Impaired hematopoiesis and leukemia development in mice with a "knock-in" allele of *U2af1*(S34F)

Dennis Liang Fei[1,2,*,#], Tao Zhen[3,*], Benjamin Durham[4], John Ferrarone[1], Tuo Zhang[5], Lisa Garrett[6], Akihide Yoshimi[4], Omar Abdel-Wahab[4, 7], Robert K. Bradley[8, 9], Paul Liu[3,#], Harold Varmus[1,2,10,#]

1. Department of Medicine, Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, United States
2. Cancer Biology Section, Cancer Genetics Branch, National Human Genome Research Institute, Bethesda, MD 20892, United States
3. Oncogenesis and Development Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, United States
4. Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States
5. Genomics Resources Core Facility, Department of Microbiology and Immunology, Weill Cornell Medicine, New York, NY 10065, United States
6. Embryonic Stem Cell and Transgenic Mouse Core, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, United States
7. Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States
8. Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, United States
9. Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States
10. New York Genome Center, New York, NY 10013, United States

* Equal contribution

Correspondence: DLF, dennisfei@hotmail.com; PL, pliu@mail.nih.gov; HV, varmus@med.cornell.edu

Keywords: U2AF1, splicing factor, S34F, Runx1, myelodysplastic syndromes, leukemia

ABSTRACT

We have generated mice that carry Cre-dependent "knock-in" alleles of *U2af1*(S34F), the murine version of a mutant allele commonly encountered in human myelodysplastic syndromes and acute myeloid leukemia (AML). Cre-mediated recombination in murine hematopoietic lineages caused changes in RNA splicing, as well as multilineage cytopenia, macrocytic anemia, decreased hematopoietic stem and progenitor cells, low-grade dysplasias, and impaired transplantability, but without lifespan shortening or leukemia development. Mice with Cre-dependent *U2af1*(S34F) and homozygous *Runx1* "knockout" alleles were mutagenized with low-dose N-Ethyl-N-Nitrosourea, and three of fourteen mice developed AML. However, AML did not arise in mice with other genotypes in our relatively small cohort. Sequencing DNA from the three AMLs revealed somatic mutations considered to be drivers of human AML, including mutations in *Tet2*, *Gata2*, *Idh1,* and *Ikzf1*. However, the *U2af1* missense mutation reverted to WT in two of the three AML cases, implying that *U2af1*(S34F) is dispensable for maintaining leukemia.

SUMMARY

A missense mutation affecting the spliceosomal protein U2AF1, *U2AF1*(S34F), is commonly found in human myeloid neoplasms. Fei *et al* describe hematopoietic defects and development of occasional leukemia in mice carrying a conditional "knock-in" allele of *U2af1*(S34F).

INTRODUCTION

Myelodysplastic syndromes (MDS) are neoplastic diseases characterized principally by deficiencies of normal cells in myeloid and erythroid lineages, myeloid dysplasia, clonal dominance of abnormal immature cells, and variable risks of developing secondary AML (Catenacci and Schiller, 2005; Troy et al., 2014; Arber et al., 2016). In over half of MDS patients, the malignant clone carries a mutation in one of four genes (*U2AF1*, *SRSF2*, *SF3B1*, and *ZRSR2*) encoding factors critical for correct splicing of pre-messenger RNA (pre-mRNA) (Yoshida et al., 2011). Mutations in these "splicing factor genes" are presumed to be causative events in MDS because of their frequency, their recurrence at a few positions in coding sequences, their high allelic ratios, and their association with clinical outcomes (Yoshida et al., 2011; Papaemmanuil et al., 2011; Wang et al., 2011; Graubert et al., 2012; Imielinski et al., 2012; Furney et al., 2013; Waterfall et al., 2014; Damm et al., 2012). Still, despite intensive studies, the mechanisms by which such mutations contribute to the initiation or maintenance of neoplasias have not been identified.

We have been studying the most common mutation observed in *U2AF1,* the gene encoding an RNA-binding protein that helps to direct the U2 small nuclear ribonucleoprotein particle (U2 snRNP) to the 3' splice acceptor site in pre-mRNA (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). This mutation replaces serine at position 34 of U2AF1 with phenylalanine (S34F) (Yoshida et al., 2011; Graubert et al., 2012), yielding a neomorphic splicing factor that changes splicing patterns for many RNAs and does not preserve enough normal splicing to permit cell survival in the absence of a wild type allele (Fei et al., 2016). In addition, *U2AF1* (S34F) is present at high allelic frequencies (near 50%) in MDS (Graubert et al., 2012), was reported to predispose patients to secondary AML (Graubert et al., 2012), and is also found, albeit at lower frequencies, in a variety of other neoplastic diseases, including solid tumors (Imielinski et al., 2012; Cancer Genome Atlas Research Network, 2014; Waterfall et al., 2014).

Progress towards understanding the oncogenic effects of U2AF1(S34F) has been impeded by a lack of appropriate biological models. Here we describe our development of genetically engineered mouse models in which *U2af1*(S34F) is assembled at the endogenous *U2af1* locus by the action of the Cre recombinase. By activating Cre in the hematopoietic compartment of these mice to produce U2af1(S34F), the mice develop impairments of blood cells with MDS-like features, accompanied by abnormal splicing patterns resembling those previously observed in human cells expressing this mutant splicing factor. In an effort to model *U2AF1*(S34F)-associated leukemia in these mice, we deprived U2af1 mutant mice of a hematopoietic transcription factor, Runx1, often co-mutated in human MDS and leukemias (Xu et al., 2014; Adema et al., 2016), and treated them with a chemical mutagen, Ethyl-Nitroso-Urea (ENU). Under those circumstances, clones of AML developed in three of fourteen mice, but not in mice lacking any of the three factors. Although AML did not occur often enough to conclude that the mutant splicing factor was required for leukemogenesis in this model, whole exome sequencing of AML cells revealed somatic mutations in genes often mutated in human AML. In summary, we report the first "knock-in" mouse lines with conditional U2af1(S34F) alleles, demonstrate the effects of mutant U2af1 on mouse hematopoiesis, and identify additional mutations that may cooperate with *U2af1*(S34F) and *Runx1* deficiency during leukemogenesis.

RESULTS

Establishing mice carrying conditional knock-in S34F alleles of *U2af1*

We used targeting vectors (termed *MGS34F* and *IES34F*; Figs 1A and S1A) to create two conditional S34F mutant alleles at the endogenous *U2af1* locus of B6/129 mice, and we documented the successful introduction of the targeting vectors at the appropriate sites by restriction mapping and Sanger sequencing (Fig. 1B, S1B, and Methods). Mice carrying either of the altered *U2af1* alleles express green fluorescent protein (GFP) in all tissues because the mouse embryonic stem cell line used to generate the mutant mice carries a *GFP* transgene driven by the ubiquitously active human UBC promoter (Schaefer et al., 2001) and the transgene is located on the same chromosome as the U*2af1* locus, with a crossover frequency of 1.95% (Fig. S1F; Table S1).

To assess whether these conditional alleles can be rearranged correctly, we crossed the two mouse strains with mice carrying a UBC-driven transgene encoding a tamoxifen-dependent Cre recombinase (UBC-CreERT2) (Ruzankina et al., 2007). Mouse embryo fibroblasts (MEFs) derived from embryos with *MGS34F; CreERT2* or *IES34F; CreERT2* were treated with 4-hydroxyl-tamoxifen (4OHT) in culture; efficient appearance of the expected configurations of *U2af1*(S34F) was confirmed by Southern blotting (Fig. 1C and S1C), and the anticipated changes in *U2af1*(S34F) mRNA were ascertained with allele-specific Taqman assays (Figs 2A, S1D, and S1I - K). *MGS34F* appeared to be more efficiently recombined by Cre than was the *IES34S* allele, and was used in all follow-up studies.

We next crossed mice heterozygous for the conditional *MGS34F* allele with mice carrying an *Mx1-Cre* transgene that is expressed in the blood lineage upon administration of poly (IC) (Kühn et al., 1995). Two weeks after poly (IC) treatment of the resulting bi-transgenic mice (*MGS34F/WT; Mx1-Cre*), Cre-mediated recombination was nearly complete in bone marrow cells, as revealed by Southern blotting (Fig.1D)

and by quantitative analysis of PCR-generated fragments (Fig. S1G and S1H). The wild type and mutant alleles appear to be expressed at equivalent levels, judging from the approximately 1:1 ratio of mutant and wild-type *U2af1* mRNA (Fig. 2B).

Effects of U2af1(S34F) on pre-mRNA splicing in mouse cells resemble effects of U2AF1(S34F) in human cells

U2AF1(S34F) is known to affect pre-mRNA splicing and possibly other RNA processing events in a variety of mostly neoplastic human cells (Fei et al., 2016; Ilagan et al., 2015; Okeyo-Owuor et al., 2015; Brooks et al., 2014; Przychodzen et al., 2013; Park et al., 2016; Chen et al., 2018; Shirai et al., 2015). We used cells and tissues from mice carrying the *U2af1(S34F)* allele to determine whether similar alterations occur in mice after expression of *U2af1*(S34F) at physiological levels. As an initial assessment, we measured changes in spliced products of pre-mRNA from six mouse genes (*H2afy*, *Gnas*, *Bcor*, *Picalm*, *Kdm6a*, and *Rac1*) homologous to human genes whose transcripts undergo changes in alternative splicing (Fei et al., 2016; Ilagan et al., 2015; Shirai et al., 2015) in the presence of U2AF1(S34F). We also examined RNA products of one mouse gene, *Atg7*, whose transcripts were reported to show changes in use of polyadenylation sites (Park et al., 2016). Isoform sensitive primers were designed to target regions of the mRNAs that are predicted to undergo changes based on previous results in human or mouse cells (Table S2). We found that similar alterations in pre-mRNA splicing occurred in the six homologous murine genes expressed in total bone marrow cells (Fig. 2C). However, the previously reported changes in *Atg7* RNA polyadenylation sites were not observed in mouse bone marrow cells expressing *U2af1*(S34F) (Fig. 2C).

We used whole cell mRNA sequencing methods to make a more extensive survey of U2af1(S34F)-induced alterations in pre-mRNA splicing in mouse bone marrow cells, looking specifically for the changes in frequency with which so-called “cassette exons” are sometimes included in the spliced products, since these are the most prominent consequences of the S34F mutant in human cells (Fei et al., 2016; Ilagan et al., 2015; Okeyo-Owuor et al., 2015; Brooks et al., 2014; Przychodzen et al., 2013; Shirai et al.,

2015). In addition, we performed these analyses in both MEFs and mouse myeloid progenitors (MPs). The MPs were isolated from the bone marrow of *MG(S34F)/WT; Mx1-Cre* mice (and from *Mx1-Cre* mice as a control) four weeks after poly (IC) treatment and were identified by their immunophenotype, $Lin^{-}Sca\text{-}1^{-}c\text{-}Kit^{+}$.

We found significant changes in the alternative splicing of cassette exons from 206 mouse genes; 43 of those genes (about 20 percent) are homologs of human genes similarly affected by *U2AF1*(S34F) or *U2AF1*(S34Y) in AML cells (Fig. 2E). Furthermore, the consensus nucleotide sequences preceding the invariant AG dinucleotide at the 5' end of the affected cassette exons match those previously identified as determinants of altered splicing patterns in human cells expressing U2AF1(S34F) (Fei et al., 2016; Ilagan et al., 2015; Okeyo-Owuor et al., 2015; Brooks et al., 2014; Przychodzen et al., 2013; Shirai et al., 2015). For example, the first nucleotide position 5' of the AG dinucleotide is frequently occupied with a C or A nucleotide upstream of exons that are more frequently included in mRNA, and is often occupied with a T nucleotide upstream of those cassette exons that are less frequently included (Figs. 2D and S1E). Based on these results, we conclude that physiological expression of *U2af1*(S34F) causes changes in pre-mRNA splicing similar but not identical to those observed by us and others in human cells expressing *U2AF1*(S34F).

U2af1(S34F) affects hematopoiesis in mice, with features that are shared with human myelodysplastic syndromes

We next examined the impact of *U2af1*(S34F) on hematopoiesis in our genetically engineered mice. Blood samples were taken before and after administration of poly (IC) to induce Cre and subjected to complete blood cell counts (CBCs) and flow cytometry. The cellular profiles of blood from wild type mice and from mice carrying only the *Mx1-Cre* transgene or only the *MGS34F* allele were similar and normal, as expected. However, *MGS34F/WT; Mx1-Cre* mice showed mild but persistent changes in red blood cells (reduced RBCs, hemoglobin concentration, and hematocrit; and increased mean red cell volume; Figs 3A, 3B, S2A, and S2B). Platelet numbers were normal (Fig. S2C),

but the numbers of white blood cells were reduced persistently by nearly 50% compared to results with mice with the control genotypes after poly (IC) treatment (Fig. 3C). Decreases were observed in all major WBC lineages as measured by flow cytometry (Fig. 3D - 3G), with the most marked reduction seen in $B220^+$ B cells (Fig. S2D - S2G). Despite the reduction in nearly all mature blood cell types upon activation of *U2af1*(S34F), bone marrow cellularity and spleen size were normal, with no evidence of pathological extramedullary hematopoiesis in mice with *U2af1*(S34F) (Figs S2Q and S2R).

We sought to determine whether these long-term changes were due to effects of the mutant splicing factor on blood stem and progenitor cells. We examined the abundance and function of hematopoietic stem cells (HSCs) and progenitors in mutant and control mice 36 weeks after poly (IC) treatment. Bone marrow cells from mice with or without *U2af1*(S34F) were stained with antibodies against markers for HSCs and for various blood cell progenitors and analyzed the cells by flow cytometry. While no differences were observed in the percentages of lineage-negative (Lin^-) cells (Fig. 3I) and MPs (Lin^- $Sca\text{-}1^-c\text{-}Kit^+$, Fig. 3H and 3J) from mutant and control animals, the percentages of short-term HSCs (marked by $Lin^-Sca\text{-}1^+c\text{-}Kit^+$ [LSK]) and, especially, long-term HSCs (marked by $CD48^-CD150^+$ within the LSK population) were decreased in the bone marrow of *U2af1*(S34F) mice (Figs 3H, 3K and 3L). These findings suggest that reductions in HSCs likely account for the persistence of multi-lineage cytopenia observed in the peripheral blood of mice expressing mutant *U2af1*.

To determine whether the effects of *U2af1*(S34F) on mouse hematopoiesis were cell autonomous, we transplanted bone marrow cells from *MGS34F/WT; Mx1-Cre* mice and from control (*MGS34F/WT*) mice non-competitively into lethally irradiated recipient mice (F1 progeny of a B6 X 129 cross; see Methods). Four weeks after transplantation, poly (IC) was used to induce production of Cre in the transplanted cells. Five weeks later, macrocytic anemia, multi-lineage cytopenia, and reduction in HSCs were observed in the recipient mice (compare Fig. 4A - 4J with Fig. 3A - 3K; Fig. S2I - S2O with Fig. S2A - S2G; and Fig. S2S and S2T with Fig. S2Q and S2R). Although the numbers of LSK-

positive cells were reduced, a higher fraction of these cells were actively cycling, as shown by a BrdU incorporation assay (Fig. 4K), suggesting a compensatory mechanism to overcome the loss of these cells. Moreover, bone marrow cells from the *U2af1*(S34F) mice formed fewer myeloid colonies, especially after re-plating (Fig. 4L), consistent with the reduced numbers of HSCs observed by flow cytometry. Overall, these results demonstrate that the effects of U2af1(S34F) on hematopoiesis are cell-autonomous.

To further characterize the HSC defects attributed to *U2af1*(S34F), we performed competitive transplantations using bone marrow cells from *MGS34F/WT; Mx1-Cre* or *MGS34F/WT* mice (test cells) mixed with bone marrow cells from WT or *Mx1-Cre* mice (competitor cells). Equal numbers of test and competitor cells were mixed and transplanted into lethally irradiated recipient mice (Fig. 5A), and the recipients received poly (IC) four weeks later. Mice that received test cells from *MGS34F/WT; Mx1-Cre* animals displayed a deficiency of *U2af1*(S34)-expressing cells in all measured mature and immature cell lineages in the peripheral blood, bone marrow, and spleen of the recipient mice, as early as two weeks after poly (IC) (Fig. 5B - 5F). In contrast, immature and mature blood cells from *MGS34F/WT* mice remained at levels of nearly 50% at all times (Fig. 5B and 5C). These results show that HSCs expressing *U2af1*(S34F) are defective in re-populating the hematopoietic compartment in a competitive transplant setting.

Using light microscopy to seek the morphological features of myelodysplasia, we occasionally observed bi-nucleated erythroid cells and hypo-segmented neutrophils in the bone marrow cells from *U2af1*(S34F) mice (Fig. 6A and 6B). Moreover, megakaryocytes from these mice occasionally formed clusters (Fig. 6C - 6E), even though platelet production was normal or only slightly diminished (Fig. S2C and S2K). These dysplastic features typically affected less than 1% of bone marrow cells, not present in the peripheral blood, and, therefore, did not meet the diagnostic criteria for human MDS. Nevertheless, the myeloid dysplasia, multi-lineage cytopenia, and macrocytic anemia found in the *U2af1*(S34F) mutant mice are all hallmarks of human MDS.

Generation of mice with conditional alleles of both *Runx1* and *U2af1*

Although *U2af1*(S34F) affects hematopoiesis in our genetically engineered mice, these mice were relatively healthy (e.g. with normal weight gain; Figs S2H and S2P), and they lived a normal life span. We reasoned that additional genetic changes might produce more severe hematologic defects or neoplasias dependent on *U2af1(S34F)*. Therefore, we combined our *MGS34F* allele and the *Mx1-Cre* transgene with a *Runx1* allele carrying LoxP sites that flank exon 4, encoding the DNA-binding Runt domain (Growney et al., 2005). When bred to homozygosity at the floxed *Runx1* locus (*Runx1*$^{F/F}$) in the presence of *Mx1-Cre* and *MGS34F*, Cre recombinase inactivates *Runx1* and allows production of the mutant splicing factor in the same cells after administration of poly (IC). Since loss-of-function mutations of *Runx1* often co-occur with *U2AF1* mutations in human myeloid neoplasms (Xu et al., 2014; Adema et al., 2016), this seemed to be a promising genetic combination for eliciting pathological phenotypes, including leukemias.

The experimental cohort used for these purposes consisted of the "compound engineered" mice (*MGS34F/WT; Runx1*$^{F/F}$; *Mx1-Cre*, called “URC mice”); mice in which poly (IC) activates *U2af1(S34F)* or inactivates *Runx1* (*MGS34F/WT; Mx1-Cre* or *Runx1*$^{F/F}$*; Mx1-Cre*); and control WT mice (or WT-equivalent mice lacking the *Mx1-Cre* transgene). After induction of Cre, URC mice developed the hematological abnormalities associated with either U2af1(S34F) (e.g. pancytopenia) or *Runx1* deficiency alone (e.g. thrombocytopenia, increased percentage of myeloid cells, and myeloid colony formation; (Growney et al., 2005; Ichikawa et al., 2004)), as shown in Fig. S3A and S3B. We also examined the effects of *U2af1*(S34F) and *Runx1* deletion on gene expression and pre-mRNA splicing in MP cells from poly (IC)-treated URC mice and compared the findings with those obtained with mice with conditional alleles at either the *U2af1* or *Runx1* loci (Fig. S3C - S3E). As anticipated from its role as a gene encoding a transcription factor, *Runx1* deficiency affected mainly gene expression, whereas U2AF1(S34F) changed predominantly mRNA splicing. As a result, MP cells

with *U2af1*(S34F) and without functional *Runx1* exhibited many changes in both mRNA splicing and gene expression, as illustrated in Fig. S3E and S3D.

