence intensity.

Fluorescent in situ hybridization (FISH): FISH was conducted at the Cytogenetics and Microscopy Core (NHGRI), as described (6). The GFP probe was made from a

fragment released from the pEGFR-C1 vector digested with Nhe I and EcoR I (New England Biolabs).

Flow cytometry: Antibodies for detection of lineage markers included those against CD3ε, CD4, CD8a, B220, IgM, IL-7Rα, NK1.1, Mac1, Gr-1 and TER119. All antibodies used for flow cytometry were purchased from BD Bioscencies or BioLegend, with targets that included: B220 (clone RA3-6B2), c-Kit (clone 2B8), CD150 (clone TC15-12F12.2), CD19 (clone 6D5 or 1D3), CD3ε (clone 145-2C11), CD4 (clone RM4-5), CD45 (clone 30-F11), CD48 (clone HM48-1), CD8a (clone 53-6.7), Gr-1 (clone RB6-8C5), IgM (clone RMM-1), IL-7Ra (clone A7R34), Mac-1 (clone M1/70), NK1.1 (clone PK136), Sca-1 (clone D7) and TER119 (TER119). Dead cells were marked by Aqua LIVE/DEAD Fixable Dead Cell Stain (Invitrogen). Intracellular staining for BrdU was performed using an APC BrdU Flow Kit according to the vendor's instruction (BD Pharmingen). Flow cytometry was performed with a three-laser (405 nm, 488 nm, 633 nm) LSR II (BD Biosciences) at the Flow Cytometry Core (NHGRI). Single color staining was used to measure spectral overlap and calculate fluorescence compensation, using the Diva software (BD Biosciences). Fluorescence-minus-one staining was used for proper gating of cell populations. FlowJo (v9.9 or v10.1, FlowJo LLC) was used for all data analysis.

High throughput nucleotide sequencing

mIMPACT gene panel sequencing: Integrated Mutation Profiling of Actionable Cancer Targets, mouse version 1 (mIMPACT) consists of a panel of 578 mouse genes homologous to human cancer-related genes (7), and was conducted and analyzed at the MSKCC Sequencing Core Facility. A pre-hybridization library was prepared using the KAPA HTP protocol (Kapa Biosystems). Targets were enriched by in-house custom DNA probes and were sequenced in HiSeq4000 to generate paired-end 126 bp reads.

Bioinformatics

Differential gene expression and splicing patterns were analyzed as described (4, 8). Briefly, reads were first mapped to known mouse or human transcripts using RSEM (9). Remaining unaligned reads were then mapped to a database of possible junctions between all 5' and 3' splice sites of relevant genes and then subsequently to the GRCm38 mouse genome assembly, using TopHat (10). The number of “reads” were then subjected to analysis by the Wagenmakers’s Bayesian alternative to the binomial proportion test (11). Differential splicing was defined as a difference in absolute isoform ratio of at least 10%, based on at least 20 relevant reads in the comparator samples, using MISO v2.0 (12). Differentially expressed genes were defined as those that exhibited at least a two-fold difference in abundance of RNA (mutant vs. WT).

Sequences logos were created using the seqLogo package from Bioconductor (13). The invariant AG at 3' splice sites was not plotted to scale to allow highlighting of the consensus nucleotides at the -3 position. Unsupervised clustering was performed using Ward’s method.

Relative levels of expression of WT and mutant alleles were estimated from the mapped mRNA sequencing data in the IGV browser (v2.3.72, Broad Institute). The Sashimi plot for *Tet2* mRNA (Fig. 5C) was also generated using the IGV browser.

Analysis of targeted gene panel sequencing (mIMPACT, Table S7) was described previously (7). The DNA-level variants were predicted using the Variant Effect Predictor (Ensembl). The variants were further filtered based on the following criteria: (1) they predict a change in the protein sequence (e.g. mis/non-sense mutations, splice-site mutations); (2) they are not in dbSNP (version 146); and (3) they are not seen in matched control samples.

LEGENDS FOR SUPPLEMENTAL FIGURES AND TABLES

Supplemental Figure S1. Additional data related to Fig. 1. (**A**) Germline transmission of the *MGS34F-Neo* allele. Southern blots of EcoR I and BamH I digested

tail DNA from three engineered mice (1, 2, 3) and of DNA from a WT mouse tail and the embryonic stem cell clone from which the mouse line originated (ES) are shown. (**B** - **D**) *U2af1*(S34F) expression and splicing alteration after Cre-mediated recombination within the *MGS34F* allele in MEFs. Panel **B**, Cre-mediated recombination documented by Southern blot using BamH I digested genomic DNA from MEFs, which produce 4OHT-dependent Cre. MEFs were treated with vehicle or escalating doses of 4OHT for 24 hr, then incubated for 48 hr without the drug. Treatment was repeated for some samples (x2) before harvest. Panel **C**, the percentages of *U2af1* RNA derived from the *U2af1*(S34F) allele in MEFs were quantified by an allele-sensitive S34F/WT Single Nucleotide Polymorphism (SNP) Taqman assay (See Fig. S3C - S3E). Panel **D**, mRNA from MEFs with or without *U2af1*(S34F) was sequenced to determine nucleotides at the 3' splice acceptor sites of cassette exons and displayed as sequence logos according to whether inclusion of the exon in mRNA was increased, decreased, or unaffected by U2af1(S34F). The MEFs were treated with 4OHT (500 ng/ml) or vehicle twice as in panels B and C. The consensus sequences at 3' splice sites in mouse genome are similar as those in human genome with U2AF1(S34F) as discussed in the Result section. (**E** - **G**) *U2af1*(S34F) expression after Cre-mediated recombination within the *MGS34F* allele in total bone marrow cells. Panel **E**, Cre-mediated recombination documented by Southern blot using BamH I digested genomic DNA from total bone marrow cells from mice with the indicated genotypes (n = 3 for each genotype). The mice were treated with poly (IC), then euthanized two weeks later for DNA extraction. Panel **F**, the percentages of *U2af1* RNA derived from the *U2af1*(S34F) allele (overall total *U2af1* RNA) were quantified by an allele-sensitive S34F/WT SNP Taqman assay (Fig. S3C - E). Panel **G**, U2af1(S34F) does not alter choice of polyadenylation sites for *Atg7* mRNA in total bone marrow cells. The same RNA samples were used to quantify splicing changes from other gene loci in Fig. 1B. Panel **H**, a GFP transgene is present on chromosome 17, the same chromosome where *U2af1* is located, in the embryonic stem cells (HG3) that were used to create mice with the *MGS34F* or *IES34F* alleles. Shown is a representative metaphase HG3 cells by fluorescence in situ hybridization (FISH). The fluorescence signal for GFP is in red (boxed). Chromosomes 17, where

U2af1 resides and colored in green, were marked by whole-chromosome painting probes. All chromosomes were stained with DAPI and colored in blue.

Supplemental Figure S2. Generation of mice and cells carrying the *IES34F* allele, and molecular characterization. (**A**) Diagram of *IES34F* and related alleles, and the strategies for Southern and PCR genotyping. The mouse line was first established to carry the *IES34F-Neo* allele, and was further crossed with mice expressing *Flippase* (*Flp*) to remove the *Neomycin* (*Neo*) cassette. Numbered boxes, exons; lines, introns; E, EcoR I site; B, BamH I site; A, Ase I site; Bs, BspH I site. Exon 2 in red carries the S34F missense mutation and was inserted in reverse orientation in the *IES34F-Neo* and *IES34F* alleles. (**B**) Germline transmission of the *IES34F-Neo* allele by Southern blot using tail DNA from the mice. Results for two positive mice (94 and 97) were shown. WT, a tail DNA sample from a WT mouse; ES, the embryonic stem cell clone that made the mouse line. (**C**) Cre-mediated recombination for *IES34F*. Southern blot was performed using genomic DNA from MEFs carrying *IES34F/WT* and UBC-*CreERT2*. MEFs were similarly treated with vehicle or 4OHT as those carrying the *MGS34F* allele in **Fig. S1**. The recombined *S34F* allele and the WT allele are similar in sizes using the 5' probe (Ase I + BspH I cuts), while the *IES34F* and *S34F* alleles are similar in sizes using the internal probe (BamH I cut). (**D**) The percentages of *U2af1* RNA derived from the *U2af1*(S34F) allele were determined using whole cell RNA of the MEFs by an allele-sensitive S34F/WT SNP Taqman assay (Fig. S3C - E). (**E**) U2af1(S34F) alters the consensus sequences at the 3' splice site in MEFs with *IES34F;CreERT2*. MEFs were treated with two rounds of 4OHT (500 ng/ml) or vehicle before harvesting their RNA for mRNA sequencing. Sequence logos were compiled using DNA sequences surrounding the preceding 3' splice site of significantly altered cassette exons in samples treated with 4OHT.

Supplemental Figure S3. Characterization of tool assays. (**A**, **B**) Fragment analysis through high resolution capillary electrophoresis is used to estimate the efficiency of Cre-mediated recombination of the *MGS34F* allele. Panel **A**, representative electrograms of samples of known genotypes using fluorescence-labelled genotyping

primers (arrows) as shown in Fig. **1A** and in the inserted cartoons. The purple lines depict heterogeneous sequences of different sizes downstream of the two LoxP sites, preceding the WT *U2af1* cDNA and mutant exon 2. Genomic PCR yields amplicons of different sizes for the *WT*, *MGS34F* and *S34F* alleles. PCR products were run through an ABI 3130xl with ROX400 size standards to achieve single base pair resolution. x axis: base pairs; y axis: fluorescence intensity of the PCR amplicons. RE: recombination efficiency. Panel **B**, quantitative fragment analysis. Genomic DNA samples from either *MGS34F/WT* or *S34F/WT* mice were mixed at different ratios for fragment analysis. RE was calculated by the formula: $FI^{S34F}/(FI^{MGS34F} + FI^{S34F})$. FI, fluorescence intensity. r, correlation coefficient. (**C** - **E**) An allele-sensitive S34F/WT Single Nucleotide Polymorphism (SNP) Taqman assay is used to estimate the fraction of *U2af1*(S34F) RNA or DNA. Panel **C**, a cartoon to depict the relative locations of the primers (arrow) and probe (bar) of the Taqman assay. The primers and probes are all within exon 2 of *U2af1*, allowing detection of *U2af1*(S34F) specific sequences in genomic DNA and mRNA. Panel **D**, S34F and WT probes are specific for their targets. Probe specificity was demonstrated using plasmid DNA carrying either the WT or S34F mutant exon 2 of *U2af1* as the template. Seven 10-fold serial dilutions of the plasmid DNA templates were used (starting concentration: 1×10^7 molecules per 10 µl reaction), and the assay was performed in triplicate. According to the qPCR amplification curves (deltaRn vs. cycle), the WT and S34F probes are specific to their targets. Rn: normalized fluorescence emission intensity. Panel **E**, efficiency of the PCR-based Taqman assay by the formula: Efficiency $=10^{(-1/slope)} - 1$. Ct, critical threshold cycle number. The sequences of the primers and probes used in these tool assays are available in **Table S2**.

Supplemental Figure S4. Additional data related to Fig. 2. Panels **A** - **G**, **K** and **L**, showed additional CBC and flow cytometry results for the same blood and bone marrow samples as in **Fig. 2**. Panel **H** described the weight of these mice over time. Panels **I** and **J**, the numbers of nucleated bone marrow cell and spleen weight were quantified 36 weeks after poly (IC) treatment. Bone marrow cells were isolated from all femurs,

tibias, humerus, iliac crest and scapula bones from each mouse and the mean values of nucleated bone marrow cells (per mouse) were shown.

Supplemental Figure S5. Hematopoietic cells from *U2af1(S34F)* mice are defective in mice receiving non-competitive transplants. (**A** - **L**) Recipient mice were lethally irradiated and received bone marrow transplants from mice with the indicated genotypes. Blood samples from the recipients before and after poly (IC) treatment (at four weeks after transplant) were evaluated as described in the legend to **Fig. 2**. (**M**) Weight of mice over time. (**N** - **R**) *U2af1*(S34F) reduces the percentage of hematopoietic stem and progenitor cells in mouse bone marrow. Bone marrow cells from mice 24 weeks after poly (IC) treatment were analyzed by flow cytometry to measure the percentages of Lin^- cells (**N**), $Lin^-Scal\text{-}1^-c\text{-}Kit^+$ (MP) cells (**O**) and $Lin^-Scal\text{-}1^+c\text{-}Kit^+$ (LSK) cells **(P**). (**Q**) LSK cells expressing *U2af1*(S34F) are more proliferative. Mice were treated with BrdU for two hours before euthanasia. BrdU positive LSK cells were then quantified by flow cytometry. In panels **E** - **L** and **N** - **Q**, only donor cells (gated on GFP^+) were quantified. (**R**) Reduced myeloid colony-forming ability by bone marrow cells expressing *U2af1*(S34F). Bone marrow cells were harvested and subject to a colony-forming cell (CFC) assay. Numbers of colonies (1^{st} plating, 30,000 nucleated cells/dish left panel; 2^{nd} plating 10,000 nucleated cells/dish, right panel) were determined seven and twelve days after plating. (**S**, **T**) The numbers of nucleated bone marrow cell and spleen weight were quantified. Asterisks indicate significant changes by t test. N.S. not significant. Error bars represent s.e.m. *MGS34F/WT* (n = 10 for panels **A** - **M**; n = 5 for panels **N** - **T**), *MGS34F/WT; Mx1-Cre* (n = 11 for panels **A** - **M**; n = 5 for panels **N** - **T**).

Supplemental Figure S6. *U2af1*(S34F) is associated with low-grade myeloid dysplasia in mouse bone marrow. (**A**, **B**) *U2af1*(S34F) expression is associated with dysplasia in the erythroid and granulocytic lineages. (**A**) Representative images of Wright-Giemsa-stained bone marrow cytospins. Bi-nucleate erythroid progenitors (black arrows) and hyposegmented neutrophils (red arrows) were occasionally observed in the bone marrows from mice with *U2af1*(S34F) expression. Scale bar = 25 μm. (**B**)

Numbers of mice with erythroid and granulocytic dysplasia. At least 500 cells were examined by light microscope to see dysplasia in cytospins from each mouse, at least 19 weeks after poly (IC) treatment. Age- and treatment-matched controls were from *WT/WT* (n = 3), *MGS34F/WT* (n = 4), and *Mx1-Cre* (n = 3) mice. (**C** - **E**) U2af1(S34F) is associated with clustering of megakaryocytes in the bone marrow. Megakaryocyte clusters (at least two adjacent megakaryocytes [arrow heads]) were stained with H&E (**C**) or immune-stained for a megakaryocyte marker, CD61 (**D**). Scale bar = 50 μm. (**E**) Megakaryocyte clustering in bone marrow from the sternum of mice 19 weeks after poly (IC) treatment. The values in the y axis represent the numbers of clusters per bone segment. At least four segments of a sternum were evaluated for each mouse. Asterisks indicate significant changes by t test. Error bars represent s.e.m. n = 5 for each genotype.

Supplemental Figure S7. Effects of *Runx1* deletion and *U2af1*(S34F) on hematopoiesis, gene expression, and splicing. (**A**) Production of blood cells *in vivo*. Flow cytometry and CBCs were performed on blood from mice with the four indicated genotypes 4 - 20 weeks after poly (IC) treatment. Statistically significant differences, as determined by multiple t tests, between the bracketed strains are marked by asterisks (FDR < 0.05). *Runx1*$^{F/F}$; *Mx1-Cre* (n = 6); *MGS34F/WT*; *Mx1-Cre* (n = 10); URC (n = 11); age- and treatment-matched mice were used as the controls (n = 9. See Result section for details.). (**B**, **C**) *U2af1*(S34F) expression was associated with myeloid dysplasia for the erythroid and granulocytic lineages in the presence of *Runx1* deletion. Panel **B** showed representative cytospin images of bone marrow cells expressing *U2af1*(S34F) and deletion of *Runx1*. Bi-nucleate erythroid progenitors (black arrows) and hyposegmented neutrophils (red arrows) were occasionally observed. Scale bar = 25 μm. Panel **C** is a summary of quantification for these myeloid dysplastic features. At least 500 bone marrow cells were quantified for each mouse. (**D**) Formation of myeloid cell colonies in cultured dish. Deletion of *Runx1* enhances the number of colonies in a myeloid CFC assay, with or without *U2af1*(S34F). 30,000 nucleated bone marrow cells, harvested from mice four weeks after poly (IC) treatment, were plated in a 35 mm dish and colonies were counted two weeks later. URC (n = 3); *Runx1*$^{F/F}$; *Mx1-Cre* (n = 3);

control genotypes include mice of *WT* (n = 1), *MGS34F/WT* (n = 1), and *MGS34F/WT; Runx1*$^{F/WT}$ (n = 1). (**E** - **G**) Deletion of *Runx1* affects mainly gene expression, while *U2af1*(S34F) causes mainly changes in mRNA splicing. MPs (Lin$^-$Scal-1$^-$c-Kit$^+$) were harvested from the mouse bone marrow four weeks after poly (IC) treatment, and RNA was sequenced (n = 4 for each genotype). Panel **E**, Non-hierarchical clustering based on RNA profiles; color codes indicate genotypes. Panels **F** and **G**, Venn diagrams depict the numbers of overlapping and non-overlapping genes with significant changes in gene expression (**F**) or mRNA splicing (**G**) as compared to samples with the control genotype (*Mx1-Cre*).