URC mice develop frank AML or AML clones after ENU mutagenesis

Although URC mice showed many changes in hematopoiesis and in gene expression in MP cells (Fig. S3), they had a normal life span and did not develop frank leukemia within one and a half years after poly (IC) treatment. To seek additional genetic alterations that might cooperate with *U2af1*(S34F) and *Runx1* deficiency to cause more profound pathology, including leukemias, we performed a forward genetic screen by treating young URC mice (6 – 16 weeks old) with the mutagen N-ethyl-N-nitrosourea (ENU) one week after inducing programmed genetic changes with poly (IC) (Fig. 7A). We used a relatively low-dose of ENU (100 mg/kg) that, based on our previous experience (Castilla et al., 1999), did not produce myeloid leukemia in otherwise normal animals. Mice that developed T cell malignancies and thymoma (known consequences of ENU) were excluded from further analysis.

All mice lacking *Runx1* died prematurely (within 15 months after poly (IC) treatment), regardless of *U2af1* status, after receiving ENU (Fig. 7B). However, ten months after administration of ENU, one of the 14 URC mice (Case #1) --- but none of the mice with *Runx1* deletion alone or any other genotypes --- developed AML (which we call primary or 1^{o} AML). Blast cells were abundant in the peripheral blood, bone marrow, and spleen, and they infiltrated multiple organs, including the liver (Fig. 7E). The malignant cells expressed the c-Kit receptor but not mature lineage markers, including Mac1 (Fig. 7D). When the bone marrow or spleen cells were transplanted from Case #1 to sub-lethally irradiated mice, the recipients died with AML within three months (Fig. 7F).

All of the other ENU-treated *Runx1*$^{F/F}$; *Mx1-Cre* mice, with or without the *U2AF1*(S34F) mutation died prematurely with myeloid pathology: increased percentages of myeloid cells in the peripheral blood, enlarged spleen (Fig. 7C), and extramedullary hematopoiesis. Low-grade dysplasia in the erythroid and neutrophil lineages was

occasionally observed in the peripheral blood and bone marrow of moribund URC mice (Fig. S4A and S4B), consistent with the myeloid dysplasia associated with *U2af1*(S34F). These results suggest that ENU-treated mice with deletions of *Runx1* develop lethal forms of myeloid proliferative disease (MPD), with occasional myeloid dysplastic features in the presence of *U2af1*(S34F).

We speculated that the animals with lethal forms of MPD but no frank AML might harbor malignant clones capable of producing myeloid leukemias if introduced into healthy recipient mice. To test this idea, spleen and/or bone marrow cells from moribund URC mice with MPD were transplanted to sub-lethally irradiated mice, and the recipients were monitored for up to one year after transplant for development of AML (Fig. 8A). None of the animals receiving transplanted cells from *Runx1*$^{F/F}$; *Mx1-Cre* mice (n = 11) developed AML or other hematopoietic malignancies. However, all recipients of spleen and/or bone marrow cells from two of the eleven ENU-treated URC donors (Cases #2 and #3 in Figs 8 and S4) died with AML 12 or 42 weeks after transplant (Fig. 8B and 8C). Several hematopoietic compartments, including blood, bone marrow and spleen, as well as other tissues such as the liver, were filled with immature c-Kit^{+} hematopoietic precursors in the recipient mice, but not in the donor mice (Figs 8D and 8E, and S4C and S4D). The c-Kit^{+} cells were GFP^{+} (data not shown), confirming that they were derived from the donor mice. Results from secondary transplants, using spleen cells from recipient mice, confirmed that the malignancies were transplantable (Fig. S4E and S4F). We refer to the AMLs derived from Cases #2 and 3 as late AMLs to distinguish them from the 1^{o} AML observed in one ENU-treated URC mouse (Case #1).

In summary, we have observed a total of three cases of AML arising from the bone marrow of URC mice after ENU treatment (Table 1). In contrast, AML did not appear in any of the ENU-treated mice with one gene alteration (i.e. only *U2af1*(S34F) or only deletion of exon 4 of *Runx1*); in recipients of bone marrow transplants from such animals; or in mice that did not receive ENU. These findings are consistent with the idea that initiation of AML in our cohort requires *U2af1*(S34F), *Runx1* deficiency, and treatment with ENU. However, due to the low incidence of AML (derived from three out

of fourteen URC mice), a larger cohort would be necessary to establish the requirement of all of these three factors, including mutant U2af1, for leukemogenesis in a statistically significant fashion (see footnote to Table 1).

Identification of probable leukemogenic, ENU-induced mutations by whole exome sequencing of DNA from Cases #1 – 3

We performed whole exome sequencing (WES) on DNA from $GFP^+Lin^-c\text{-}Kit^+$ (LK) cells from the spleens of Cases #1, 2, and 3 mice to seek additional genetic changes caused by ENU. DNA from $GFP^+CD3\varepsilon^+$ or GFP^+CD19^+ lymphocytes from each of the three affected URC mice was used to determine a reference sequence. In each leukemic sample, we identified more than 1000 somatic variants by WES (Table S5). The most common variants were single nucleotide substitutions, among which G to A (or C to T) transitions, A to G (or T to C) transitions, and G to T (or C to A) transversions were the most frequent (Fig. S5A). These changes are consistent with previously reported effects of ENU (Table S6).

Case #1: By reviewing the variants for likely participants in leukemogenesis in primary AML cells from Case #1, we identified a splice donor site mutation in *Tet2* that is predicted to disrupt correct splicing, and a missense mutation in *Gata2* homologous to the *GATA2*(R362Q) mutation found in human AML (Fig. 9A - 9F).

Case #2: In the leukemic cells from Case #2, we found a missense point mutation in *Idh1* homologous to *IDH1*(R132Q) (Fig. 9G – 9J).

Case #3: In the leukemia cells from case #3, we found a missense mutation affecting a cysteine in a C2H2 Zinc finger domain of the transcription factor Ikzf1 (C175R, Fig. 9K – 9M).

In general, the VAFs of these potentially pathogenic variants were low (0.2 or less) in the donor LK cells, absent in the donor's normal lymphocytes, and near or even above 0.5 in the LK cells recovered from the recipient mice that had developed AML (Figs 9A, 9D, 9G, and 9K).

We next validated the presence of these variants and others by deep sequencing of several DNA samples from leukemic cells from Cases #1 - 3, using a focused panel of cancer-related genes (Table S7). This approach detected additional variants at low VAFs that are likely to be subclonal (Fig. S5B and S5C). We further confirmed the expression or functional consequences of these variants by mRNA sequencing. For example, the splice site mutation in *Tet2* resulted in exon skipping (Fig. 9C) and the *Gata2*(R362Q), *Idh1*(R132Q) and *Ikzf1*(C175R) mutations were observed in mRNA (Fig. 9F, 9I, and 9M). These identified variants are likely drivers of leukemogenesis, because *TET2* loss-of-function mutations, *GATA2*(R362Q), and *IDH1*(R132) mutations are often seen in human AML (see Discussion).

The *U2af1*(S34F) allele is absent in AML cells derived from Cases #2 and #3.

Encountering mutations in these mouse AMLs that are identical or similar to mutations frequently reported in human AML was surprising, but analysis of the DNA sequencing results revealed one additional surprise: *U2af1*(S34F) was absent in the DNA of LK cells from the transplant recipients from cases #2 and #3 that developed late AML (Fig. S5F and S5G, top panels). This was true despite the facts that Cre-mediated recombination was complete in the late AML cells and there was no evidence of focal deletions by PCR (Fig S5F and S5G, bottom panels; data not shown). The LK cells from these recipients were also GFP^+ (data not shown) and showed complete deletion of the floxed exon 4 in the $Runx1^F$ alleles (Fig. S5D), further confirming that these cells originated from the donors. The absence of *U2af1(S34F)* was also confirmed by the allele-sensitive Taqman assay, Sanger sequencing, and mRNA sequencing (Fig. S5H and data not shown). In contrast, *U2af1(S34F)* was present in LK cells from donor mice (albeit at a lower VAF), in matched lymphocytes and tails, and in the LK cells from Case #1 with primary AML (Figs S5E and S5H and data not shown). Moreover, the AML cells from Case #1, but not those from Cases #2 and #3, exhibited *U2af1*(S34F)-associated splicing changes in the consensus sequences preceding the altered cassette exons (data not shown), providing functional evidence for the presence or absence of *U2af1*(S34F) in these samples. These findings are considered further in the Discussion.

DISCUSSION

Here we report the establishment of genetically modified mouse strains that express a commonly observed mutant allele, *U2af1*(S34F), from the endogenous *U2af1* locus in a Cre-dependent manner (Figs 1A and S1A). We show that production of U2af1(S34F) from one of the two *U2af1* loci in heterozygous (S34F/WT) mice alters pre-mRNA splicing in mouse cells in ways similar to those observed in *U2AF1*(S34F)-expressing human cells, as previously reported by us and others (Fig. 2 and Table S3) (Fei et al., 2016; Ilagan et al., 2015; Okeyo-Owuor et al., 2015; Brooks et al., 2014; Przychodzen et al., 2013; Shirai et al., 2015). In the heterozygous mice, hematopoiesis is impaired, producing multi-lineage cytopenia, macrocytic anemia, and myeloid dysplasia (Figs 3 – 6, and S2); these features are hallmarks of human MDS, a class of disorders in which mutant U2AF1, including U2AF1(S34F), or other mutant splicing factors are often found.

We also observed myeloid leukemogenesis in three of fourteen mice in which *U2af1*(S34F) is combined with homozygous deletion of *Runx1* and exposure to a strong mutagen, ENU, but not in mice lacking any of those three factors (Figs 7, 8, S4, and Table 1). Whole exome sequencing of DNA from the three AMLs identified additional mutations associated with human AML. However, the *U2af1*(S34F) missense mutation was absent in two of the three AML cases (Figs 9 and S5; Tables S5 and S7), suggesting that *U2af1*(S34F) may be involved in the initiation but not the maintenance of AML.

Overall, the mice we describe here are useful for studying the contributions of mutant U2AF1 to MDS and MDS-related AML, and they may provide disease-relevant models for testing therapeutic options.

U2af1(S34F) changes RNA processing in ways both common to and different from changes observed in human cells carrying U2AF1(S34F)

Specific alterations in pre-mRNA splicing are the most widely characterized consequences of expressing *U2AF1*(S34F) in human cells (Fei et al., 2016; Ilagan et al.,

2015; Okeyo-Owuor et al., 2015; Brooks et al., 2014; Przychodzen et al., 2013; Shirai et al., 2015). In this report, we show that *U2af1*(S34F) causes similar but not identical changes in mRNA splicing in both mouse fibroblasts and MP cells. Less than 5% of cassette exons show at least a 10% change in rates of inclusion in mouse cells expressing *U2af1*(S34F) (Figs 2D and S1E), numbers similar to those we previously observed in human cells expressing *U2AF1*(S34F) (Fei et al., 2016; Ilagan et al., 2015). Moreover, the consensus 3' splice site sequences that precede the affected exons in mouse and human cells are very similar or identical (Figs 2D and S1E) (Fei et al., 2016; Ilagan et al., 2015). These similarities likely reflect the highly conserved sequence of U2AF1 protein and other components of the spliceosome machinery. For example, human and mouse U2AF1 proteins are nearly (more than 99%) identical, and the S34F mutations probably have similar effects on the two homologous proteins, recognizing highly similar 3' splice sites in mouse and human transcriptomes.

Despite the similar effects of S34F mutations on sequence recognition and genres of altered splicing, the mouse mRNAs altered by U2af1(S34F) are not necessarily the homologs of the human mRNAs altered by U2AF1(S34F) (Fig. 2C and E). This is not entirely surprising, considering that the consensus motif recognized by U2AF1(S34F)-containing complex is intronic, and introns are generally not as highly conserved as coding sequences. Similar observations have been made previously in transgenic mice expressing *U2AF1*(S34F) (Shirai et al., 2015) and in mice expressing other mutant splicing factors (Kim et al., 2015; Kon et al., 2018; Obeng et al., 2016; Mupo et al., 2016). Despite many differences (roughly 20% of genes with mRNAs significantly affected by the S34F mutation in mouse and human cells are homologs), a few of the genes with altered mRNAs in both species are often mutated in human MDS and AML or are partners in gene fusions in myeloid disorders (*PICALM*, *GNAS*, *KDM6A* and *BCOR*) (Caudell and Aplan, 2008; Bejar et al., 2011; Jankowska et al., 2011; Damm et al., 2013). Another gene (*RAC1*) plays an important role in HSCs (Gu et al., 2003; Cancelas et al., 2005), and one more (*H2AFY*) encodes isoforms reported to affect erythropoiesis (Yip et al., 2017). It remains to be determined, however, whether any of the changes in spliced mRNAs from these genes has a role in the hematologic abnormalities we have observed in our *U2af1*(S34F) mouse model.

Mutant U2AF1 has also been claimed to cause neoplasia by compromising mitophagy through alternative polyadenylation (APA) of *ATG7* mRNA (Park et al., 2016) or by increasing the formation of R-loops (Chen et al., 2018). While we observed no evidence of effects of *U2af1*(S34F) on *Atg7* APA selection in mouse bone marrow cells (Fig. 2C), it remains to be determined whether R-loop formation is enhanced in mouse cells containing *U2af1*(S34F).

U2af1(S34F) confers MDS-like features on mouse hematopoiesis but does not cause full-fledged MDS

We have found that expression of *U2af1*(S34F) from the endogenous locus in our mice impairs hematopoiesis, with features reminiscent of human MDS (Figs 3 – 6). We also note similarities and differences in the hematopoietic phenotypes in our mice and the phenotypes in a previously reported transgenic mouse model (Shirai et al., 2015), in which U2AF1(S34F) is driven by a tetracycline-responsive element in the *Col1a1* locus. Cytopenia is a common feature of both models, but we observed cytopenia in all leukocyte lineages, whereas in the transgenic model, the affected cell populations are restricted to monocytes and B cells. Both models manifest defective HSC function since they compete poorly with co-transplanted normal HSCs in irradiated recipients. However, we observed decreased numbers of hematopoietic stem and progenitor cells (including LSK, ST-HSC and LT-HSC populations) in our mice, while the number of LSK cells was increased in the transgenic model. Furthermore, our mice developed mild macrocytic anemia and low-grade myeloid dysplasia, which were not present in the transgenic model. Factors contributing to these differences might include the levels of expression of the mutant proteins (mutant and WT proteins are at equal levels in our model but amounts may differ in the transgenic model) and the methods for inducing the mutant protein (treatment with poly (IC) to activate the *Mx1-Cre* transgene in our mice as opposed to addition of doxycycline to activate the rtTA transcriptional regulator in the transgenic model).

Mice engineered to produce mutant splicing factors other than mutant U2AF1

(Kim et al., 2015; Kon et al., 2018; Obeng et al., 2016; Mupo et al., 2016) display similar features, including cytopenia and a competitive disadvantage of mutant-containing HSCs in re-population assays. Therefore, different mutant splicing factors may elicit effects on hematopoiesis that are both common to splicing factor mutations and specific for certain factors.

Regardless of the hematopoietic phenotypes produced by different mutant splicing factors, none of the mouse models (including ours) recapitulates all aspects of human MDS. These results suggest that a mutant splicing factor is insufficient to drive human MDS on its own. Similarly, expression of an MDS-associated *SRSF2* mutant in human induced pluripotent stem cells did not affect the production of hematopoietic progenitor cells and only mildly affected hematopoiesis. In contrast, a much more drastic phenotype was produced by deletion of chromosome 7q, another common but much more extensive genetic alteration in MDS (Chang et al., 2018). These observations are consistent with findings that mutant *U2AF1* is often associated with other genetic alterations, such as deletion of chromosome 20q and mutations in *ASXL1*, *DNMT3A*, *TET2*, and *RUNX1* in human MDS (Graubert et al., 2012; Yoshida et al., 2011; Thol et al., 2012; Damm et al., 2012; Papaemmanuil et al., 2013; Xu et al., 2014; Adema et al., 2016).

On the other hand, mutations in *U2AF1* or other splicing factor genes are found at a lower frequency in healthy individuals with so-called Clonal Hematopoiesis of Indeterminate Potential (or CHIP) than they are in MDS and secondary AML (Steensma et al., 2015). HSCs from individuals with CHIP often carry loss-of-function mutations in *DNMT3A* and *TET2*, mutations that often co-occur with *U2AF1* mutations in MDS. Since loss of *DNMT3A* or *TET2* enhances HSC renewal and transplantability, the combination of *U2af1*(S34F) and loss of *DNMT3A* or *TET2* might produce hematopoietic phenotypes in mice that more precisely resemble human MDS. The occurrence of a *Tet2* splice site mutation in a case of AML in our mouse model (Fig. 9A) may reflect such cooperativity between mutant *U2af1* and loss of *Tet2*.

Does U2af1(S34F) promote initiation of myeloid leukemia?

We have pursued the hypothesis that additional genetic changes are necessary to produce severe, progressive hematological disease in mice expressing *U2af1*(S34F) in the blood lineage. We first sought evidence of cooperativity between *U2af1*(S34F) and deletion of *Runx1* because *U2AF1* missense mutations and *RUNX1* loss-of-function mutations are reported to co-occur in human MDS and AML (Xu et al., 2014; Adema et al., 2016), and we found that mice carrying both gene alterations showed hematopoietic defects observed in mice with each of the individual alterations. But mice with both mutations did not develop leukemia until we further treated these mice with a chemical mutagen (ENU) to cause additional genetic changes that may be necessary for leukemic transformation. While *U2af1*(S34F), *Runx1* deficiency, or ENU treatment alone, and combinations of any two of these, were insufficient to cause leukemia in our relatively small cohorts, *bona fide* AML did arise at low frequency in compound URC mice treated with ENU, either in an ENU-treated animal in one case or in syngeneic recipients transplanted with bone marrow or spleen cells from two other ENU-treated URC mice (Figs 7 and 8; Table 1). Although a larger cohort is necessary to reach statistical significance and confirm these findings (see footnote to Table 1), our results are consistent with the idea that *U2af1*(S34F) plays a necessary but not sufficient role in leukemogenesis in this setting. The results are also consistent with the clinical observation that MDS patients with a *U2AF1* mutation are more likely to undergo disease progression toward AML (Graubert et al., 2012).

We were surprised to find that in two cases of AML---the two detected after bone marrow transplantation---the leukemic cells contained only *U2af1* RNA with a wild-type coding sequence (Fig. S5E - S5H). The cells retained the Cre-recombined *MGS34F* allele (Fig. S5F and S5G) but codon 34 was composed of TCT (serine) rather than TTT (phenylalanine). A T-to-C transition might have occurred during exposure to ENU treatment or randomly during the leukemogenic process, and conferred a selective advantage of the kind demonstrated in the competitive transplantation experiments shown in Fig. 5. Regardless of the mechanism and significance of the loss of *U2af1*(S34F), our results suggest that *U2af1*(S34F) is dispensable for maintenance of murine AML in this setting. This is consistent with our previous observation that

U2AF1(S34F) is dispensable for the growth of established human lung adenocarcinoma cell lines carrying this mutation (Fei et al., 2016).

ENU-induced mutations that likely cooperate with *U2af1*(S34F) and *Runx1* deletion in leukemogenesis

Exome sequencing of leukemic cells from the three cases of AML derived from our mouse model identified at least four somatic mutations that are likely to have contributed functionally to leukemogenesis (Fig. 9):

(i) The *Tet2* variant in AML case #1 affects a splice donor site of a constitutive exon (exon 9, as shown in Fig. 9B) so that exon 9 is skipped during splicing (Fig. 9C). Loss of exon 9 (138 bp in length) does not change the reading frame in the Tet2 coding sequence, but loss of the 2-oxoglutarate-Fe(II)-dependent dioxygenase domain, partly encoded by exon 9, likely eliminated normal Tet2 function.

(ii) The murine *Gata2* variant in case #1 AML is equivalent to *GATA2*(R362Q) (Fig. 9E), which is one of the recurrent *GATA2* mutations in AML [COSMIC ID: COSM87004; (Luesink et al., 2012; Yan et al., 2011)]. *GATA2* encodes a transcription factor that plays a critical role during hematopoiesis.

(iii) The *Idh1* mutation identified in one of the late AML cases (Case #2) is equivalent to human *IDH1*(R132Q) (Fig. 9H). R132 mutants of IDH1, including IDH1(R132Q) produce the "onco-metabolite" 2-hydroxyglutarate (Fig. 9J) that is commonly found elevated in several human cancers including AML (Dang et al., 2009).

(iv) The variant allele, *Ikzf1*(C175R), found in one case of late AML (Case #3), affects a conserved cysteine in a zinc finger domain of the transcription factor it encodes (Fig. 9L) and thus likely compromises (or alters) the DNA binding capacity of Ikzf1. *IKZF1*, a known tumor suppressor gene for lymphoid neoplasms (Mullighan et al., 2009), is also recurrently deleted in myeloproliferative neoplasms (Jäger et al., 2010) and pediatric AML (de Rooij et al., 2015). In combination with our results, it seems likely that *IKZF1* is a tumor suppressor gene for myeloid cancers as well.

These likely driver mutations from the three AML cases in our mice occurred largely independently, but some appear to be functionally related, implying that they may affect hematopoietic cell functions that must be altered to produce murine AML. For example, mutant IDH1 is known to inhibit TET2 enzyme activity (Figueroa et al., 2010), and Gata*2* and Ikzf1 (as well as Runx1) are transcription factors critical for hematopoietic lineage specification and/or HSC function. Mutant U2AF1 might cooperate with alterations of the epigenetic modifiers and lineage transcription factors to initiate leukemogenesis.

We have used the mice described here to study the effects of U2AF1(S34F) on hematopoiesis and to attempt to model human MDS and AML. Because production of U2af1(S34F) is dependent on Cre, which can be made in a lineage-specific manner in any cell type carrying an appropriate Cre transgene, our models can be used to study the roles of U2af1(S34F) in other cancers, such as lung adenocarcinoma and hairy cell leukemia in which *U2AF1(S34F)* is recurrently found (Imielinski et al., 2012; Cancer Genome Atlas Research Network, 2014; Waterfall et al., 2014).

MATERIALS AND METHODS

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 > TCT). 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. S1A, 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 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 to identify those clones that had undergone correct homologous recombination (Figs 1B and S1B) 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 blots (Figs 1B and S1B). 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 $Runx1^{F}$ (Growney et al., 2005) alleles as specified in the Results section for desired genotypes. All alleles, except for $Runx1^{F}$, were used in a heterozygous (or hemizygous) state in this study. Mice carrying the *U2af1* targeted alleles were all backcrossed to B6 mice for at least three generations before they were used for the reported experiments, and were considered to have a mixed B6,129 background. Littermates of animals with experimental genotypes were used as controls in all described studies.

Mouse procedures

All procedures related to mouse husbandry were approved by the Institutional Animal Care and Use Committees at the National Human Genome Research Institute (NHGRI, Protocols: G-13-1 and G-95-8) and at Weill Cornell Medicine (WCM, Protocol: 2015-017).

Peripheral blood sampling and CBC analysis: Blood was collected by either submandibular or retro-orbital bleeding. The former procedure was done as described (Golde et al., 2005), 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 (Figs 4 and S2), 2 - 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 via the tail vein. For competitive transplants (Fig. 5), 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 via the tail vein. For transplantation of bone marrow or spleen cells in Figs. 7, 8, 9 and S4, 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 instructions. Poly (IC) (three doses of 250ug per mouse) was given every other day via intraperitoneal injection. ENU treatment was as previously described (Castilla et al., 1999). 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*, *IES34F/WT* or *mTmG/WT* (JAX stock number: 007576) 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 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.

Cell culture, reagents, and assays

MEFs: MEFs were cultured as described above. To induce Cre activity, sub-confluent cells were treated with 4OHT (H7904, Sigma-Aldrich) at the indicated doses for 24 hr. The drug was then removed, and cells were cultured for an additional 48 hr before harvest. In other instances, the same treatment was repeated before harvest to improve the efficacy of Cre-mediated recombination.

Colony forming cell (CFC) assays: 30,000 nucleated bone marrow cells were seeded in MethoCult GF M3434 medium (StemCell Technologies) in a 60 mm dish in triplicates, according to the vendor's instructions. The number of colonies --- including erythroid (BFU-E), granulocyte-monocyte (CFU-GM) and multipotential (CFU-GEMM) colonies --- were counted 14 days later (or 7 days later when secondary plating was performed). When secondary plating was performed, primary CFC cultures from each set of triplicate cultures were harvested on day 8, pooled, and washed with IMDM containing

2% FBS. 30,000 or 10,000 nucleated cells were again seeded in a 60 mm dish with MethoCult GF M3434 medium in triplicate. The colonies were counted 12 days later. CFC assays was performed either as a contracted service (StemCell Technologies) or by the investigators.

Southern blot: Southern blot was performed as previously described (Fei et al., 2016). 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 S1A), 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 (Fei et al., 2016). 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 QuantaSoft Software (Bio-Rad) and results were presented as percent inclusion of the indicated exon. The allele-sensitive S34F/WT SNP Taqman assay was custom synthesized (Applied Biosystems) and characterized in Fig. S1I - S1K. The plasmids for calibrating the assay contained *U2af1*(WT) or *U2af1*(S34) cDNA, and were described previously (Fei et al., 2016).

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 (Growney et al., 2005).

Fragment analysis: Fragment analysis was used as an alternative method to Southern blotting to measure the extent of Cre-mediated recombination (Figs S1G and S1H). 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 (Carrington et al., 2015), 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 (Dutra et al., 2010). The GFP probe was made from a GFP cDNA released from the pEGFR-C1 vector by digestion with Nhe I and EcoR I (New England Biolabs).

Cell surface and intracellular staining and flow cytometry: Cell surface staining was conducted as previously described (Zhen et al., 2017). RBCs in blood, bone marrow or spleen cell suspensions were lysed using ACK lysing buffer (Quality Biological). 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 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.

Metabolite quantification by LC-MS/MS: Two million live splenocytes from Case #1 - #3 recipient mice at the moribund stage were used to quantify for soluble metabolite by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) at the Proteomics and Metabolics core lab (WCM) as a paid service.

High throughput nucleotide sequencing

mRNA sequencing: mRNA sequencing (RNAseq) was conducted at either the Sequencing Facility (NCI) or the Genome Resources Core Facility (WCM). RNA quality was assessed by a Bioanalyzer (Agilent). Total RNA with good integrity values (RIN > 9.0) was used for poly A selection and library preparation using the Illumina TruSeq RNA library preparation kit. The MEFs samples were run on a HiSeq2500 instrument and sequenced to the depth of 100 million paired-end 101 bp reads per sample. All other samples were run on a HiSeq4000 instrument and sequenced to the depth of 50 million paired-end 51 bp reads per sample.

Whole exome sequencing (WES): WES was performed in the WCM Genome Resources Core Facility. DNA quality was assessed with a Bioanalyzer (Agilent). A pre-hybridization library was prepared using the KAPA LTP library preparation kit (Kapa Biosystems). Targets were enriched by NimbleGen SeqCap EZ Exome Library V2.0. Up to 18 samples were sequenced in one HiSeq4000 lane to generate paired-end 101 bp reads.

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 (Cheng et al., 2015), 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 (Ilagan et al., 2015; Fei et al., 2016). Briefly, reads were first mapped to known mouse or human transcripts using RSEM (Li and Dewey, 2011). 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

(Trapnell et al., 2009). The number of “reads” were then subjected to analysis by the Wagenmakers’s Bayesian alternative to the binomial proportion test (Wagenmakers et al., 2010). 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 (Katz et al., 2010). 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 (Gentleman et al., 2004). 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. 9C) was also generated using the IGV browser.

For the analysis of whole exome sequencing data, raw sequencing reads were edited to remove low-quality bases, and adapter sequences were identified using cutadapt (Martin, 2011). The processed reads were then mapped to the mouse GRCm38 reference genome using BWA (Li, 2013). The alignments were further processed for local realignment and base quality score recalibration using GATK (McKenna et al., 2010). Somatic SNVs and small insertions and deletions (“Indels”) were detected using Varscan2 (Koboldt et al., 2012).

Analysis of targeted gene panel sequencing (mIMPACT, Table S7) was described previously (Cheng et al., 2015). 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.

Statistics

Statistical significance was determined by two-tailed Student's t test using GraphPad Prism 6 or otherwise stated. In all analyses, p values ≤ 0.05 are considered statistically significant.

LEGENDS

Figure 1. Creation of a conditional allele, *MGS34F*, at the endogenous *U2af1* locus. (**A**) Diagrams of the endogenous *U2af1* locus (top line), the *MGS34F* targeting vector (second line) and modified alleles, with sites used for Southern- and PCR-based genotyping as identified below. After verification of mice carrying an *MGS34F-Neo* allele (see Materials and methods), the animals were crossed with mice expressing *Flippase* (*Flp*) to remove the *Neomycin* (*Neo*) cassette. Numbers in boxes indicate exons or exonic sequences in cDNA; lines represent introns; STOP denotes a 3x transcriptional stop signal from SV40; E, EcoRI site; B, BamHI site. The red version of Exon 2 encodes the S34F mutation. (**B**) Germline transmission of the *MGS34F-Neo* allele. Southern blots of EcoRI and BamHI digested tail DNA from three engineered mice (1, 2, 3) and of DNA from WT mouse tails and the embryonic stem cell clone from which the mouse line originated (ES) are shown. (**C, D**) Cre-mediated recombination within the *MGS34F* allele documented by Southern blot. Panel **C**, genomic DNA from MEFs with the indicated genotype, producing 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 (x 2) before harvest. Panel **D**, 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.

Figure 2. U2af1(S34F)-associated changes in RNA splicing profiles in mouse cells resemble changes observed in human cells expressing *U2AF1*(S34F). (**A, B**) Expression of *U2af1*(S34F) as measured by an allele-specific Taqman assay (characterized in **Fig. S1I - S1K**). (A) The percentages of *U2af1* RNA derived from the *U2af1*(S34F) allele were determined using whole cell RNA from the same MEFs tested in Fig. **1C**. (**B**) The percentages of *U2af1* RNA derived from the mutant allele were determined with RNA from the bone marrow cells used for Fig.**1D**. (**C**) U2af1(S34F) changes the inclusion levels of indicated exons (or portions of exons, see the inserted

cartoons) found at six loci of genes (named in boxes) but does not alter choice of polyadenylation site for *Atg7* mRNA (top right) in total bone marrow cells as in Panel **B**. Genotypes are indicated by color-coding (right lower corner). As noted in the text, alternative splicing of human homologs of these mRNAs was previously reported to be affected by U2AF1(S34F). Asterisks indicate statistically significant changes compared to other genotypes by t test ($p < 0.05$). Error bars represent standard error of the mean (s.e.m). n = 3 for each genotype. (**D**) U2af1(S34F) recognizes similar consensus sequences at 3' splice sites in mouse genome as U2AF1(S34F) does in human genome. mRNA from MEFs and MPs with or without *U2af1*(S34F) was sequenced (as described in Methods) 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 panel **A,** and the MPs were isolated from *MGS34F/WT*; *Mx1-Cre* or *Mx1-Cre* mice, treated with poly (IC) one month before harvest (n = 4 for each genotype). The resemblance of these logos to those previously determined with human materials is discussed in the text. (**E**) The Venn diagram indicates the numbers of orthologous genes from mouse and human datasets that show at least 10% change in cassette exon inclusion levels in the presence of mutant U2AF1. The human dataset was published previously (Ilagan et al., 2015), using AML samples from TCGA (The Cancer Genome Atlas Research Network, 2013). Genes with low levels of expression (median transcripts per million sequenced RNAs < 10) or have no mouse or human ortholog were excluded from the analysis.

Figure 3. U2af1(S34F) causes multi-lineage cytopenia, macrocytic anemia, and reduction in the hematopoietic stem and progenitor cell populations. (**A** - **G**) Multilineage cytopenia and macrocytic anemia in mice expressing *U2af1*(S34F). Blood samples taken from mice of the indicated genotypes (see keys in panel **A**) before and up to 24 weeks after poly (IC) treatment were subjected to complete blood count to determine numbers of red blood cells (**A**), mean red blood cell volume (**B**), and white blood cells (**C**); and used for flow cytometry (panels **D** - **G**) to determine concentrations of cells with indicated lineage markers. Asterisks indicate significant changes for mice

with the *MGS34F/WT; Mx1-Cre* genotype as compared to mice of any other genotype by multiple t test (FDR < 0.05). (**H** - **L**) *U2af1*(S34F) reduces the percentage of hematopoietic stem and progenitor cells in the bone marrow. Bone marrow cells from mice of the indicated genotypes were harvested 36 weeks after poly (IC) treatment, and the percentages of hematopoietic stem and progenitor cells were determined by flow cytometry using the stated markers. (**H**) Representative results of flow cytometry with numbers indicating the percentages of live nucleated cells in boxed areas. (**I** – **L**) Bar heights indicate the mean percentages of cells with the designated phenotypes (vertical axes) in the bone marrow. A dot indicated the percentage of cells in a mouse. Asterisks denote significant changes by t test ($p < 0.05$). N.S. not significant. Error bars represent s.e.m. See **Supplemental Figure S2** (panels **A** - **H**, **Q**, and **R**) for additional characterization of cells from these mice. *WT/WT* (n = 10), *WT/WT; Mx1-Cre* (n = 10), *MGS34F/WT* (n = 10), *MGS34F/WT; Mx1-Cre* (n = 11).

Figure 4. Hematopoietic cells from *U2af1(S34F)* mice are defective in mice receiving non-competitive transplants. (**A** - **G**) 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. 3**. (**H** - **L**) *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 (**H**), $Lin^-Scal\text{-}1^-c\text{-}Kit^+$ (LK) cells (**I**) and $Lin^-Scal\text{-}1^+c\text{-}Kit^+$ (LSK) cells **(J**). (**K)** 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 **D** - **K**, only donor cells (gated on GFP^+) were quantified. (**L)** 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. Asterisks indicate significant changes by t test. N.S. not significant. Error bars represent s.e.m. *MGS34F/WT* (n = 10 for

panels **A** - **G**; n = 5 for panels **H** - **L**), *MGS34F/WT; Mx1-Cre* (n = 11 for panels **A** - **G**; n = 5 for panels **H** - **L**). See **Fig. S2**, panels **I** - **P**, **S**, and **T** for additional characterization of hematopoietic cells from these mice.

Figure 5. Hematopoietic stem and progenitor cells with U2af1(S34F) are eliminated in a competitive transplant assay with cells lacking mutant U2af1. (**A**) Assay design. Equal numbers (0.5 million nucleated cells) of donor (*MGS34F/WT* or *MGS34F/WT; Mx1-Cre,* both GFP^+) and competitor (*WT* or *Mx1-Cre* alone, GFP^-) bone marrow cells were mixed and transplanted to lethally irradiated, wild-type recipient mice. Recipient mice were treated with poly (IC) four weeks after transplant. (**B** - **F**) Contributions of donor cells to hematopoietic compartments of recipient mice. (**B** and **C**) *U2af1*(S34F)-containing BM cells from *MGS34F/WT; Mx1-Cre* mice, but not control BM cells from *MGS34F/WT* mice, constituted less of the white blood cells with BM competitors from *Mx1-Cre* (**B**) or wildtype mice (**C**) as measured by percentages of GFP^+ cells in peripheral blood in the recipient mice in a competitive transplant. (**D** - **F**) All measured cell lineages showed competitive disadvantages of *U2af1*(S34F)-containing cells in the indicated cell lineages in the blood (**D**), bone marrow (**E**), and spleen (**F**) 21 weeks after poly (IC) treatment. n = 5 - 10 for each group of mice. Asterisks indicate significant changes by t test. Error bars represent s.e.m.

Figure 6. *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 bone segments were evaluated for each mouse. Asterisks indicate significant changes by t test. Error bars represent s.e.m. n = 5 for each genotype.

Figure 7. Myeloid hyperplasia and reduced life span in ENU-treated, *RUNX1* deficient mice and a case of primary AML in one URC mouse. (**A**) Strategy for testing leukemogenesis. Mice with the genotypes in Table 1 were treated with poly (IC) and, one week later, with ENU (100 mg/kg). The mice were monitored until they were moribund or up to 1.5 years after receiving poly (IC) and then examined for hematological abnormalities. (**B**) 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 (See **Table 1** for greater detail about the genotypes). (**C**) 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. (**D** - **F**) A case of primary AML in one of fourteen poly (IC)- and ENU-treated URC mice (Case #1). (**D**) A high percentage of c-Kit^{+}/Mac1^{-} cells in the peripheral blood by flow cytometry. (**E**) Infiltration of morphologically defined myeloblasts (arrow head) in a Wright-Giemsa stained blood smear, and H&E stained sections of bone marrow, spleen, and liver. Scale bar = 50 μm. (**F**) Rapid death of sub-lethally irradiated mice after transplants of bone marrow (three mice) or spleen cells (four mice) from Case #1. Day count started since after the transplantation.

Figure 8. Secondary late AML in mice after transplant of splenic cells from two poly(IC)- and ENU-treated URC mice. (**A**) Strategy. Splenic or bone marrow cells from

moribund mice without frank AML were transplanted to sub-lethally irradiated, wild-type recipient mice. Three to five recipient mice were used for each transplant and monitored for up to one year for the development of AML. (**B**, **C**) Early mortality of recipient mice (n=3) that received spleen or bone marrow cells from two end-stage URC mice --- Cases #2 (**B**) and #3 (**C**). Day count started since after the transplantation. (**D**) Representative flow cytometry results for the Case #3 donor (top row) and one of transplant recipients at end stage (bottom row). Percentages of c-Kit$^+$ cells were markedly increased in the peripheral blood, bone marrow and spleen of the recipient mouse, compared to those in the donor. (**E**) Representative images of the Wright-Giemsa stained blood smear and H&E stained sections of bone marrow, spleen and liver from Case #3 mouse (top panel) and from one of the transplant recipients (bottom panel) when they were moribund. Abnormally nucleated red blood cells (arrow) were observed in the blood smear from both the donor and recipient, while myeloblasts (arrow head) were only observed in blood from the recipient. Blast cells also infiltrated the bone marrow, spleen and liver of the recipient mouse. Scale bar = 50 μm. See **Fig. S4** (panels **C** - **F**) for additional characterization of these cases of late AML.