Supplemental Figure S8. Reduced life span and myeloid hyperplasia in ENU-treated, *RUNX1* deficient mice. (**A**) Survival curves for mice with the designated genotypes, showing that mice without *Runx1* have a reduced life span since the start of poly (IC) treatment. *Runx1*$^{F/F}$*; Mx1-Cre* (n = 14); *MGS34F/WT*; *Mx1-Cre* (n = 9); URC (n = 16); control animals (n = 10) lack both *U2af1*(S34F) and *Runx1* deletions. (**B**) Mice lacking *Runx1* appear to have myeloproliferative disease, regardless of *U2af1* status. Left panel, percentages of WBCs with the indicated markers in the blood of end-stage *Runx1*$^{F/F}$; *Mx1-Cre* (n = 5) or URC (i.e. *MGS34F/WT*; *Runx1*$^{F/F}$; *Mx1-Cre***.** n = 8) mice. Age- and treatment-matched control mice lacking *U2af1*(S34F) and *Runx1* deletions (n = 3). Right panel, weights of spleens from the same mice.

Supplemental Figure S9. Further characterization of AML in ENU-treated URC mice related to Fig. 4. (**A**) Evidence of myeloblast infiltration in the spleen and liver (H&E stained) of the URC mouse with 1^o AML. (**B** - **G**) Further characterization of late AML cases. The left side of panels **B** and **C** showed increased percentage of c-Kit$^+$ cells in peripheral blood, bone marrow or spleen of a recipient mouse in comparison to those of the donor mouse (Case #2 and #3 respectively) by flow cytometry, when they were moribund. In the right side of panels **B** and **C**, the histology of H&E-stained bone marrow, spleen, or liver sections was compared between the donor and one of the recipient mice, when they were moribund. Blast cells occupied the bone marrow, spleen, and liver in the moribund recipient mice. Scale bar = 50 μm. (**D**, **E**) The cKit$^+$

cells in the recipients of late AML cases are GFP^+. Representative flow cytometry plots showing high percentages of GFP^+ cells in the $cKit^+$ gate of spleen cells from the recipient mice. Spleen cells from wild-type mice were used as the negative control. (**F**, **G**) Survival curves for secondary transplant. The donor cells were the moribund recipient mice that developed late AML in either Case #2 (panel **F**, n = 5) or Case #3 (panel **G**, n = 4). Day count started since after the transplantation.

Supplemental Figure S10. Genetic alterations in AMLs from Cases #1 - 3. (**A**) Frequencies of single nucleotide substitutions in the $GFP^+Lin^-c\text{-}Kit^+$ cell population from donor and recipient mice. TV, transversion; TI, transition. (**B** - **C**) Confirmation of somatic mutations by targeted panel sequencing using mIMPACT (see Supplemental Methods and Materials). Panel **B**, overlap of variants identified by WES and mIMPACT in DNA from $GFP^+Lin^-c\text{-}Kit^+$ cells from a Case #1 recipient mouse (R2). Panel **C**, VAF of shared variants and those uniquely identified by mIMPACT. (**D** - **H**) *Runx1* and *U2af1* alterations in the AML samples. Panel **D**, complete or nearly complete Cre-mediated recombination in the *Runx1* locus. Exon 4 (red box), which is flanked by LoxP sites, was absent in $GFP^+Lin^-c\text{-}Kit^+$ cells in all cases, but present in the tail DNA. An adjacent exon (Exon 3, black box) is shown as a control. Panels **E** - **G** depict the VAF of *U2af1*(S34F) (from WES, top) and the recombination efficiency (RE) of the *U2af1*(S34F) allele (bottom), as determined by fragment analysis (see **Fig. S3**). *U2af1*(S34F) was present in all donor- and recipient-derived samples from Case #1 (**E**) but absent in $GFP^+Lin^-c\text{-}Kit^+$ recipient cells (late AML) in Cases #2 and #3 (**F**, **G**), although Cre-mediated recombination of the *U2af1*(S34F) allele was nearly complete. (**H**) Representative results from the analysis of DNA fragments shown in panel F.

Supplemental Figure S11. AMLs from Case #1 - 3 exhibited expected changes at the RNA level. (A) *U2af1*(S34F) mRNA in $GFP^+Lin^-c\text{-}Kit^+$ cells from transplant recipients. *U2af1*(S34F) mRNA was present in $GFP^+Lin^-c\text{-}Kit^+$ cells from a Case #1 recipient mouse but not from recipients of cells from Cases #2 and #3 by mRNA sequencing. (**B**) *U2af1*(S34F) alters the consensus sequences at the 3' splice site in AMLs from Case #1 but the changes are not apparent in AML cells from Cases #2 and

#3. Comparisons were made to the transcriptome of AML derived from a *CBFB-MYH11* mouse (3), a non-U2af1-mutant model. Sequence logos were compiled in the same way as done in Figs 1E, S1D and S2E. These results are consistent with the status of *U2af1* DNA (see Fig. S10F and S10G).

Supplemental Table S1: Crossover of the linked *U2af1*(S34F) targeted alleles and the GFP transgene. The number of mice in each genotype category is shown. S34F in the column headers represents either the *MGS34F* or *IES34F* allele.

Supplemental Table S2: Primer and probe sequences.

Supplemental Table S3: Splicing alterations in mouse MP cells. Each row of the table corresponds to an isoform of a splicing event that is differentially spliced in MP cells from mice of different genotypes (n = 4 of each genotype) for the indicated comparisons. In each case, the isoform for which information is given is as follows: the inclusion isoform for cassette exons (type “se”), the intron-proximal isoform for competing 5' or 3' splice site events (types “a5ss” or “a3ss”), inclusion of the upstream exon for mutually exclusive exon events (type “mxe”), splicing of retained introns annotated as alternative (type “ri”) or constitutive (type “ci”), or canonical splicing of an annotated constitutive junction (type “cj”). The columns specify relevant annotations for each event as follows: “coords” specifies the genomic coordinates of the event; “gene” specifies the Ensembl gene ID; “geneName” specifies the gene name; “geneDescription” is a description of the gene as given by Ensembl; “type” specifies the type of splicing event; “nmdTarget” specifies whether the isoform is a predicted substrate for degradation by nonsense mediated decay (NMD) (NA indicates that either all or no isoforms of the event are predicted NMD substrates); "deltaPsi” give the absolute difference in isoform ratio (Percent Spliced In or “PSI” value) for the indicated comparisons; "logFoldChange", Log2 fold change values of Psi for the indicated comparisons; “p-value”, p-value computed with the Mann-Whitney U test for group comparisons.

Supplemental Table S4: Gene expression changes in mouse MP cells. Same as Supplemental Table S3 except that the levels of gene expression were compared.

Supplemental Table S5: Acquired somatic variants by whole exome sequencing (WES). Each row of the table corresponds to a variant of a gene, of which the allelic frequency (VAF) is significantly higher in the Lin-c-Kit+ cells originating from the donor (column A), as compared to the matched lymphocytes (column B). The columns specify relevant annotations for each variant as a standard output from Varscan2 (14).

Supplemental Table S6: Comparison of the frequencies of base substitutions in ENU-treated biological samples.

Supplemental Table S7: Acquired somatic variants by targeted gene panel sequencing (mIMPACT). Each row of the table corresponds to a variant of a gene, of which the allelic frequency (VAF) is seen in the Lin-c-Kit+ cells originating from the donor (column A), but not in the matched lymphocytes (column B). The columns specify relevant annotations for each variant as a standard output from Variant Effect Predictor (Ensembl), as described previously (7).

REFERENCES

1. Growney JD, et al. (2005) Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood* 106(2):494–504.
2. Golde WT, Gollobin P, Rodriguez LL (2005) A rapid, simple, and humane method for submandibular bleeding of mice using a lancet. *Lab Anim (NY)* 34(9):39–43.
3. Castilla LH, et al. (1999) The fusion gene Cbfb-MYH11 blocks myeloid differentiation and predisposes mice to acute myelomonocytic leukaemia. *Nature Genetics* 23(2):144–146.
4. Fei DL, et al. (2016) Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival. *PLoS Genet* 12(10):e1006384.
5. Carrington B, Varshney GK, Burgess SM, Sood R (2015) CRISPR-STAT: an easy

and reliable PCR-based method to evaluate target-specific sgRNA activity. *Nucleic Acids Res* 43(22):e157.

6. Dutra A, et al. (2010) Nuclear bud formation: a novel manifestation of Zidovudine genotoxicity. *Cytogenet Genome Res* 128(1-3):105–110.

7. Cheng DT, et al. (2015) Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J Mol Diagn* 17(3):251–264.

8. Ilagan JO, et al. (2015) U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res* 25(1):14–26.

9. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

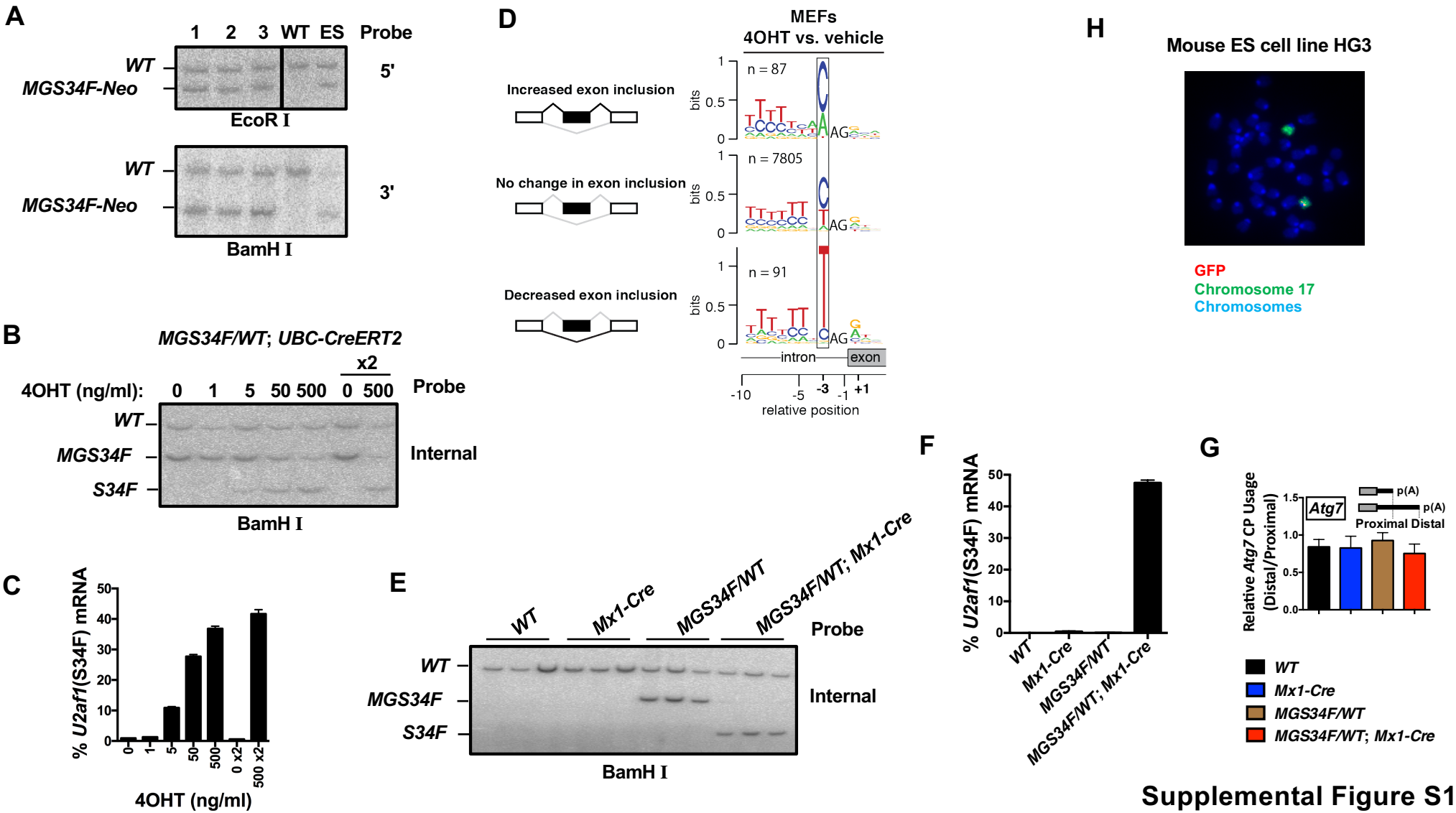
10. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.

11. Wagenmakers E-J, Lodewyckx T, Kuriyal H, Grasman R (2010) Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. *Cogn Psychol* 60(3):158–189.

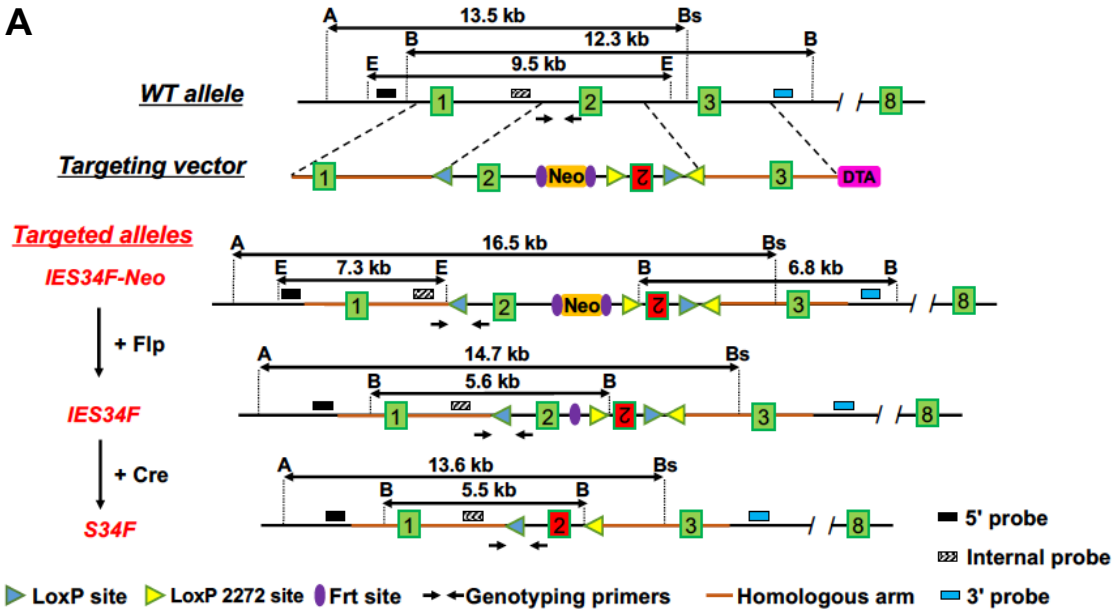
12. Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7(12):1009–1015.

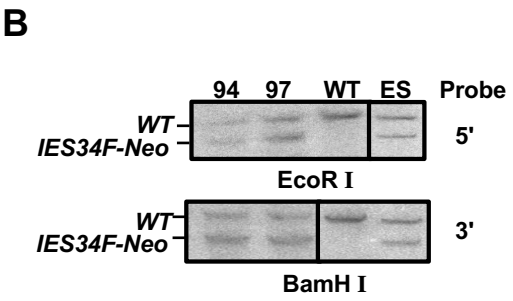
13. Gentleman RC, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80.

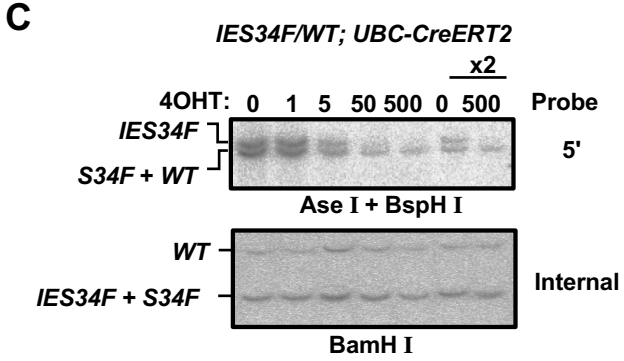
14. Koboldt DC, et al. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576.