Figure 9. Acquired somatic mutations in Cases #1 - 3 AML cells affecting mouse homologs of human cancer genes. DNA was prepared from donor-derived (i.e. GFP$^+$) Lin$^-$c-Kit$^+$ cells and lymphocytes (CD3ε$^+$ or CD19$^+$) from the spleens of donor and matched transplant recipient mice from Cases #1 - 3 (see Figs **8** and **9**) and used for whole exome sequencing (WES) to identify somatic mutations (see Methods). Sequencing of RNA from some samples was performed to confirm the WES findings. (**A** - **F**) Acquired mutations in *Tet2* and *Gata2* in AML cells from Case #1. (**A** and **D**) Variant allelic frequencies (VAF) for *Tet2* and *Gata2* mutations in the indicated cells from the donor (D) and transplant recipient (R) mice. (**B** and **C**) The T to A mutation in *Tet2* disrupts a splice donor site at the 3' end of exon 9, producing transcripts in which exon 9 is omitted. (**E**) Comparison of mouse Gata2 and human GATA2 proteins, highlighting in red the R362 residue that was changed to Q by the missense mutation. (**F**) Confirmation of the *Gata2* (R362Q) mutation by mRNA sequencing from AML cells from a recipient mouse. (**G** - **J**) An *Idh1*(R132Q) mutation in AML cells from Case #2. (**G**)

VAFs for the R132Q allele in DNA from donor and recipient mice. (**H**) Position of the mutated amino acid in mouse Idh1 and human IDH1 proteins. (**I**) Confirmation of *Idh1*(R132Q) by sequencing of RNA from AML from a recipient mouse. (**J**) AML cells with *Idh1*(R132Q) had elevated levels of 2-hydroxyglutarate, as measured by LC-MS/MS. (**K** - **M**) An *Ikzf1*(C175R) mutation in Case #3. (**K**) VAFs of the C175R allele in donor and recipient mice. (**L**) A cartoon illustrating the position of C175 in the third zinc finger domain of Ikzf1. (**M**) Confirmation of the *Ikzf1*(C175R) mutation by sequencing RNA from AML cells from a recipient mouse. Additional characterizations of the acquired mutations are presented in **Fig. S5** and **Tables S5**, **S6** and **S7**.

Table 1. Summary of AML incidences in mice with ENU and poly (IC) treatment.

Supplemental Figure S1. The *IES34F* allele and genotyping assays. (**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 in **Fig. 1C**. 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 from the same MEFs tested in **Fig S1C**. (**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. (**F**) A GFP transgene was 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 was a representative metaphase HG3 cells by fluorescence in situ hybridization (FISH). The fluorescence signal for GFP was in red (boxed). Chromosomes 17, where *U2af1* resides, were marked in green. All chromosomes were stained with DAPI and colored in blue. (**G**, **H**) Estimation of efficiency of Cre-mediated recombination of the *MGS34F* allele by fragment analysis through high resolution capillary electrophoresis. (**G**) 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 wild-type *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. (**H**) 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. (**I** - **K**) Estimation of the *U2af1*(S34F) allele by an allele-sensitive S34F/WT Single Nucleotide Polymorphism (SNP) Taqman assay. (**I**) The primers (arrows) and probes (bar) are all within exon 2 *U2af1*, allowing detection of *U2af1*(S34F) specific sequences in genomic DNA and mRNA. (**J**) 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. (**K**) Efficiency of the PCR-based Taqman assay. Data from panel **K** were used to estimate the primer efficiencies

by the formula: Efficiency $=10^{(-1/slope)} - 1$. Ct, critical threshold cycle number. The sequences of the primers and probes used in the assays are available in **Table S2**.

Supplemental Figure S2. U2af1(S34F) affects hematopoiesis in mice. Panels **A** - **H, Q**, and **R** were for the same cohort of mice as described in **Fig. 3**. Panels **I** - **P**, **S** and **T** were for the same cohort of mice as described in **Fig. 4**. Panels **A** - **G**; **I** - **O** showed additional CBC and flow cytometry results for the blood samples. Panels **H** and **P** described the weight of mice over time. The numbers of nucleated bone marrow cell and spleen weight were quantified either 36 weeks (**Q**, **R**) or 24 weeks (**S**, **T**) after poly (IC) treatment from respective mouse cohort. In panel **Q**, 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. In panel **S**, bone marrow cells were isolated from one femur and two tibias of each mouse. Asterisks indicate significant changes by t test. Error bars represent s.e.m.

Supplemental Figure S3. 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**) 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). (**C** - **E**) Deletion of *Runx1* affects mainly gene expression, while *U2af1*(S34F) causes mainly changes in mRNA splicing. MP cells (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). (**C**) Non-hierarchical clustering based on RNA profiles; color codes indicate genotypes. (**D** and **E**) Venn diagrams depict the numbers of overlapping and non-overlapping genes with significant changes in gene expression (**D**) or mRNA splicing (**E**) as compared to samples with the control genotype *Mx1-Cre*.

Supplemental Figure S4. Consequences of *U2af1*(S34F) expression and *Runx1* deficiency in mouse hematopoiesis and leukemogenesis after ENU treatment. (**A**, **B**) *U2af1*(S34F) expression was associated with myeloid dysplasia for the erythroid and granulocytic lineages in the presence of *Runx1* deletion. Panel **A** 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 **B** was a summary of quantification for these myeloid dysplastic features. At least 500 cells were quantified for each mouse. (**C** - **F**) Further characterization of late AML cases related to **Fig. 8**. Panel **C** showed increased percentage of c-Kit$^+$ cells in peripheral blood, bone marrow and spleen of a recipient mouse in comparison to those of the donor mouse (Case #2) by flow cytometry, when they were moribund. In panel **D**, the histology of H&E-stained bone marrow and spleen section was compared between the donor and one of the recipient mice, when they were moribund. Blast cells occupied the bone marrow and spleen in the moribund recipient mice. Scale bar = 50 μm. (**E** and **F**) Survival curves for secondary transplant. The donor cells were the moribund recipient mice that developed late AML in either Case #2 (panel **E**, n = 5) or Case #3 (panel **F**, n = 4). Day count started since after the transplantation.

Supplemental Figure S5. Genetic alterations in AMLs from Cases #1 - 3. (**A**) Frequencies of single nucleotide substitutions in the Lin$^-$c-Kit$^+$ cell population from donor and recipient mice. TV, transversion; TI, transition. (**B** - **C**) Confirmation of somatic mutations by targeted panel sequencing using the Integrated Mutation Profiling of Actionable Cancer Targets, mouse version 1 (mIMPACT, see Methods). (**B)** Overlap of variants identified by WES and mIMPACT in DNA from Lin$^-$c-Kit$^+$ cells from a Case #1 recipient mouse (R2). (**C**) VAF of shared variants and those uniquely identified by

mIMPACT. (**D** - **H**) *Runx1* and *U2af1* alterations in the AML samples. (**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 Lin$^-$c-Kit$^+$ cells in all cases, but present in all tail DNAs. 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. S1**). *U2af1*(S34F) was present in all donor- and recipient-derived samples from Case #1 (**E**) but absent in Lin$^-$c-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**) *U2af1*(S34F) mRNA in Lin$^-$c-Kit$^+$ cells from transplant recipients. *U2af1*(S34F) mRNA was present in Lin$^-$c-Kit$^+$ cells from a Case #1 recipient but not in Lin$^-$c-Kit$^+$ cells from recipients of Cases #2 and #3.

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 (Koboldt et al., 2012).

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 (Cheng et al., 2015).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: DLF, TZ (Zhen), PL, HV.

Performed the experiments: DLF, TZ (Zhen).

Analyzed the data: DLF, TZ (Zhen), BD, TZ (Zhang), RKB, PL, HV.

Contributed reagents/materials/analysis tools: JF, TZ (Zhang), LG, AY, OAW, RKB.

Wrote the paper: DLF, HV.

ACKNOWLEDGMENTS

We thank Sukanya Goswami (WCM), Shao Ning Yang (WCM), Hayley Motowski (NHGRI), Jackie Idol (NHGRI), Danielle Miller-O'Mard (NHGRI), Ursula Harper (NHGRI), Amalia Dutra (NHGRI), Evgenia Pak (NHGRI), Stacie Anderson (NHGRI), Martha Kirby (NHGRI), David Bodine (NHGRI), Shelley Hoogstraten-Miller (NHGRI), Cecilia Rivas (NHGRI), and other members of the transgenic mouse core at NHGRI for exceptional technical assistance. We thank Nancy Speck (UPenn) for providing the conditional *Runx1* knockout mice. We thank members of the Varmus lab, the Liu lab, Guoan Zhang (WCM), Timothy A. Graubert (MGH), Pavankumar N.G. Reddy (MGH), Borja Saez (MGH), Matthew J. Walter (Wash.U.), Cara Lunn Shirai (Wash.U.), Dan Larson (NCI), Murali Palangat (NCI), and Markus Hafner (NIAMS) for helpful discussions during the course of the study.

HV was supported by the Intramural Program at the National Institutes of Health and is now supported by the Meyer Cancer Center at Weill Cornell Medicine and the Edward P. Evans Foundation. PL is supported by the Intramural Program at the National Institutes of Health.

REFERENCE

Adema, V., C.M. Hirsch, B.P. Przychodzen, A. Nazha, T. Kuzmanovic, E. Negoro, D. You, H. Makishima, M.J. Clemente, H.E. Carraway, M.A. Sekeres, V. Visconte, and J.P. Maciejewski. 2016. U2AF1 Mutations in S34 and Q157 Create Distinct Molecular and Clinical Contexts. *Blood*. 128:3155–3155.

Arber, D.A., A. Orazi, R. Hasserjian, J. Thiele, M.J. Borowitz, M.M. Le Beau, C.D. Bloomfield, M. Cazzola, and J.W. Vardiman. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 127:2391–2405. doi:10.1182/blood-2016-03-643544.

Bejar, R., K. Stevenson, O. Abdel-Wahab, N. Galili, B. Nilsson, G. Garcia-Manero, H. Kantarjian, A. Raza, R.L. Levine, D. Neuberg, and B.L. Ebert. 2011. Clinical effect of point mutations in myelodysplastic syndromes. *N. Engl. J. Med.* 364:2496–2506. doi:10.1056/NEJMoa1013343.

Brooks, A.N., P.S. Choi, L. de Waal, T. Sharifnia, M. Imielinski, G. Saksena, C.S. Pedamallu, A. Sivachenko, M. Rosenberg, J. Chmielecki, M.S. Lawrence, D.S. DeLuca, G. Getz, and M. Meyerson. 2014. A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events. *PLoS ONE*. 9:e87361. doi:10.1371/journal.pone.0087361.s008.

Cancelas, J.A., A.W. Lee, R. Prabhakar, K.F. Stringer, Y. Zheng, and D.A. Williams. 2005. Rac GTPases differentially integrate signals regulating hematopoietic stem cell localization. *Nat Med*. 11:886–891. doi:10.1038/nm1274.

Cancer Genome Atlas Research Network. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 511:543–550. doi:10.1038/nature13385.

Carrington, B., G.K. Varshney, S.M. Burgess, and R. Sood. 2015. CRISPR-STAT: an easy and reliable PCR-based method to evaluate target-specific sgRNA activity. *Nucleic Acids Res.* 43:e157. doi:10.1093/nar/gkv802.

Castilla, L.H., L. Garrett, N. Adya, D. Orlic, A. Dutra, S. Anderson, J. Owens, M. Eckhaus, D. Bodine, and P.P. Liu. 1999. The fusion gene Cbfb-MYH11 blocks myeloid differentiation and predisposes mice to acute myelomonocytic leukaemia. *Nature Genetics*. 23:144–146. doi:10.1038/13776.

Catenacci, D.V.T., and G.J. Schiller. 2005. Myelodysplasic syndromes: a comprehensive review. *Blood Rev.* 19:301–319. doi:10.1016/j.blre.2005.01.004.

Caudell, D., and P.D. Aplan. 2008. The role of CALM-AF10 gene fusion in acute leukemia. *Leukemia*. 22:678–685. doi:10.1038/sj.leu.2405074.

Chang, C.-J., A.G. Kotini, M. Olszewska, M. Georgomanoli, J. Teruya-Feldstein, H. Sperber, R. Sanchez, R. DeVita, T.J. Martins, O. Abdel-Wahab, R.K. Bradley, and

E.P. Papapetrou. 2018. Stem Cell Reports. *Stem Cell Reports*. 10:1610–1624. doi:10.1016/j.stemcr.2018.03.020.

Chen, L., J.-Y. Chen, Y.-J. Huang, Y. Gu, J. Qiu, H. Qian, C. Shao, X. Zhang, J. Hu, H. Li, S. He, Y. Zhou, O. Abdel-Wahab, D.-E. Zhang, and X.-D. Fu. 2018. The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. *Mol. Cell*. 1–30. doi:10.1016/j.molcel.2017.12.029.

Cheng, D.T., T.N. Mitchell, A. Zehir, R.H. Shah, R. Benayed, A. Syed, R. Chandramohan, Z.Y. Liu, H.H. Won, S.N. Scott, A.R. Brannon, C. O'Reilly, J. Sadowska, J. Casanova, A. Yannes, J.F. Hechtman, J. Yao, W. Song, D.S. Ross, A. Oultache, S. Dogan, L. Borsu, M. Hameed, K. Nafa, M.E. Arcila, M. Ladanyi, and M.F. Berger. 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*. 17:251–264. doi:10.1016/j.jmoldx.2014.12.006.

Damm, F., O. Kosmider, V. Gelsi-Boyer, A. Renneville, N. Carbuccia, C. Hidalgo-Curtis, V. Della Valle, L. Couronné, L. Scourzic, V. Chesnais, A. Guerci-Bresler, B. Slama, O. Beyne-Rauzy, A. Schmidt-Tanguy, A. Stamatoullas-Bastard, F. Dreyfus, T. Prébet, S. de Botton, N. Vey, M.A. Morgan, N.C.P. Cross, C. Preudhomme, D. Birnbaum, O.A. Bernard, M. Fontenay, Groupe Francophone des Myélodysplasies. 2012. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. 119:3211–3218. doi:10.1182/blood-2011-12-400994.

Damm, F., V. Chesnais, Y. Nagata, K. Yoshida, L. Scourzic, Y. Okuno, R. Itzykson, M. Sanada, Y. Shiraishi, V. Gelsi-Boyer, A. Renneville, S. Miyano, H. Mori, L.-Y. Shih, S. Park, F. Dreyfus, A. Guerci-Bresler, E. Solary, C. Rose, S. Cheze, T. Prébet, N. Vey, M. Legentil, Y. Duffourd, S. de Botton, C. Preudhomme, D. Birnbaum, O.A. Bernard, S. Ogawa, M. Fontenay, and O. Kosmider. 2013. BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders. *Blood*. 122:3169–3177. doi:10.1182/blood-2012-11-469619.

Dang, L., D.W. White, S. Gross, B.D. Bennett, M.A. Bittinger, E.M. Driggers, V.R. Fantin, H.G. Jang, S. Jin, M.C. Keenan, K.M. Marks, R.M. Prins, P.S. Ward, K.E. Yen, L.M. Liau, J.D. Rabinowitz, L.C. Cantley, C.B. Thompson, M.G. Vander Heiden, and S.M. Su. 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. 462:739–744. doi:10.1038/nature08617.

de Rooij, J.D.E., E. Beuling, M.M. van den Heuvel-Eibrink, A. Obulkasim, A. Baruchel, J. Trka, D. Reinhardt, E. Sonneveld, B.E.S. Gibson, R. Pieters, M. Zimmermann, C.M. Zwaan, and M. Fornerod. 2015. Recurrent deletions of IKZF1 in pediatric acute myeloid leukemia. *Haematologica*. 100:1151–1159. doi:10.3324/haematol.2015.124321.

Dutra, A., E. Pak, S. Wincovitch, K. John, M.C. Poirier, and O.A. Olivero. 2010. Nuclear bud formation: a novel manifestation of Zidovudine genotoxicity. *Cytogenet. Genome Res.* 128:105–110. doi:10.1159/000298794.

Fei, D.L., H. Motowski, R. Chatrikhi, S. Prasad, J. Yu, S. Gao, C.L. Kielkopf, R.K. Bradley, and H. Varmus. 2016. Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival. *PLoS Genet*. 12:e1006384. doi:10.1371/journal.pgen.1006384.

Figueroa, M.E., O. Abdel-Wahab, C. Lu, P.S. Ward, J. Patel, A. Shih, Y. Li, N. Bhagwat, A. Vasanthakumar, H.F. Fernandez, M.S. Tallman, Z. Sun, K. Wolniak, J.K. Peeters, W. Liu, S.E. Choe, V.R. Fantin, E. Paietta, B. Löwenberg, J.D. Licht, L.A. Godley, R. Delwel, P.J.M. Valk, C.B. Thompson, R.L. Levine, and A. Melnick. 2010. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 18:553–567. doi:10.1016/j.ccr.2010.11.015.

Furney, S.J., M. Pedersen, D. Gentien, A.G. Dumont, A. Rapinat, L. Desjardins, S. Turajlic, S. Piperno-Neumann, P. de la Grange, S. Roman-Roman, M.H. Stern, and R. Marais. 2013. SF3B1 Mutations Are Associated with Alternative Splicing in Uveal Melanoma. *Cancer Discovery*. 3:1122–1129. doi:10.1158/2159-8290.CD-13-0330.

Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y.H. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80. doi:10.1186/gb-2004-5-10-r80.

Golde, W.T., P. Gollobin, and L.L. Rodriguez. 2005. A rapid, simple, and humane method for submandibular bleeding of mice using a lancet. *Lab Anim (NY)*. 34:39–43. doi:10.1038/laban1005-39.

Graubert, T.A., D. Shen, L. Ding, T. Okeyo-Owuor, C.L. Lunn, J. Shao, K. Krysiak, C.C. Harris, D.C. Koboldt, D.E. Larson, M.D. McLellan, D.J. Dooling, R.M. Abbott, R.S. Fulton, H. Schmidt, J. Kalicki-Veizer, M. O'Laughlin, M. Grillot, J. Baty, S. Heath, J.L. Frater, T. Nasim, D.C. Link, M.H. Tomasson, P. Westervelt, J.F. DiPersio, E.R. Mardis, T.J. Ley, R.K. Wilson, and M.J. Walter. 2012. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nature Genetics*. 44:53–57. doi:10.1038/ng.1031.

Growney, J.D., H. Shigematsu, Z. Li, B.H. Lee, J. Adelsperger, R. Rowan, D.P. Curley, J.L. Kutok, K. Akashi, I.R. Williams, N.A. Speck, and D.G. Gilliland. 2005. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*. 106:494–504. doi:10.1182/blood-2004-08-3280.

Gu, Y., M.-D. Filippi, J.A. Cancelas, J.E. Siefring, E.P. Williams, A.C. Jasti, C.E. Harris, A.W. Lee, R. Prabhakar, S.J. Atkinson, D.J. Kwiatkowski, and D.A. Williams. 2003.

Hematopoietic cell regulation by Rac1 and Rac2 guanosine triphosphatases. *Science*. 302:445–449. doi:10.1126/science.1088485.

Ichikawa, M., T. Asai, T. Saito, S. Seo, I. Yamazaki, T. Yamagata, K. Mitani, S. Chiba, S. Ogawa, M. Kurokawa, and H. Hirai. 2004. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med*. 10:299–304. doi:10.1038/nm997.

Ilagan, J.O., A. Ramakrishnan, B. Hayes, M.E. Murphy, A.S. Zebari, P. Bradley, and R.K. Bradley. 2015. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res.* 25:14–26. doi:10.1101/gr.181016.114.

Imielinski, M., A.H. Berger, P.S. Hammerman, B. Hernandez, T.J. Pugh, E. Hodis, J. Cho, J. Suh, M. Capelletti, A. Sivachenko, C. Sougnez, D. Auclair, M.S. Lawrence, P. Stojanov, K. Cibulskis, K. Choi, L. de Waal, T. Sharifnia, A. Brooks, H. Greulich, S. Banerji, T. Zander, D. Seidel, F. Leenders, S. Ansén, C. Ludwig, W. Engel-Riedel, E. Stoelben, J. Wolf, C. Goparju, K. Thompson, W. Winckler, D. Kwiatkowski, B.E. Johnson, P.A. Jänne, V.A. Miller, W. Pao, W.D. Travis, H.I. Pass, S.B. Gabriel, E.S. Lander, R.K. Thomas, L.A. Garraway, G. Getz, and M. Meyerson. 2012. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 150:1107–1120. doi:10.1016/j.cell.2012.08.029.

Jankowska, A.M., H. Makishima, R.V. Tiu, H. Szpurka, Y. Huang, F. Traina, V. Visconte, Y. Sugimoto, C. Prince, C. O'Keefe, E.D. Hsi, A. List, M.A. Sekeres, A. Rao, M.A. McDevitt, and J.P. Maciejewski. 2011. Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood*. 118:3932–3941. doi:10.1182/blood-2010-10-311019.

Jäger, R., H. Gisslinger, F. Passamonti, E. Rumi, T. Berg, B. Gisslinger, D. Pietra, A. Harutyunyan, T. Klampfl, D. Olcaydu, M. Cazzola, and R. Kralovics. 2010. Deletions of the transcription factor Ikaros in myeloproliferative neoplasms. *Leukemia*. 24:1290–1298. doi:10.1038/leu.2010.99.

Katz, Y., E.T. Wang, E.M. Airoldi, and C.B. Burge. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods*. 7:1009–1015. doi:10.1038/nmeth.1528.

Kim, E., J.O. Ilagan, Y. Liang, G.M. Daubner, S.C.W. Lee, A. Ramakrishnan, Y. Li, Y.R. Chung, J.-B. Micol, M.E. Murphy, H. Cho, M.-K. Kim, A.S. Zebari, S. Aumann, C.Y. Park, S. Buonamici, P.G. Smith, H.J. Deeg, C. Lobry, I. Aifantis, Y. Modis, F.H.T. Allain, S. Halene, R.K. Bradley, and O. Abdel-Wahab. 2015. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell*. 27:617–630. doi:10.1016/j.ccell.2015.04.006.

Koboldt, D.C., Q. Zhang, D.E. Larson, D. Shen, M.D. McLellan, L. Lin, C.A. Miller, E.R. Mardis, L. Ding, and R.K. Wilson. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22:568–

576. doi:10.1101/gr.129684.111.

Kon, A., S. Yamazaki, Y. Nannya, K. Kataoka, Y. Ota, M.M. Nakagawa, K. Yoshida, Y. Shiozawa, M. Morita, T. Yoshizato, M. Sanada, M. Nakayama, H. Koseki, H. Nakauchi, and S. Ogawa. 2018. PhysiologicalSrsf2P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood*. 131:621–635. doi:10.1182/blood-2017-01-762393.

Kühn, R., F. Schwenk, M. Aguet, and K. Rajewsky. 1995. Inducible gene targeting in mice. *Science*. 269:1427–1429.

Li, B., and C.N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12:323. doi:10.1186/1471-2105-12-323.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. q-bio.GN.

Luesink, M., I.H.I.M. Hollink, V.H.J. van der Velden, R.H.J.N. Knops, J.B.M. Boezeman, V. de Haas, J. Trka, A. Baruchel, D. Reinhardt, B.A. van der Reijden, M.M. van den Heuvel-Eibrink, C.M. Zwaan, and J.H. Jansen. 2012. High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. *Blood*. 120:2064–2075. doi:10.1182/blood-2011-12-397083.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* 17:10. doi:10.14806/ej.17.1.200.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303. doi:10.1101/gr.107524.110.

Merendino, L., S. Guth, D. Bilbao, C. Martínez, and J. Valcárcel. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature*. 402:838–841. doi:10.1038/45602.

Mullighan, C.G., X. Su, J. Zhang, I. Radtke, L.A.A. Phillips, C.B. Miller, J. Ma, W. Liu, C. Cheng, B.A. Schulman, R.C. Harvey, I.-M. Chen, R.J. Clifford, W.L. Carroll, G. Reaman, W.P. Bowman, M. Devidas, D.S. Gerhard, W. Yang, M.V. Relling, S.A. Shurtleff, D. Campana, M.J. Borowitz, C.-H. Pui, M. Smith, S.P. Hunger, C.L. Willman, J.R. Downing, Children's Oncology Group. 2009. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* 360:470–480. doi:10.1056/NEJMoa0808253.

Mupo, A., M. Seiler, V. Sathiaseelan, A. Pance, Y. Yang, A.A. Agrawal, F. Iorio, R. Bautista, S. Pacharne, K. Tzelepis, N. Manes, P. Wright, E. Papaemmanuil, D.G. Kent, P.C. Campbell, S. Buonamici, N. Bolli, and G.S. Vassiliou. 2016. Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS

but develop anemia without ring sideroblasts. *Leukemia*. doi:10.1038/leu.2016.251.

Obeng, E.A., R.J. Chappell, M. Seiler, M.C. Chen, D.R. Campagna, P.J. Schmidt, R.K. Schneider, A.M. Lord, L. Wang, R.G. Gambe, M.E. McConkey, A.M. Ali, A. Raza, L. Yu, S. Buonamici, P.G. Smith, A. Mullally, C.J. Wu, M.D. Fleming, and B.L. Ebert. 2016. Physiologic Expression of Sf3b1. *Cancer Cell*. 30:404–417. doi:10.1016/j.ccell.2016.08.006.

Okeyo-Owuor, T., B.S. White, R. Chatrikhi, D.R. Mohan, S. Kim, M. Griffith, L. Ding, S. Ketkar-Kulkarni, J. Hundal, K.M. Laird, C.L. Kielkopf, T.J. Ley, M.J. Walter, and T.A. Graubert. 2015. U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia*. 29:909–917. doi:10.1038/leu.2014.303.

Papaemmanuil, E., M. Cazzola, J. Boultwood, L. Malcovati, P. Vyas, D. Bowen, A. Pellagatti, J.S. Wainscoat, E. Hellstrom-Lindberg, C. Gambacorti-Passerini, A.L. Godfrey, I. Rapado, A. Cvejic, R. Rance, C. McGee, P. Ellis, L.J. Mudie, P.J. Stephens, S. McLaren, C.E. Massie, P.S. Tarpey, I. Varela, S. Nik-Zainal, H.R. Davies, A. Shlien, D. Jones, K. Raine, J. Hinton, A.P. Butler, J.W. Teague, E.J. Baxter, J. Score, A. Galli, M.G. Della Porta, E. Travaglino, M. Groves, S. Tauro, N.C. Munshi, K.C. Anderson, A. El-Naggar, A. Fischer, V. Mustonen, A.J. Warren, N.C.P. Cross, A.R. Green, P.A. Futreal, M.R. Stratton, P.J. Campbell, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. 2011. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* 365:1384–1395. doi:10.1056/NEJMoa1103283.

Papaemmanuil, E., M. Gerstung, L. Malcovati, S. Tauro, G. Gundem, P. Van Loo, C.J. Yoon, P. Ellis, D.C. Wedge, A. Pellagatti, A. Shlien, M.J. Groves, S.A. Forbes, K. Raine, J. Hinton, L.J. Mudie, S. McLaren, C. Hardy, C. Latimer, M.G. Della Porta, S. O'Meara, I. Ambaglio, A. Galli, A.P. Butler, G. Walldin, J.W. Teague, L. Quek, A. Sternberg, C. Gambacorti-Passerini, N.C.P. Cross, A.R. Green, J. Boultwood, P. Vyas, E. Hellstrom-Lindberg, D. Bowen, M. Cazzola, M.R. Stratton, P.J. Campbell, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. 2013. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 122:3616–27– quiz 3699. doi:10.1182/blood-2013-08-518886.

Park, S.M., J. Ou, L. Chamberlain, T.M. Simone, H. Yang, C.-M. Virbasius, A.M. Ali, L.J. Zhu, S. Mukherjee, A. Raza, and M.R. Green. 2016. U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3. *Mol. Cell*. 1–34. doi:10.1016/j.molcel.2016.04.011.

Przychodzen, B., A. Jerez, K. Guinta, M.A. Sekeres, R. Padgett, J.P. Maciejewski, and H. Makishima. 2013. Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood*. 122:999–1006. doi:10.1182/blood-2013-01-480970.

Ruzankina, Y., C. Pinzon-Guzman, A. Asare, T. Ong, L. Pontano, G. Cotsarelis, V.P. Zediak, M. Velez, A. Bhandoola, and E.J. Brown. 2007. Deletion of the

Developmentally Essential Gene ATR in Adult Mice Leads to Age-Related Phenotypes and Stem Cell Loss. *Cell Stem Cell*. 1:113–126. doi:10.1016/j.stem.2007.03.002.

Schaefer, B.C., M.L. Schaefer, J.W. Kappler, P. Marrack, and R.M. Kedl. 2001. Observation of Antigen-Dependent CD8+ T-Cell/ Dendritic Cell Interactions in Vivo. *Cellular Immunology*. 214:110–122. doi:10.1006/cimm.2001.1895.

Shirai, C.L., J.N. Ley, B.S. White, S. Kim, J. Tibbitts, J. Shao, M. Ndonwi, B. Wadugu, E.J. Duncavage, T. Okeyo-Owuor, T. Liu, M. Griffith, S. McGrath, V. Magrini, R.S. Fulton, C. Fronick, M. O'Laughlin, T.A. Graubert, and M.J. Walter. 2015. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*. 27:631–643. doi:10.1016/j.ccell.2015.04.008.

Steensma, D.P., R. Bejar, S. Jaiswal, R.C. Lindsley, M.A. Sekeres, R.P. Hasserjian, and B.L. Ebert. 2015. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 126:9–16. doi:10.1182/blood-2015-03-631747.

The Cancer Genome Atlas Research Network. 2013. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* doi:10.1056/NEJMoa1301689.

Thol, F., S. Kade, C. Schlarmann, P. Löffeld, M. Morgan, J. Krauter, M.W. Wlodarski, B. Kölking, M. Wichmann, K. Görlich, G. Göhring, G. Bug, O. Ottmann, C.M. Niemeyer, W.-K. Hofmann, B. Schlegelberger, A. Ganser, and M. Heuser. 2012. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 119:3578–3584. doi:10.1182/blood-2011-12-399337.

Trapnell, C., L. Pachter, and S.L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 25:1105–1111. doi:10.1093/bioinformatics/btp120.

Troy, J.D., E. Atallah, J.T. Geyer, and W. Saber. 2014. Myelodysplastic syndromes in the United States: an update for clinicians. *Ann. Med.* 46:283–289. doi:10.3109/07853890.2014.898863.

Wagenmakers, E.-J., T. Lodewyckx, H. Kuriyal, and R. Grasman. 2010. Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. *Cogn Psychol*. 60:158–189. doi:10.1016/j.cogpsych.2009.12.001.

Wang, L., M.S. Lawrence, Y. Wan, P. Stojanov, C. Sougnez, K. Stevenson, L. Werner, A. Sivachenko, D.S. DeLuca, L. Zhang, W. Zhang, A.R. Vartanov, S.M. Fernandes, N.R. Goldstein, E.G. Folco, K. Cibulskis, B. Tesar, Q.L. Sievers, E. Shefler, S. Gabriel, N. Hacohen, R. Reed, M. Meyerson, T.R. Golub, E.S. Lander, D. Neuberg, J.R. Brown, G. Getz, and C.J. Wu. 2011. SF3B1and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* 365:2497–2506. doi:10.1056/NEJMoa1109016.

Waterfall, J.J., E. Arons, R.L. Walker, M. Pineda, L. Roth, J.K. Killian, O.D. Abaan, S.R. Davis, R.J. Kreitman, and P.S. Meltzer. 2014. High prevalence of MAP2K1 mutations in variant and IGHV4-34-expressing hairy-cell leukemias. *Nature Genetics*. 46:8–10. doi:10.1038/ng.2828.

Wu, S., C.M. Romfo, T.W. Nilsen, and M.R. Green. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*. 402:832–835. doi:10.1038/45590.

Xu, L., Z.H. Gu, Y. Li, J.L. Zhang, C.K. Chang, C.M. Pan, J.Y. Shi, Y. Shen, B. Chen, Y.Y. Wang, L. Jiang, J. Lu, X. Xu, J.L. Tan, Y. Chen, S.Y. Wang, X. Li, Z. Chen, and S.J. Chen. 2014. Genomic landscape of CD34+ hematopoietic cells in myelodysplastic syndrome and gene mutation profiles as prognostic markers. *Proc Natl Acad Sci U S A*. 111:8589–8594. doi:10.1073/pnas.1407688111.

Yan, X.-J., J. Xu, Z.-H. Gu, C.-M. Pan, G. Lu, Y. Shen, J.-Y. Shi, Y.-M. Zhu, L. Tang, X.-W. Zhang, W.-X. Liang, J.-Q. Mi, H.-D. Song, K.-Q. Li, Z. Chen, and S.-J. Chen. 2011. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature Genetics*. 43:309–315. doi:10.1038/ng.788.

Yip, B.H., V. Steeples, E. Repapi, R.N. Armstrong, M. Llorian, S. Roy, J. Shaw, H. Dolatshad, S. Taylor, A. Verma, M. Bartenstein, P. Vyas, N.C.P. Cross, L. Malcovati, M. Cazzola, E. Hellstrom-Lindberg, S. Ogawa, C.W.J. Smith, A. Pellagatti, and J. Boultwood. 2017. The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. *J. Clin. Invest.* doi:10.1172/JCI91363DS9.

Yoshida, K., M. Sanada, Y. Shiraishi, D. Nowak, Y. Nagata, R. Yamamoto, Y. Sato, A. Sato-Otsubo, A. Kon, M. Nagasaki, G. Chalkidis, Y. Suzuki, M. Shiosaka, R. Kawahata, T. Yamaguchi, M. Otsu, N. Obara, M. Sakata-Yanagimoto, K. Ishiyama, H. Mori, F. Nolte, W.-K. Hofmann, S. Miyawaki, S. Sugano, C. Haferlach, H.P. Koeffler, L.-Y. Shih, T. Haferlach, S. Chiba, H. Nakauchi, S. Miyano, and S. Ogawa. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 478:64–69. doi:10.1038/nature10496.

Zhen, T., E.M. Kwon, L. Zhao, J. Hsu, R.K. Hyde, Y. Lu, L. Alemu, N.A. Speck, and P.P. Liu. 2017. Chd7 deficiency delays leukemogenesis in mice induced by Cbfb-MYH11. *Blood*. 130:2431–2442. doi:10.1182/blood-2017-04-780106.

Zorio, D.A., and T. Blumenthal. 1999. Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature*. 402:835–838. doi:10.1038/45597.

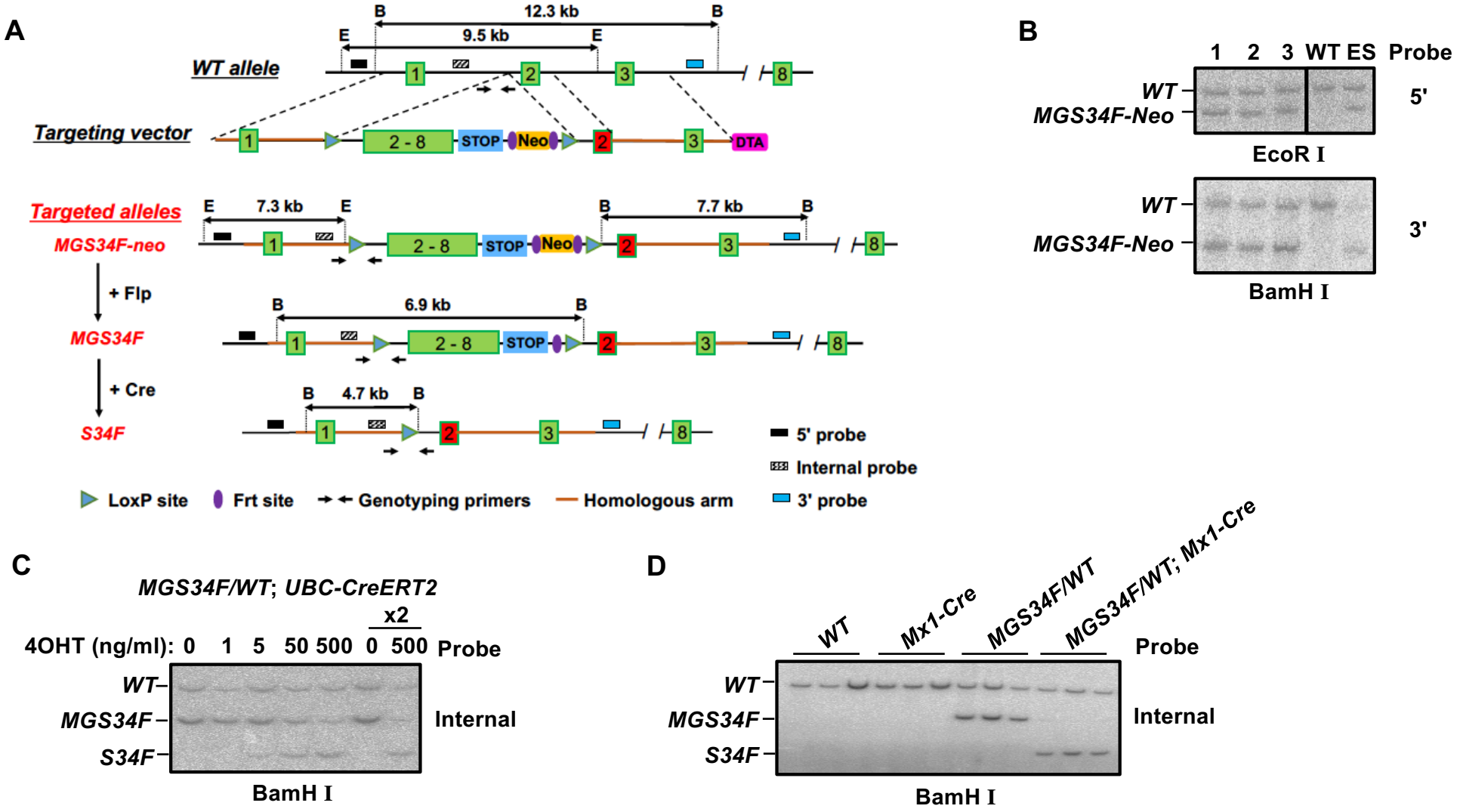

A
B
12.3 kb
9.5 kb
WT allele
Targeting vector
2 - 8
STOP
Neo
DTA
Targeted alleles
MGS34F-neo
7.3 kb
7.7 kb
+ Flp
MGS34F
6.9 kb
+ Cre
S34F
4.7 kb
5' probe
Internal probe
3' probe
LoxP site
Frt site
Genotyping primers
Homologous arm
1 2 3 WT ES Probe
WT
MGS34F-Neo
5'
EcoR I
3'
BamH I
C
MGS34F/WT; UBC-CreERT2
x2
4OHT (ng/ml): 0 1 5 50 500 0 500
Probe
WT
MGS34F
S34F
Internal
BamH I
D
WT
Mx1-Cre
MGS34F/WT
MGS34F/WT; Mx1-Cre
Probe
WT
MGS34F
S34F
Internal
BamH I