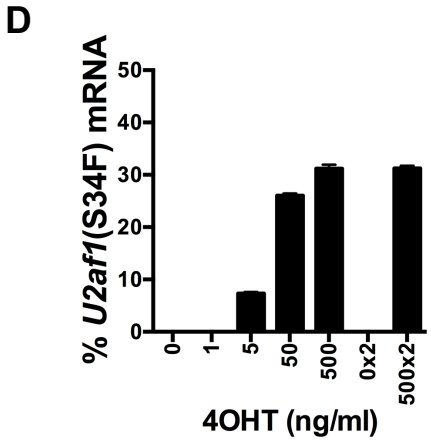


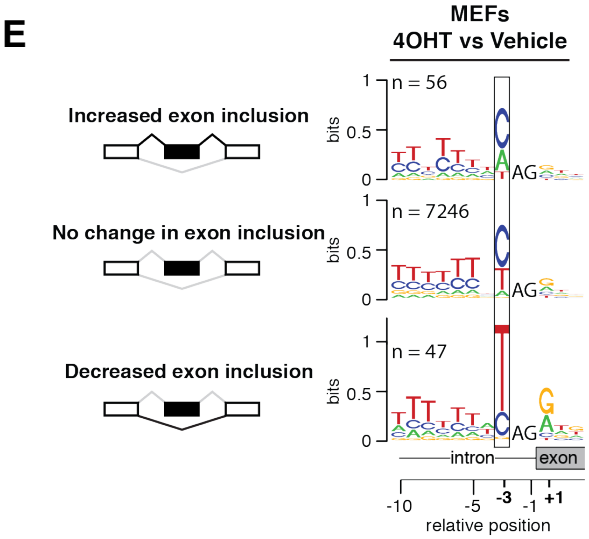
Supplemental Figure S1



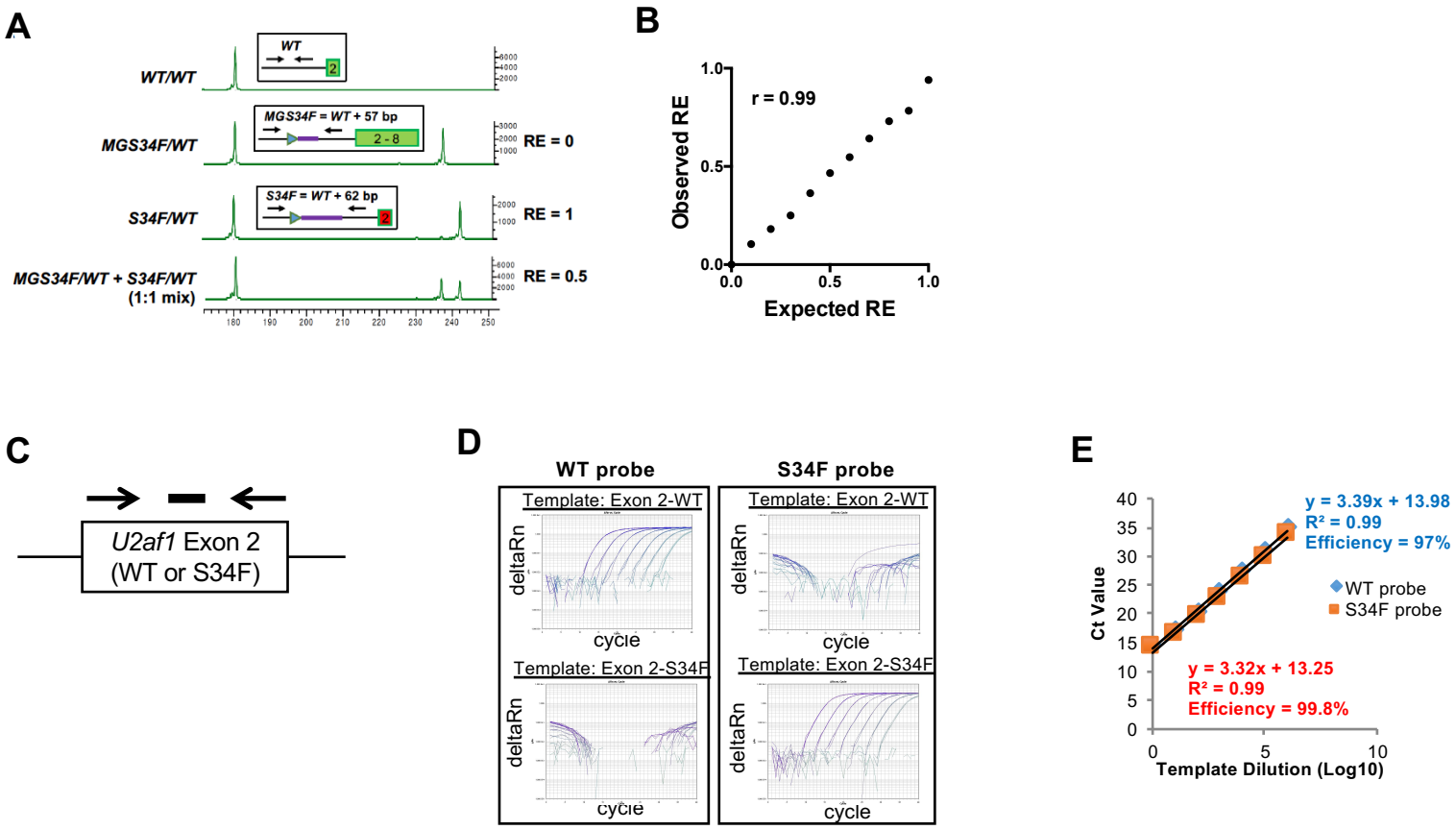




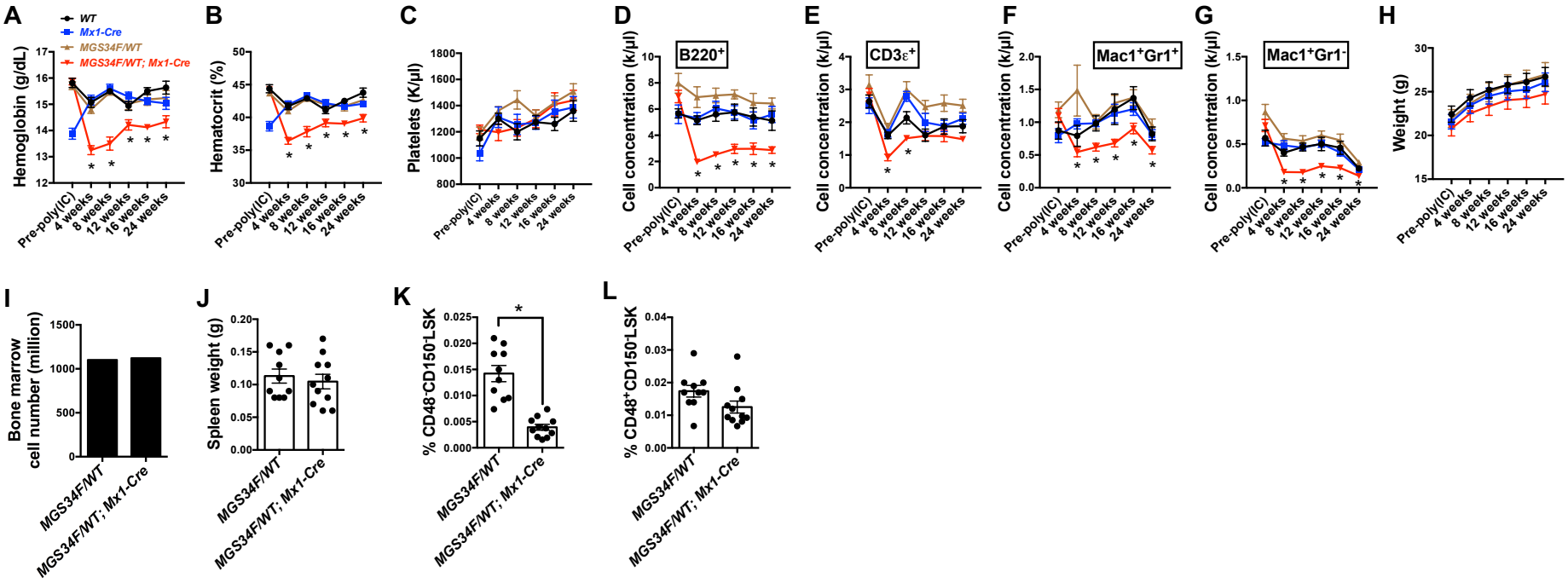




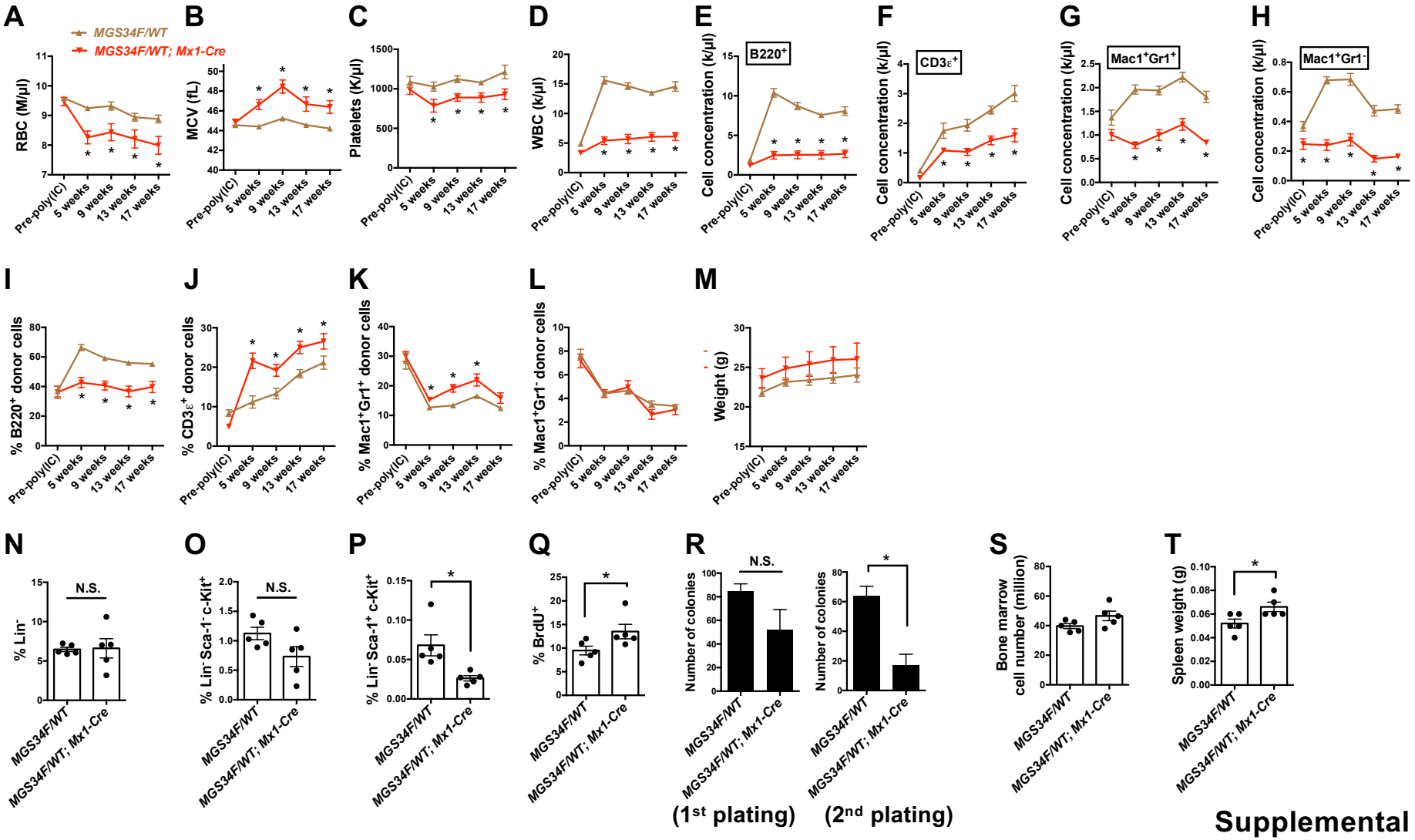