Figure 1

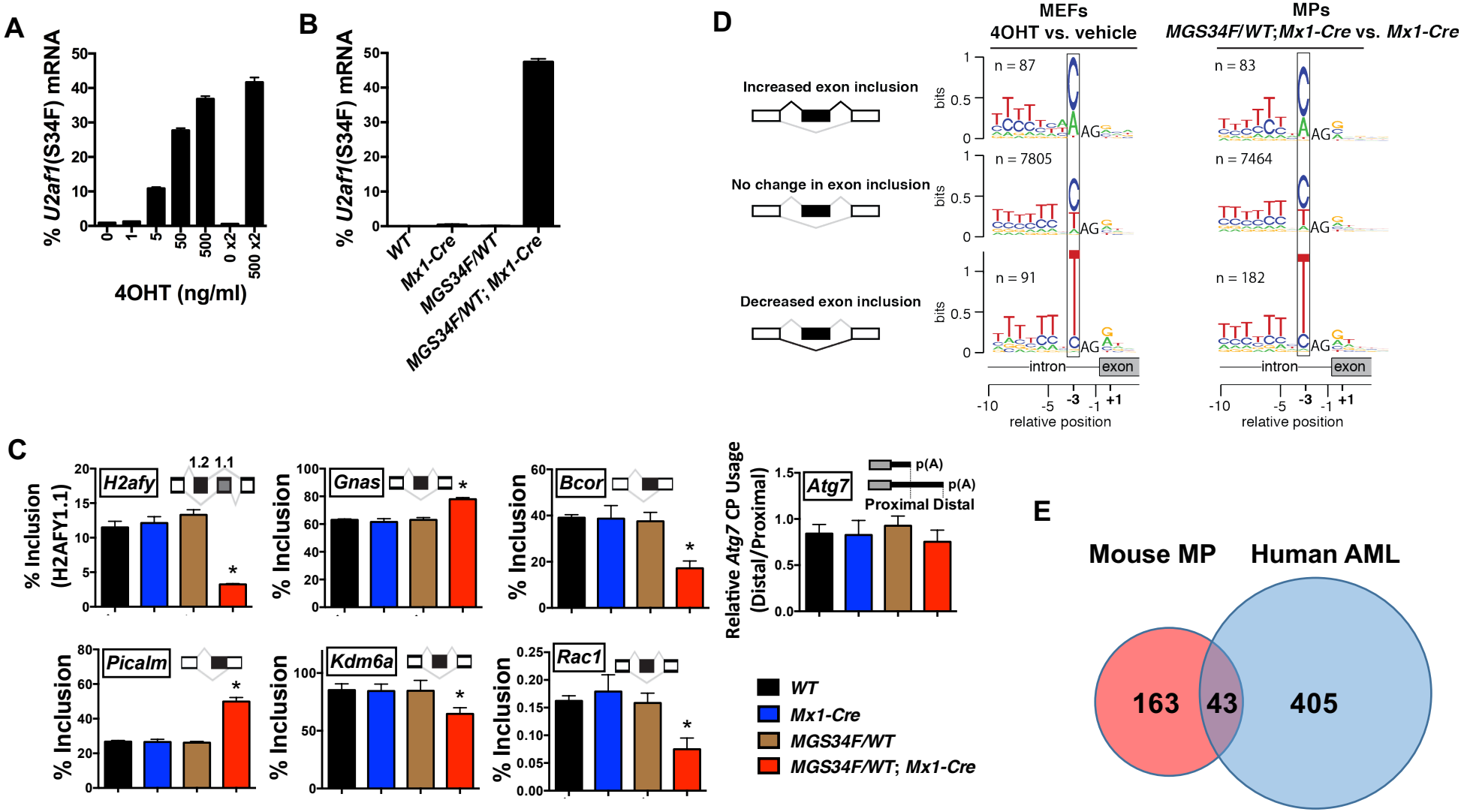

A
% U2af1(S34F) mRNA
4OHT (ng/ml)
0
1
5
50
500
0 x2
500 x2
B
% U2af1(S34F) mRNA
WT
Mx1-Cre
MGS34F/WT
MGS34F/WT; Mx1-Cre
D
MEFs
4OHT vs. vehicle
MPs
MGS34F/WT;Mx1-Cre vs. Mx1-Cre
Increased exon inclusion
No change in exon inclusion
Decreased exon inclusion
n = 87
n = 7805
n = 91
n = 83
n = 7464
n = 182
bits
intron
exon
relative position
C
H2afy
1.2 1.1
Gnas
Bcor
Picalm
Kdm6a
Rac1
% Inclusion
% Inclusion (H2AFY1.1)
Atg7
p(A)
Proximal Distal
Relative Atg7 CP Usage (Distal/Proximal)
WT
Mx1-Cre
MGS34F/WT
MGS34F/WT; Mx1-Cre
E
Mouse MP
Human AML
163
43
405

Figure 2

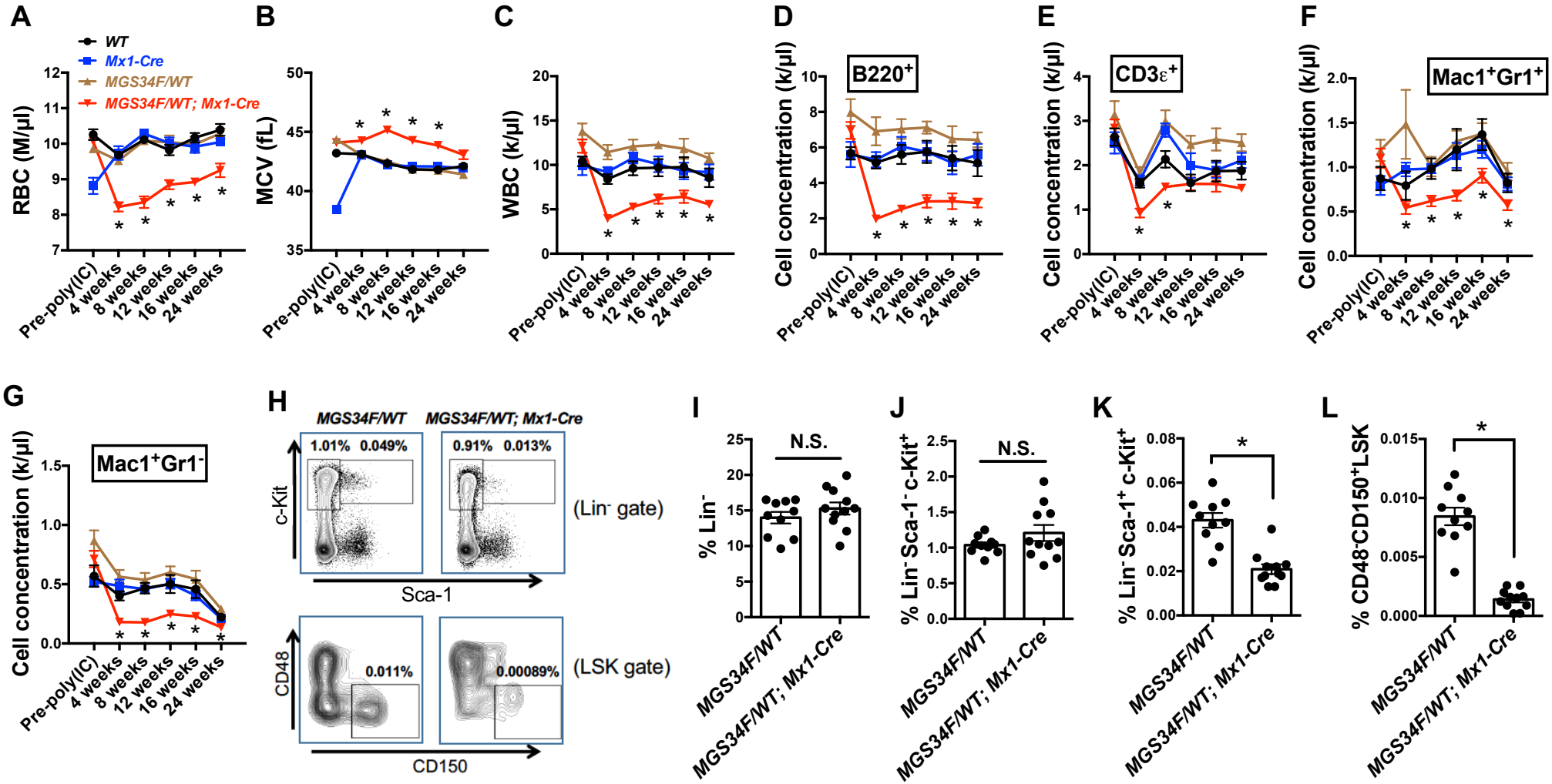

A
B
C
D
E
F
G
H
I
J
K
L
WT
Mx1-Cre
MGS34F/WT
MGS34F/WT; Mx1-Cre
RBC (M/μl)
MCV (fL)
WBC (k/μl)
Cell concentration (k/μl)
B220+
CD3ε+
Mac1+Gr1+
Mac1+Gr1-
Pre-poly(IC)
4 weeks
8 weeks
12 weeks
16 weeks
24 weeks
MGS34F/WT
MGS34F/WT; Mx1-Cre
1.01% 0.049%
0.91% 0.013%
c-Kit
Sca-1
(Lin- gate)
0.011%
0.00089%
CD48
CD150
(LSK gate)
% Lin-
% Lin-Sca-1-c-Kit+
% Lin-Sca-1+c-Kit+
% CD48-CD150+LSK
N.S.
*

Figure 3

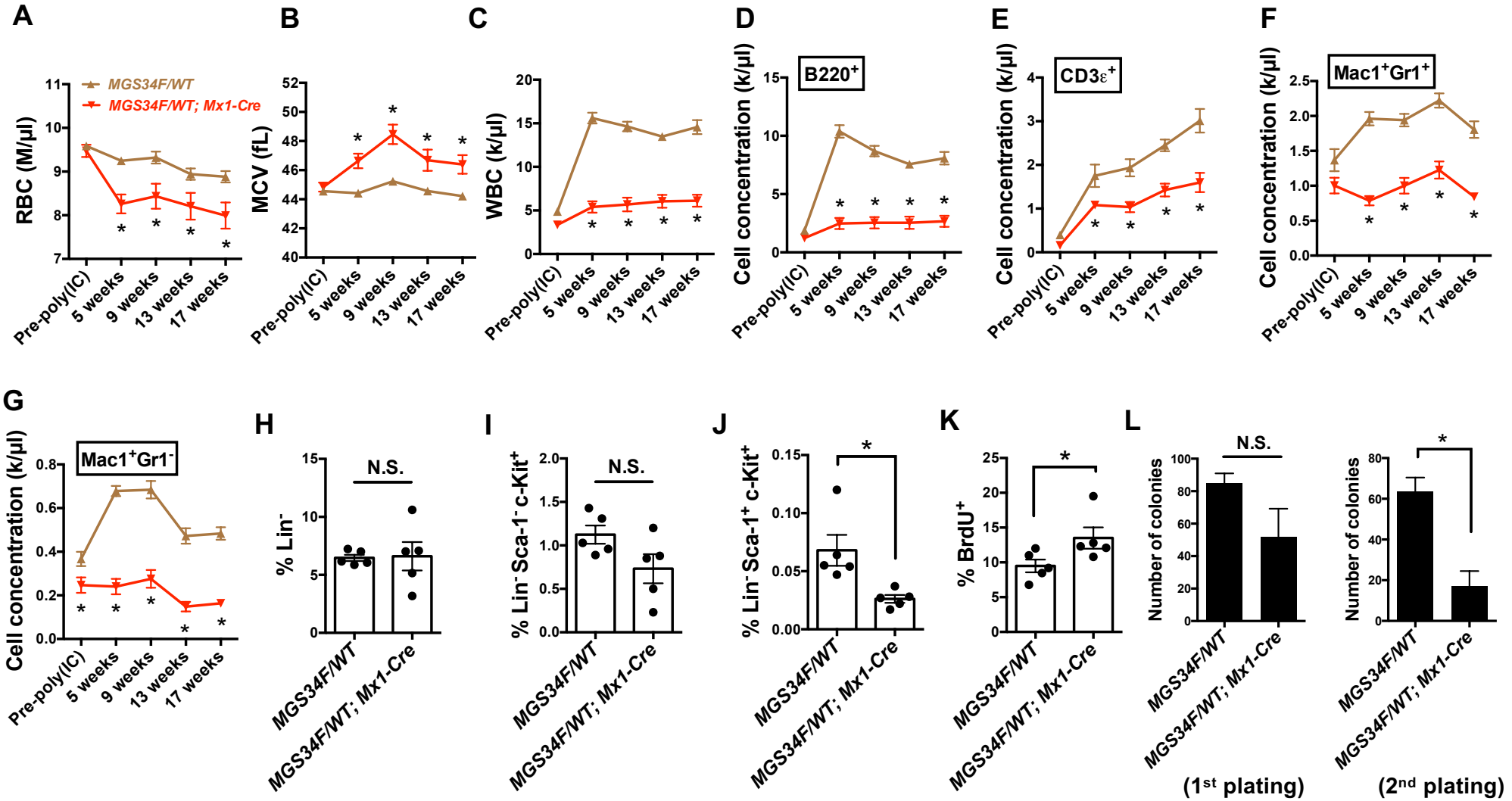

A
B
C
D
E
F
G
H
I
J
K
L
MGS34F/WT
MGS34F/WT; Mx1-Cre
RBC (M/μl)
MCV (fL)
WBC (k/μl)
Cell concentration (k/μl)
B220+
CD3ε+
Mac1+Gr1+
Mac1+Gr1-
Pre-poly(IC)
5 weeks
9 weeks
13 weeks
17 weeks
% Lin-
% Lin-Sca-1-c-Kit+
% Lin-Sca-1+c-Kit+
% BrdU+
N.S.
Number of colonies
(1st plating)
(2nd plating)

Figure 4

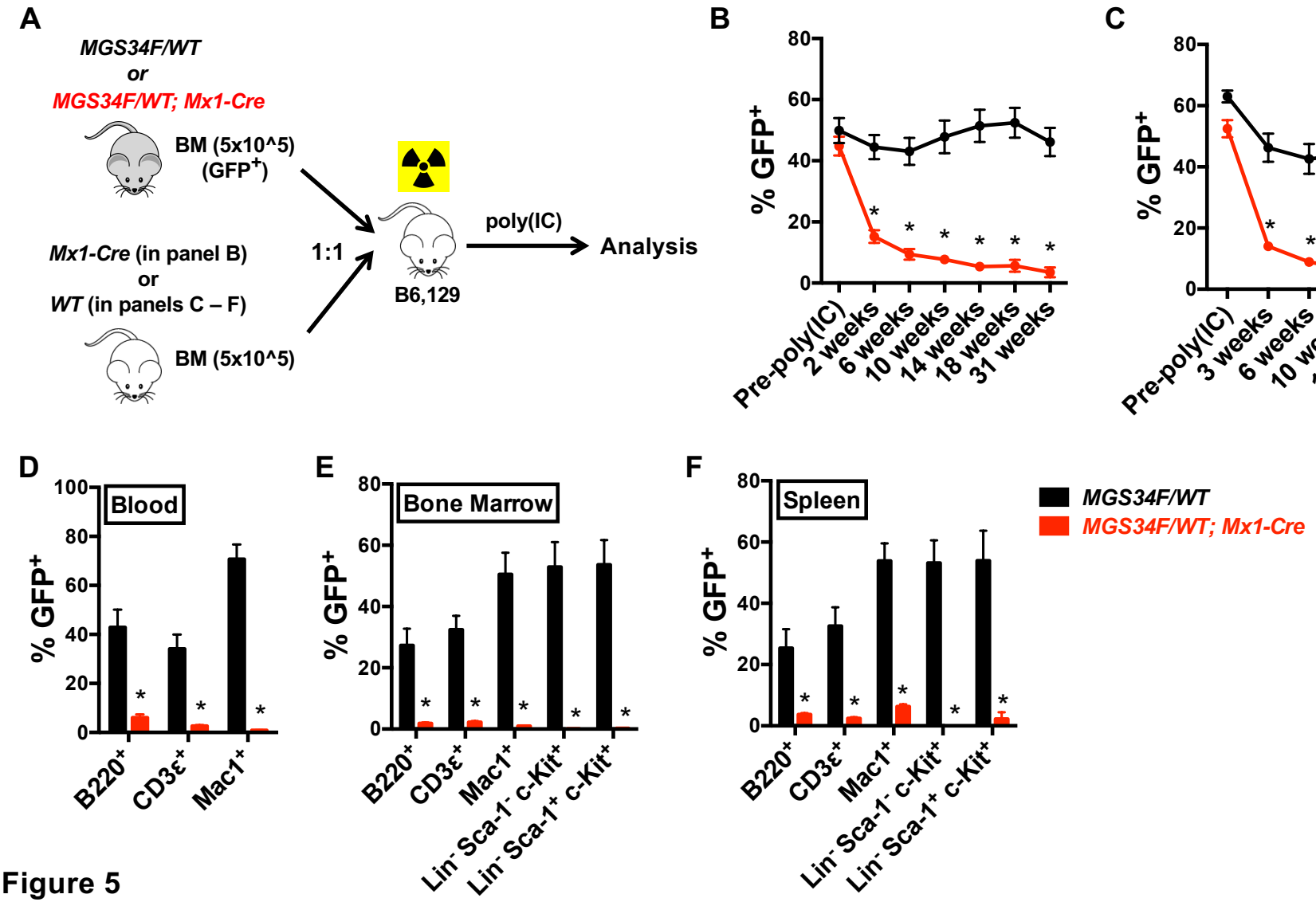

A
MGS34F/WT
or
MGS34F/WT; Mx1-Cre
BM (5x10^5)
(GFP+)
Mx1-Cre (in panel B)
or
WT (in panels C – F)
BM (5x10^5)
1:1
B6,129
poly(IC)
Analysis
B
% GFP+
Pre-poly(IC)
2 weeks
6 weeks
10 weeks
14 weeks
18 weeks
31 weeks
C
% GFP+
Pre-poly(IC)
3 weeks
6 weeks
10 weeks
15 weeks
21 weeks
D
Blood
% GFP+
B220+
CD3ε+
Mac1+
E
Bone Marrow
% GFP+
B220+
CD3ε+
Mac1+
Lin- Sca-1- c-Kit+
Lin- Sca-1+ c-Kit+
F
Spleen
% GFP+
B220+
CD3ε+
Mac1+
Lin- Sca-1- c-Kit+
Lin- Sca-1+ c-Kit+
MGS34F/WT
MGS34F/WT; Mx1-Cre

Figure 5

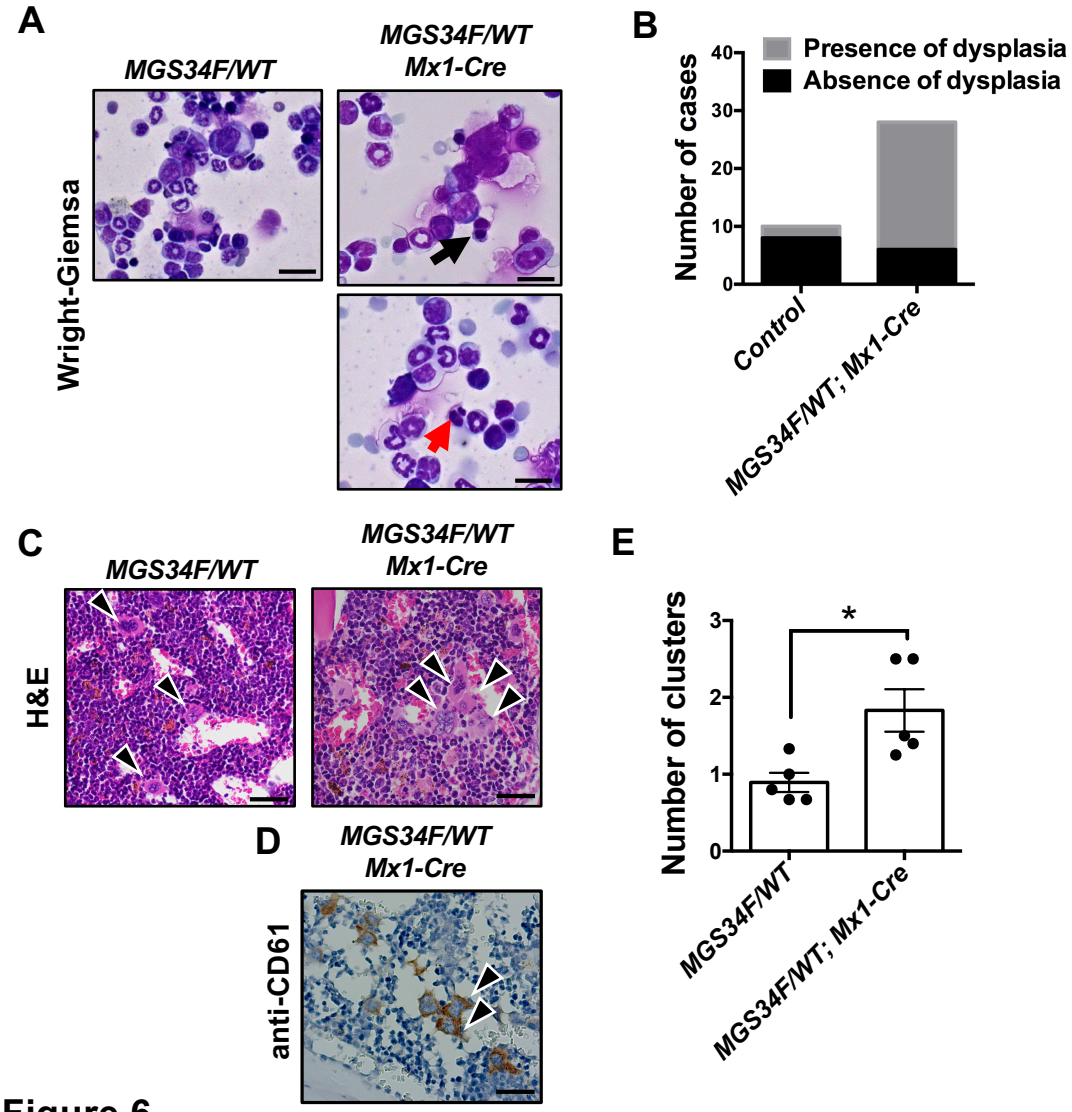

A
MGS34F/WT
MGS34F/WT
Mx1-Cre
Wright-Giemsa
B
Presence of dysplasia
Absence of dysplasia
Number of cases
40
30
20
10
0
Control
MGS34F/WT; Mx1-Cre
C
MGS34F/WT
MGS34F/WT
Mx1-Cre
H&E
D
MGS34F/WT
Mx1-Cre
anti-CD61
E
Number of clusters
3
2
1
0
*
MGS34F/WT
MGS34F/WT; Mx1-Cre

Figure 6

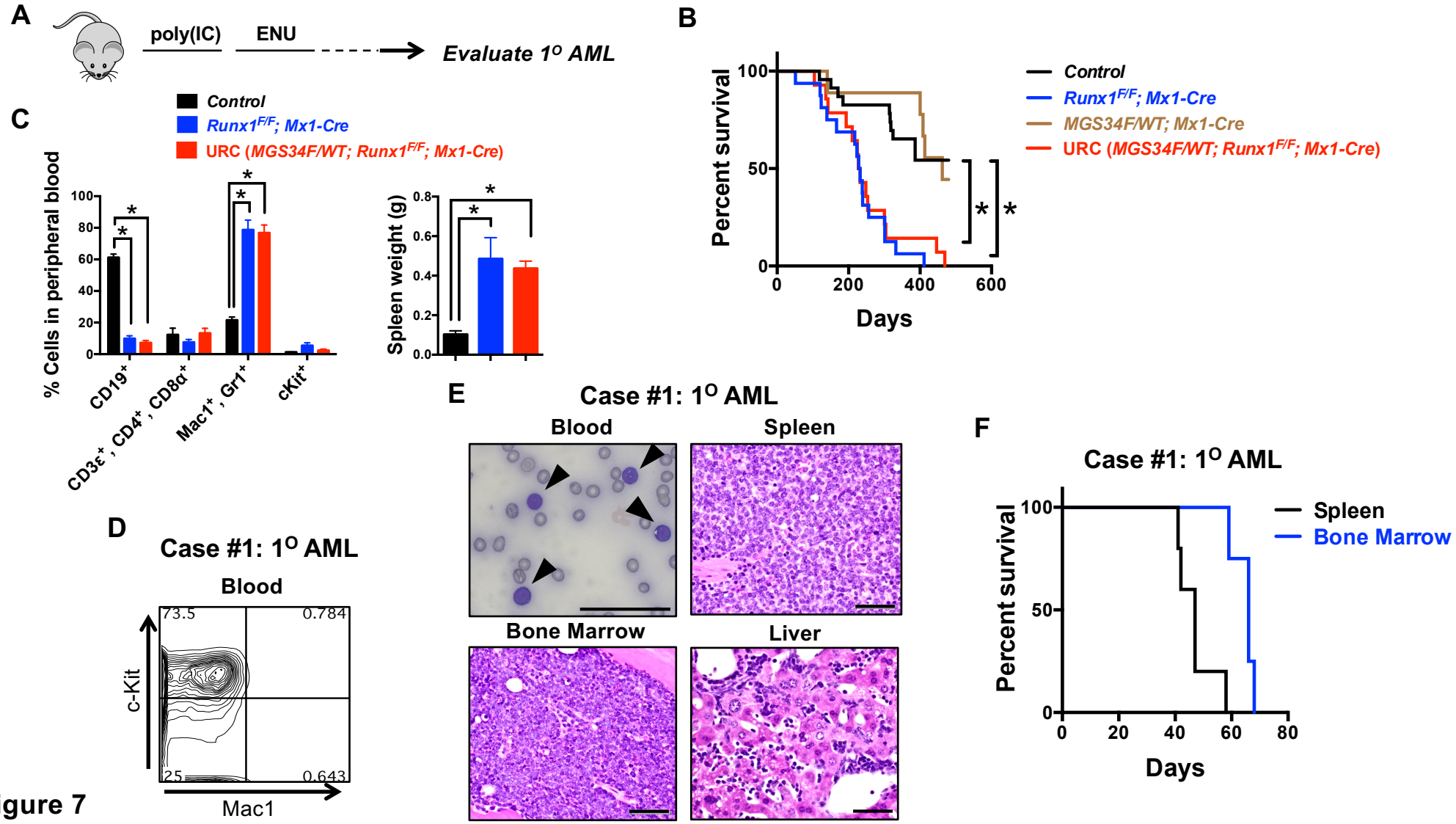
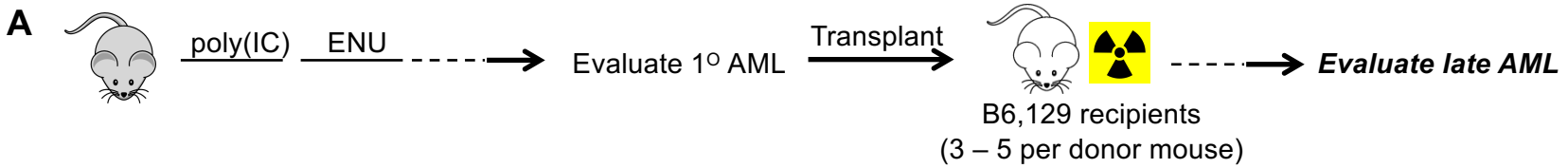
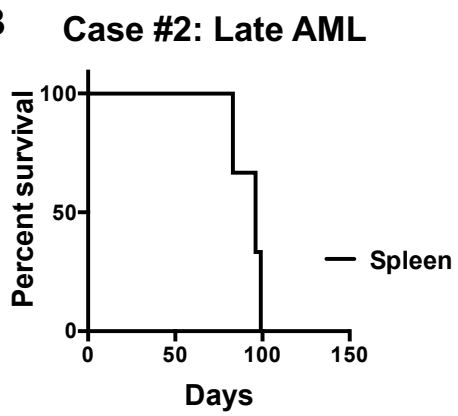
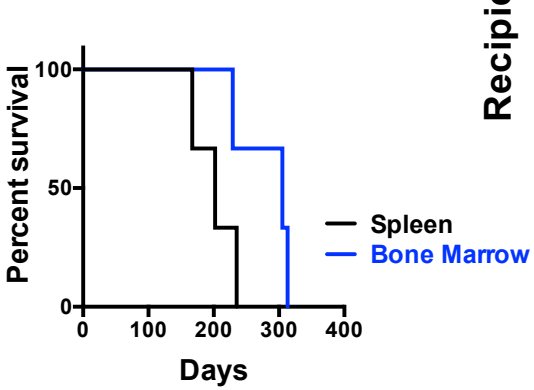

A
poly(IC)
ENU
Evaluate 1° AML
B
Percent survival
Days
Control
Runx1F/F; Mx1-Cre
MGS34F/WT; Mx1-Cre
URC (MGS34F/WT; Runx1F/F; Mx1-Cre)
C
Control
Runx1F/F; Mx1-Cre
URC (MGS34F/WT; Runx1F/F; Mx1-Cre)
% Cells in peripheral blood
CD19+
CD3ε+, CD4+, CD8α+
Mac1+, Gr1+
cKit+
Spleen weight (g)
D
Case #1: 1° AML
Blood
c-Kit
Mac1
73.5
0.784
25
0.643
E
Case #1: 1° AML
Blood
Spleen
Bone Marrow
Liver
F
Case #1: 1° AML
Percent survival
Days
Spleen
Bone Marrow

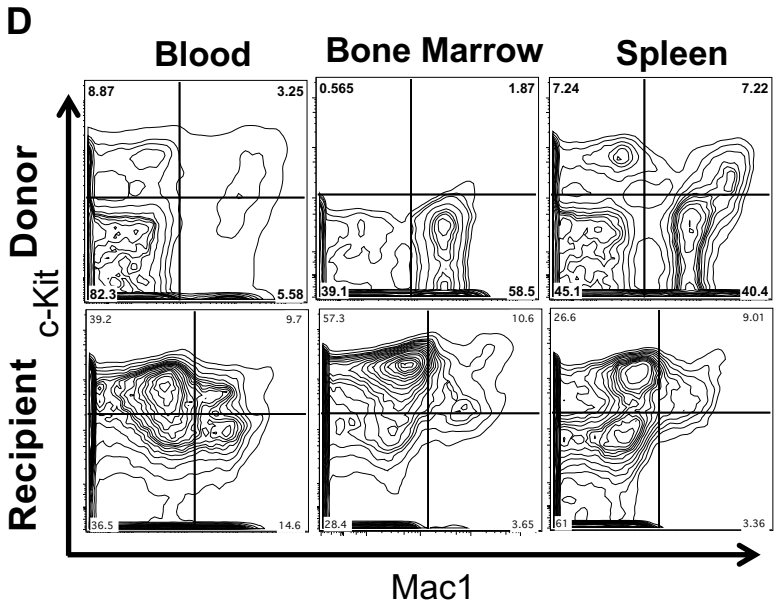
Figure 7





Case #3: Late AML (C – E)





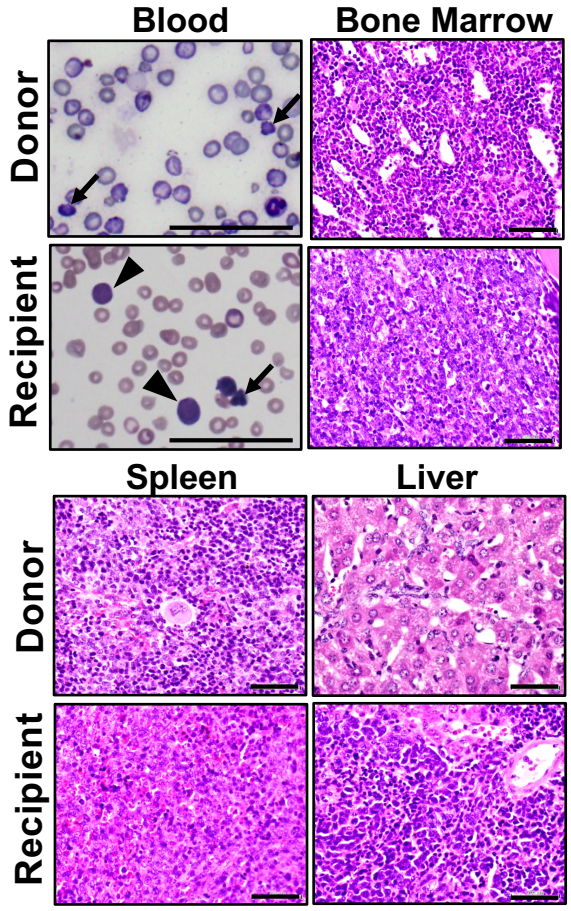


Figure 8

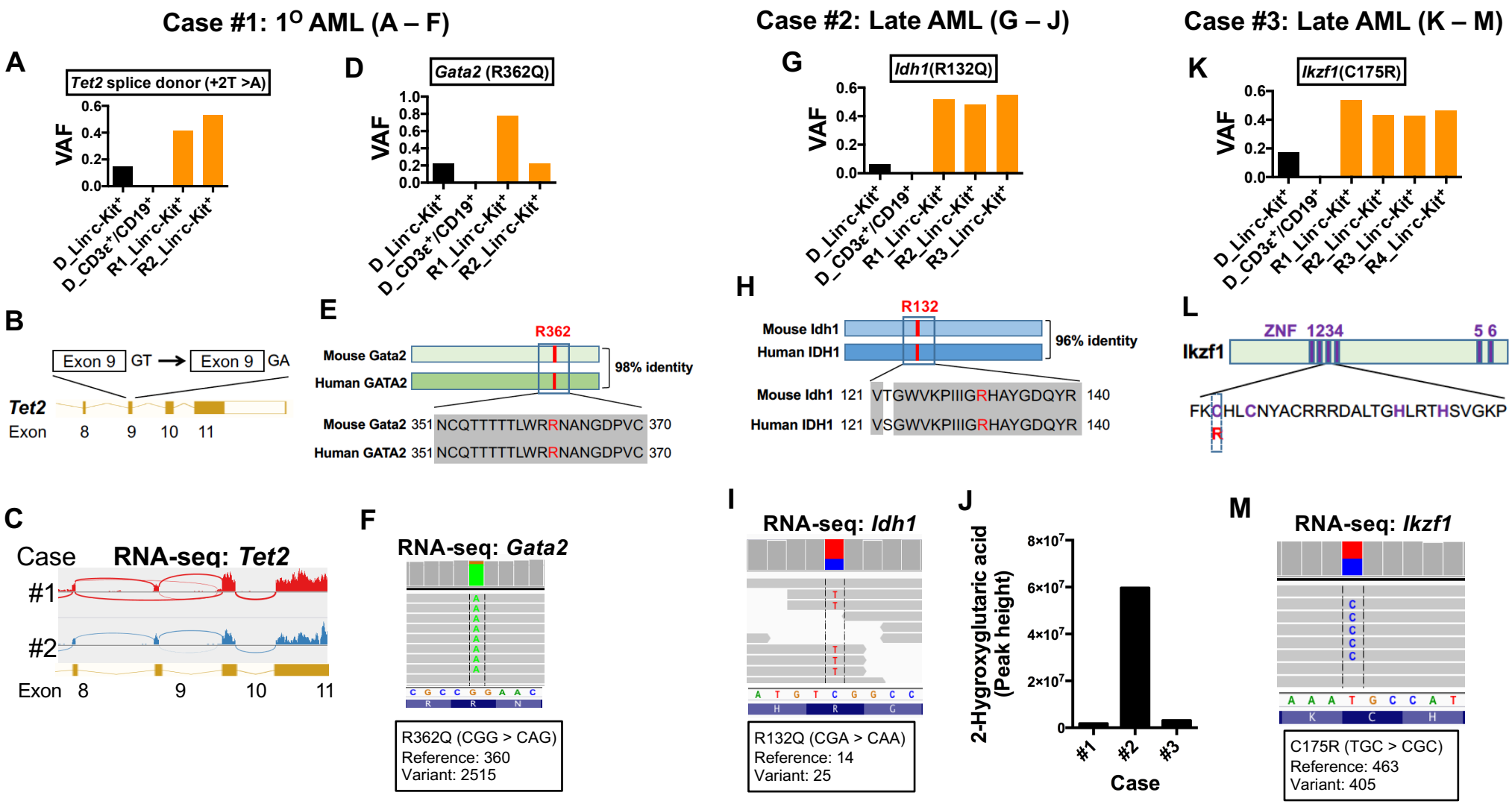

Case #1: 1O AML (A – F)
Case #2: Late AML (G – J)
Case #3: Late AML (K – M)
A
Tet2 splice donor (+2T >A)
VAF
D_Lin-c-Kit+
D_CD3ε+/CD19+
R1_Lin-c-Kit+
R2_Lin-c-Kit+
D
Gata2 (R362Q)
G
Idh1(R132Q)
R3_Lin-c-Kit+
K
Ikzf1(C175R)
R4_Lin-c-Kit+
B
Exon 9
GT
GA
Tet2
Exon 8 9 10 11
E
R362
Mouse Gata2
Human GATA2
98% identity
Mouse Gata2 351 NCQTTTTTLWRRNANGDPVC 370
Human GATA2 351 NCQTTTTTLWRRNANGDPVC 370
H
R132
Mouse Idh1
Human IDH1
96% identity
Mouse Idh1 121 VTGWVKPIIIGRHAYGDQYR 140
Human IDH1 121 VSGWVKPIIIGRHAYGDQYR 140
L
ZNF 1234
5 6
Ikzf1
FKCHLCNYACRRRDALTGHLRTHSVGKP
R
C
Case
RNA-seq: Tet2
#1
#2
Exon 8 9 10 11
F
RNA-seq: Gata2
R362Q (CGG > CAG)
Reference: 360
Variant: 2515
I
RNA-seq: Idh1
R132Q (CGA > CAA)
Reference: 14
Variant: 25
J
2-Hygroxyglutaric acid
(Peak height)
8×10^7
6×10^7
4×10^7
2×10^7
0
#1
#2
#3
Case
M
RNA-seq: Ikzf1
C175R (TGC > CGC)
Reference: 463
Variant: 405

Figure 9

Table 1. Summary of AML incidences in mice with ENU and poly (IC) treatment.