Supplemental Figure S2



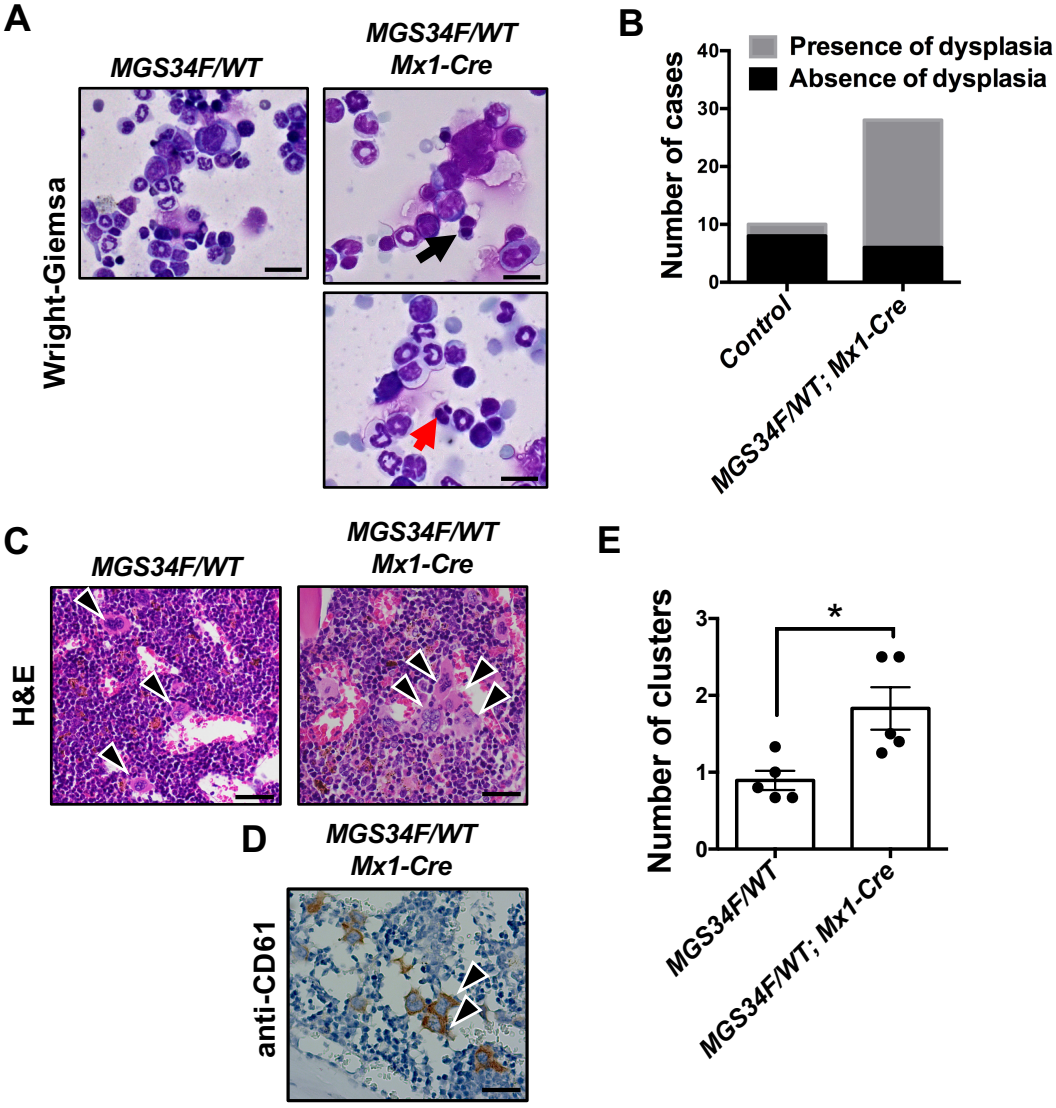
Supplemental Figure S3



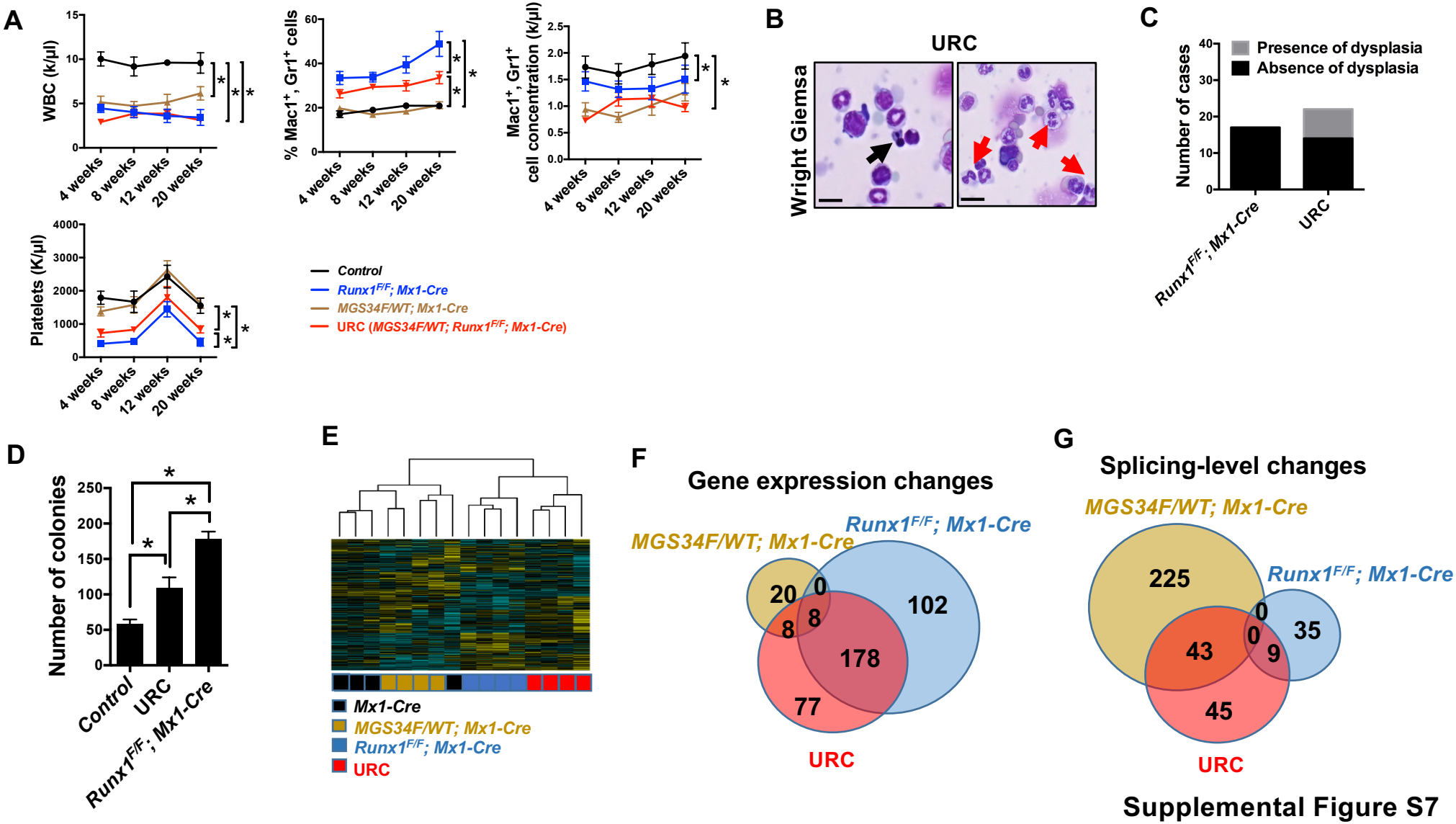
Supplemental Figure S4



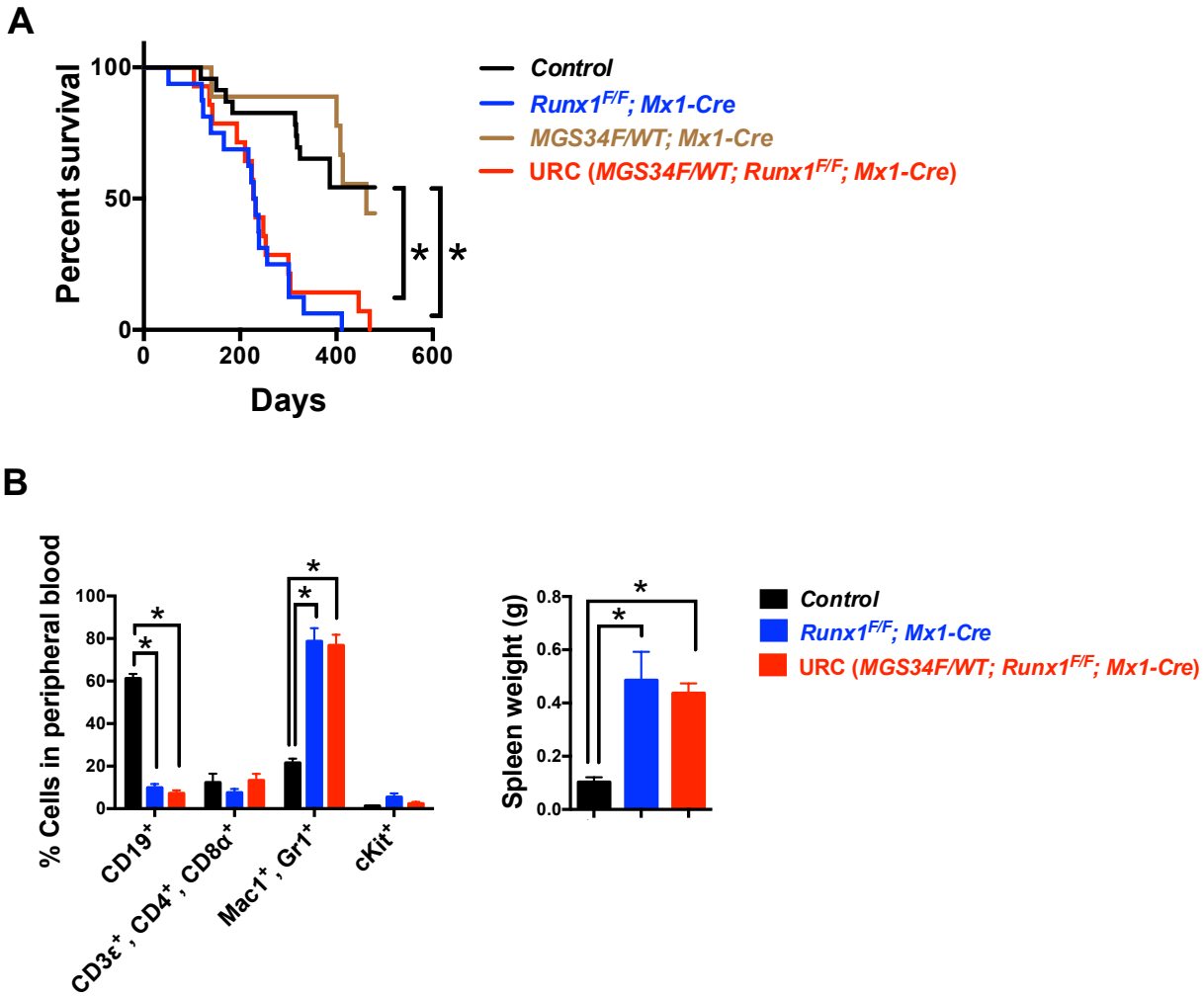
Supplemental Figure S5



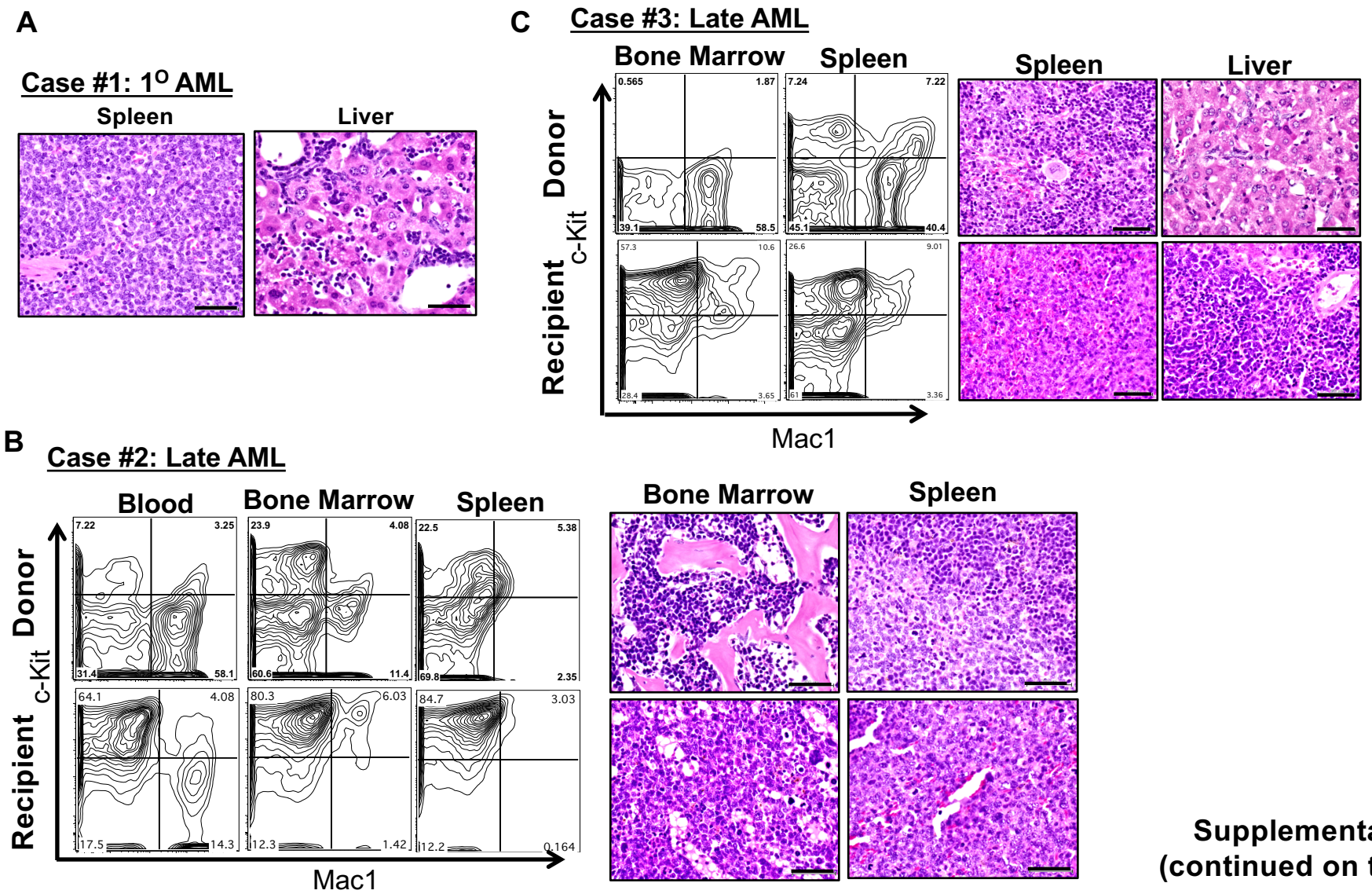
Supplemental Figure S6



Supplemental Figure S7

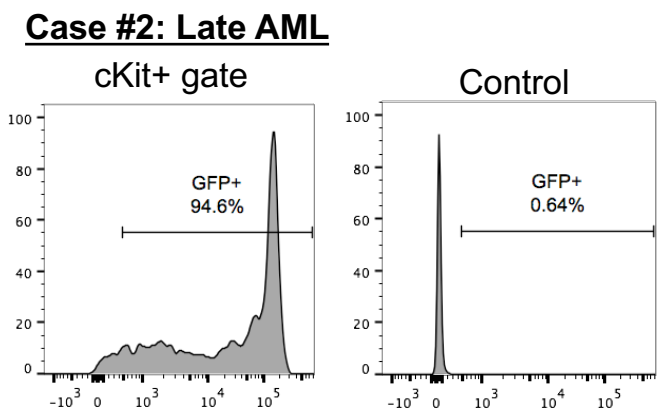


Supplemental Figure S8

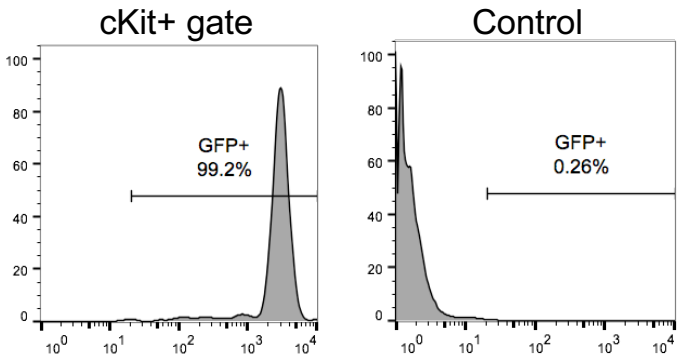


Supplemental Figure S9
(continued on the next page)

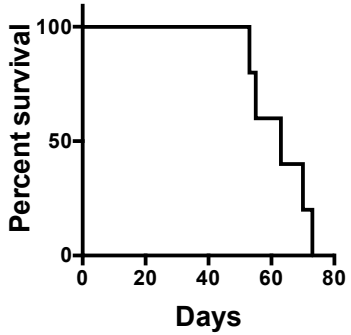
D



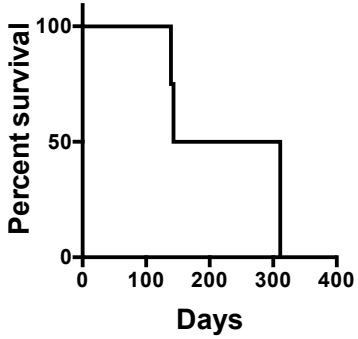
E



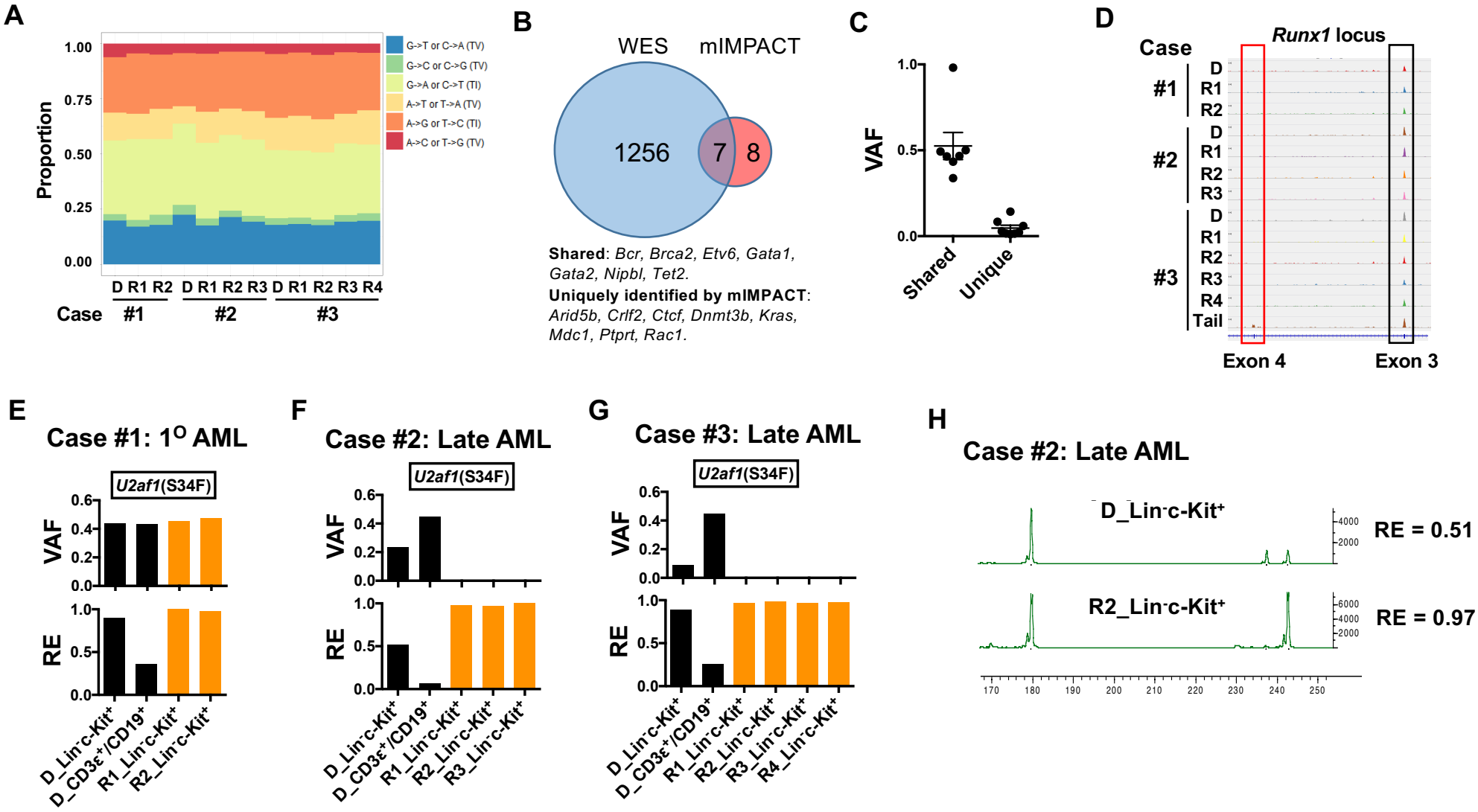
F



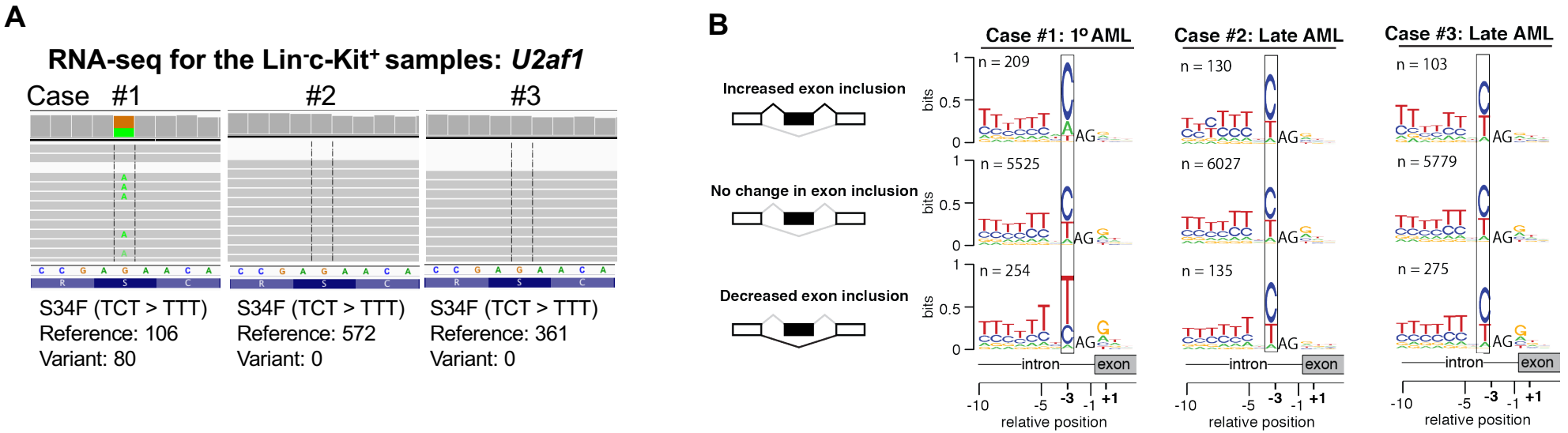
G



Supplemental Figure S9



Supplemental Figure S10



Supplemental Figure S11