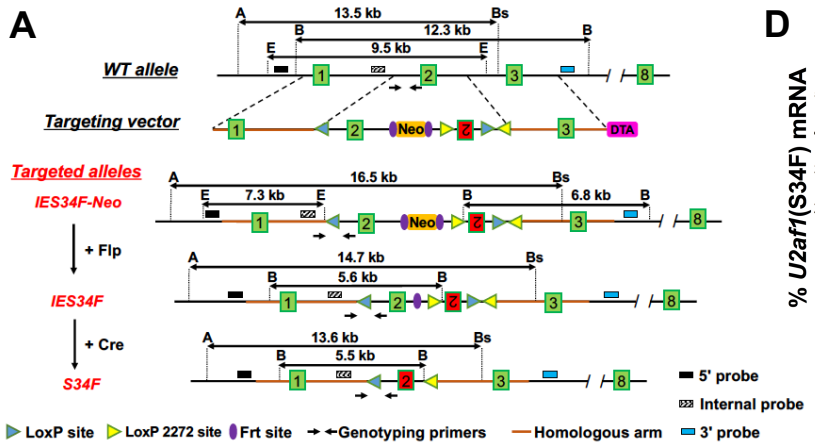
Genotype	Cohort size	Mice with AML	1° AML	Late AML	Censorship^
Control[&]	10	0	0	NA[$]	3
MGS34F/WT; Mx1Cre	9	0	0	0	1
Runx1$^{F/F}$*; Mx1Cre*	14	0	0	0	0
URC (MGS34F/WT; Runx1$^{F/F}$*; Mx1Cre)*	16	3[#]	1	2	0

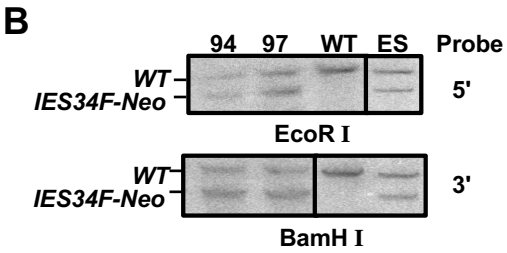
^ The number of mice that were excluded from the study because they developed either T cell lymphoma or thymoma.

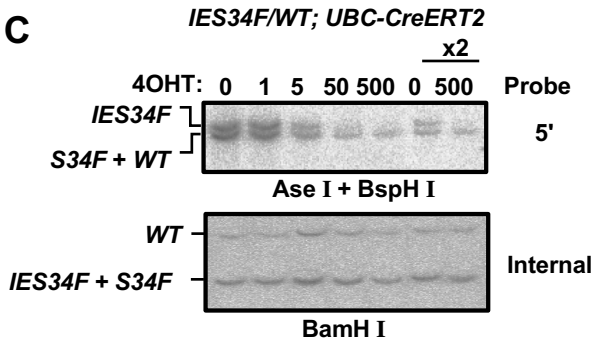
& The "control" genotype includes mice with the following genotypes: *WT/WT* (n = 3). *MGS34F/WT* (n = 2). *Runx1*$^{F/WT}$ (n = 2). *MGS34F/WT*; *Runx1*$^{F/WT}$ (n = 1). *MGS34F/WT*; *Runx1*$^{F/F}$ (n = 1). *Runx1*$^{F/F}$ (n = 1).

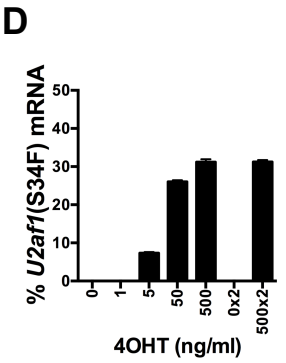
$ Not available. No mouse with the "control" genotype was not used for transplant.

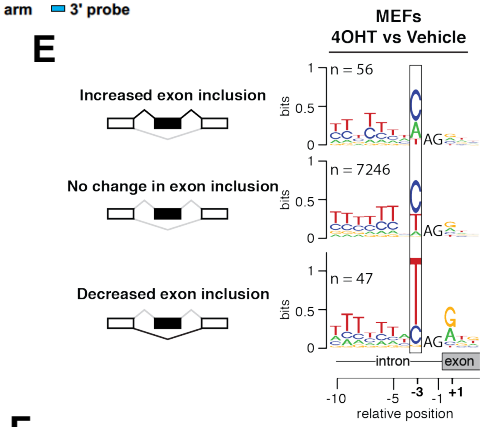
p-value = 0.13 (four separate genotypes as shown) or 0.03 (URC vs. others), by Fisher's exact test.

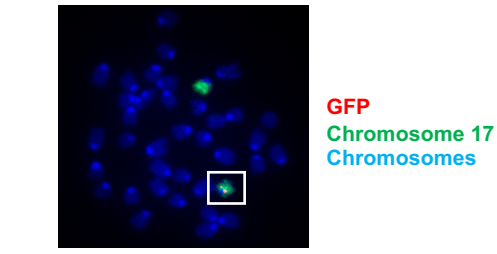


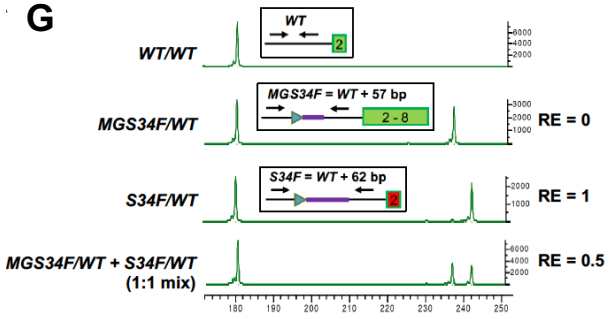


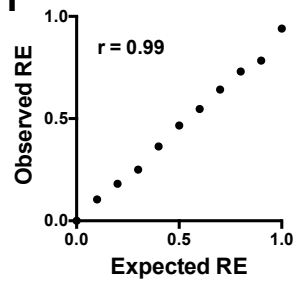


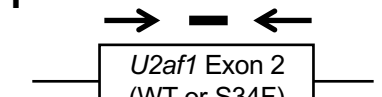


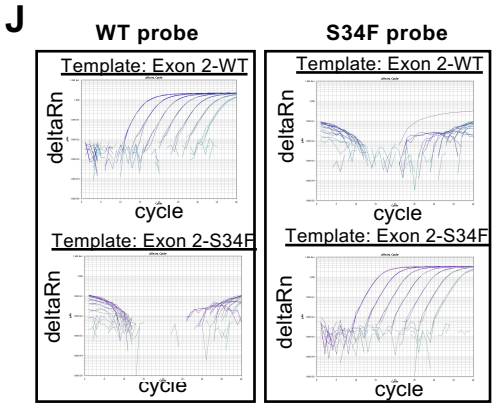


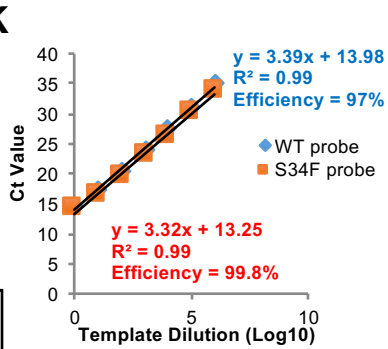




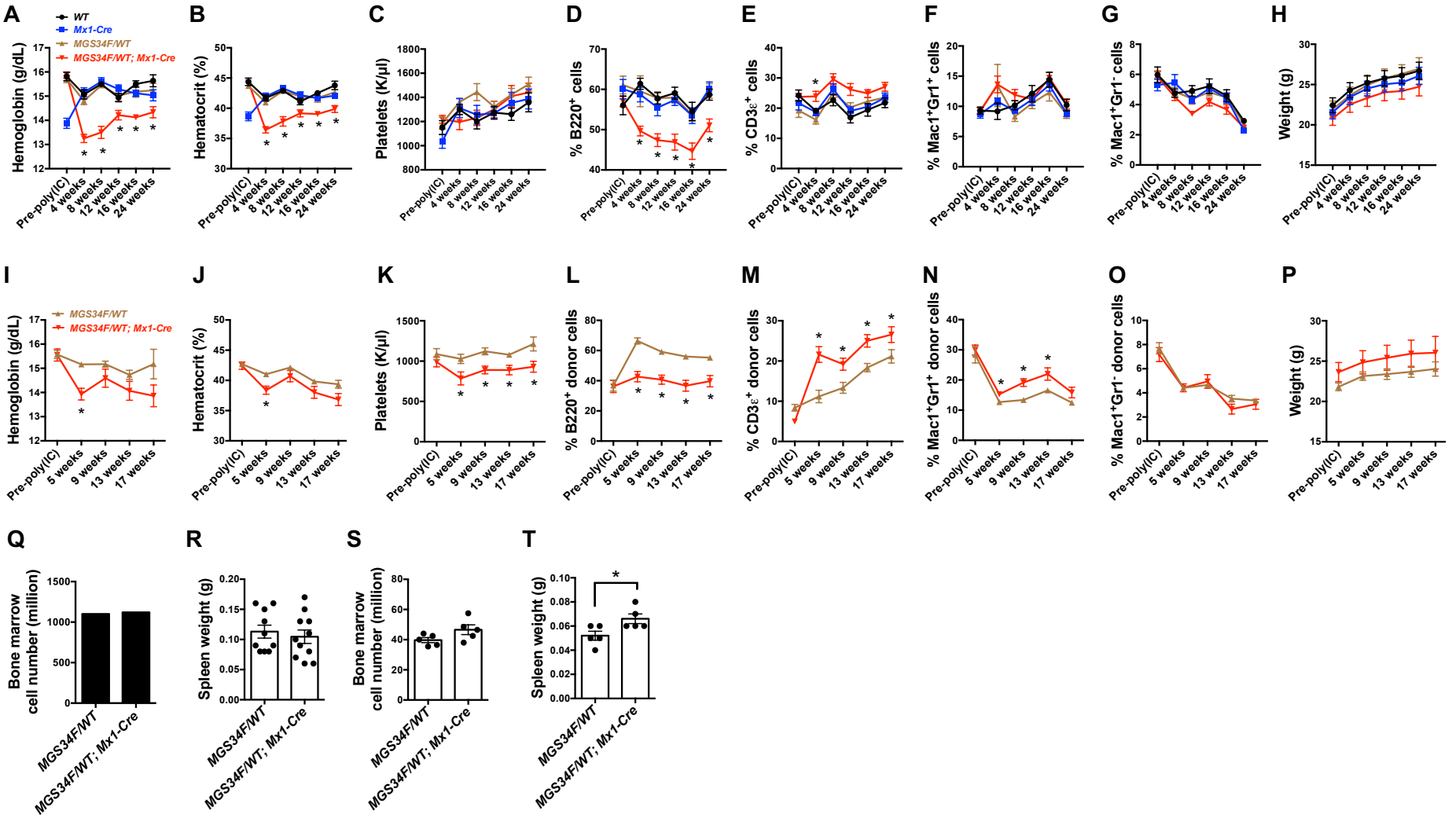




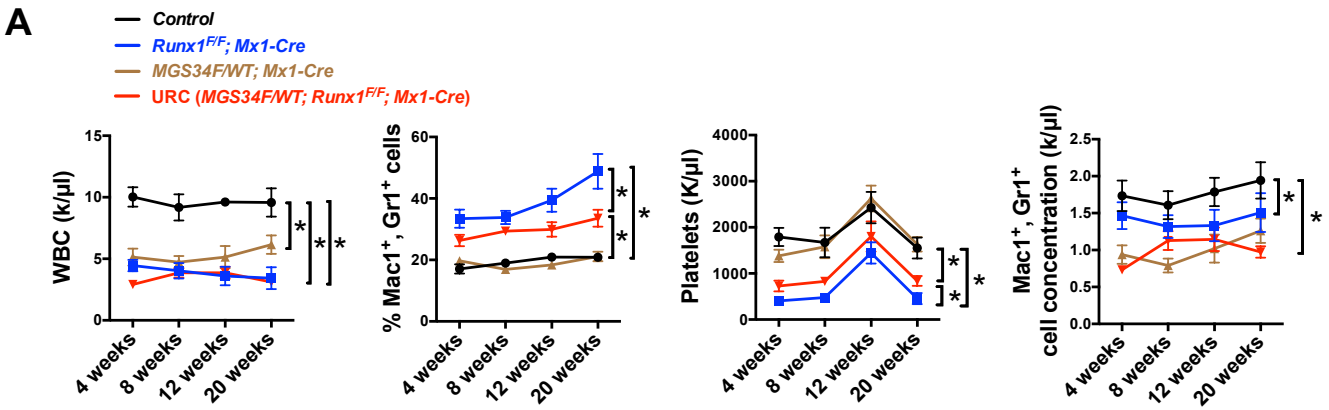


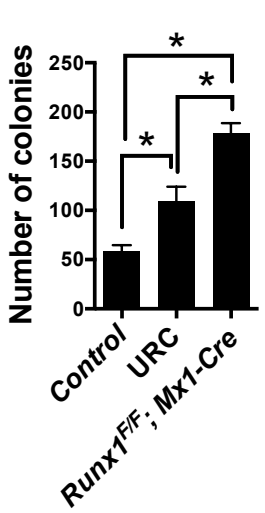


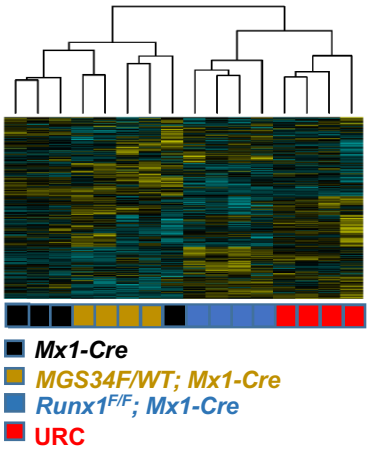
Supplemental Figure S1

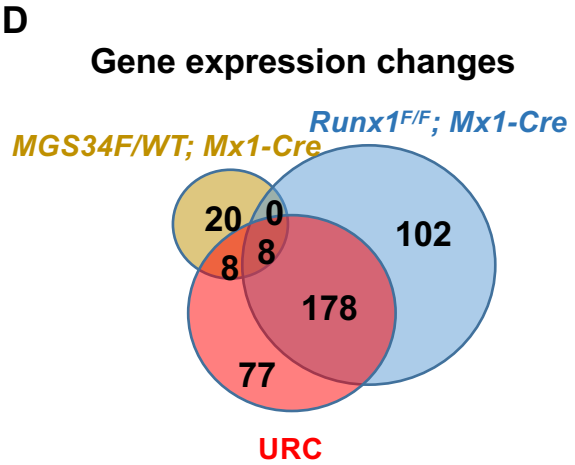


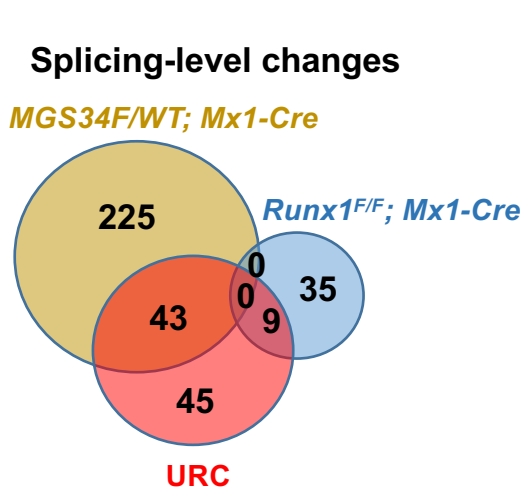
Supplemental Figure S2



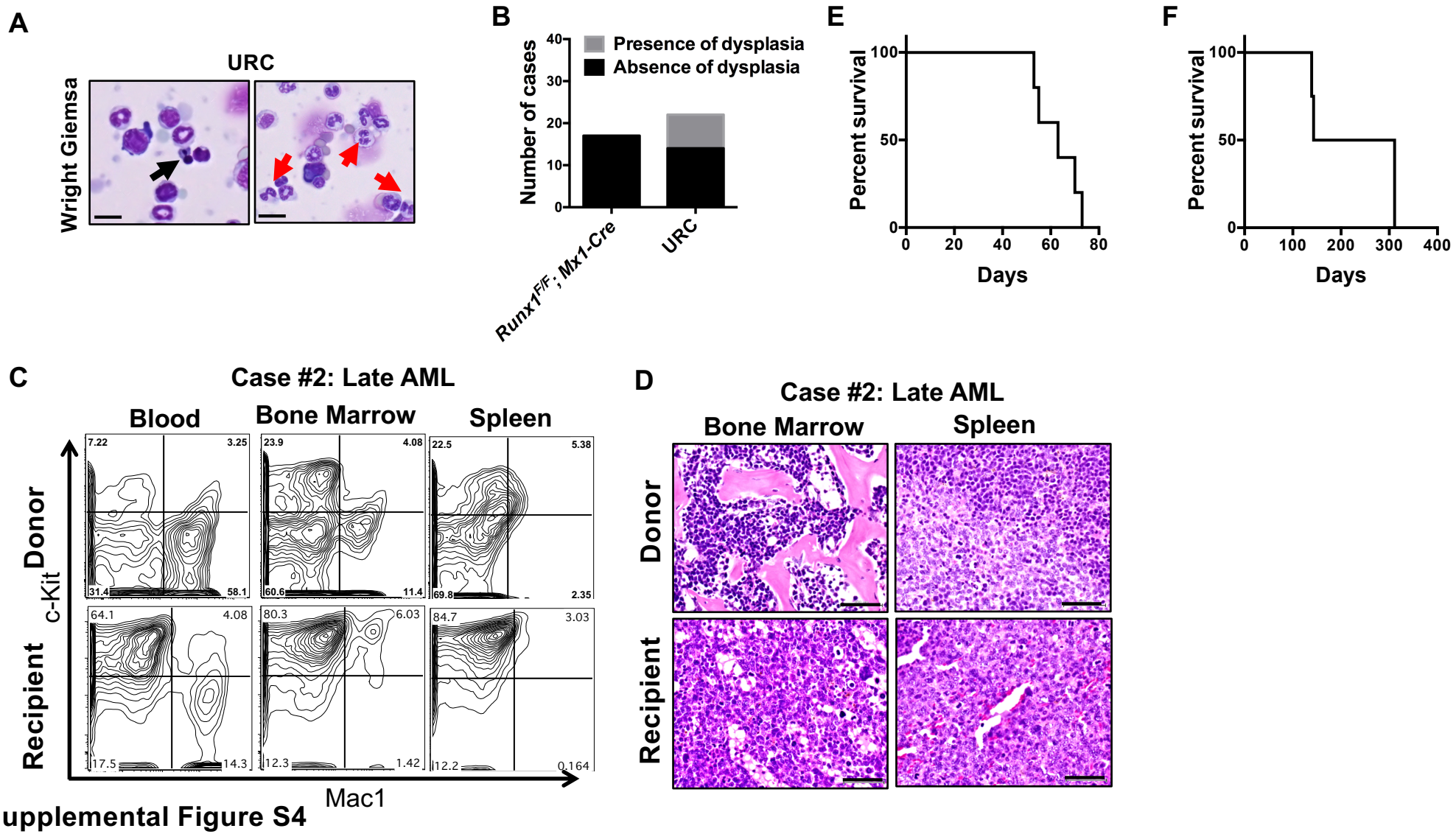




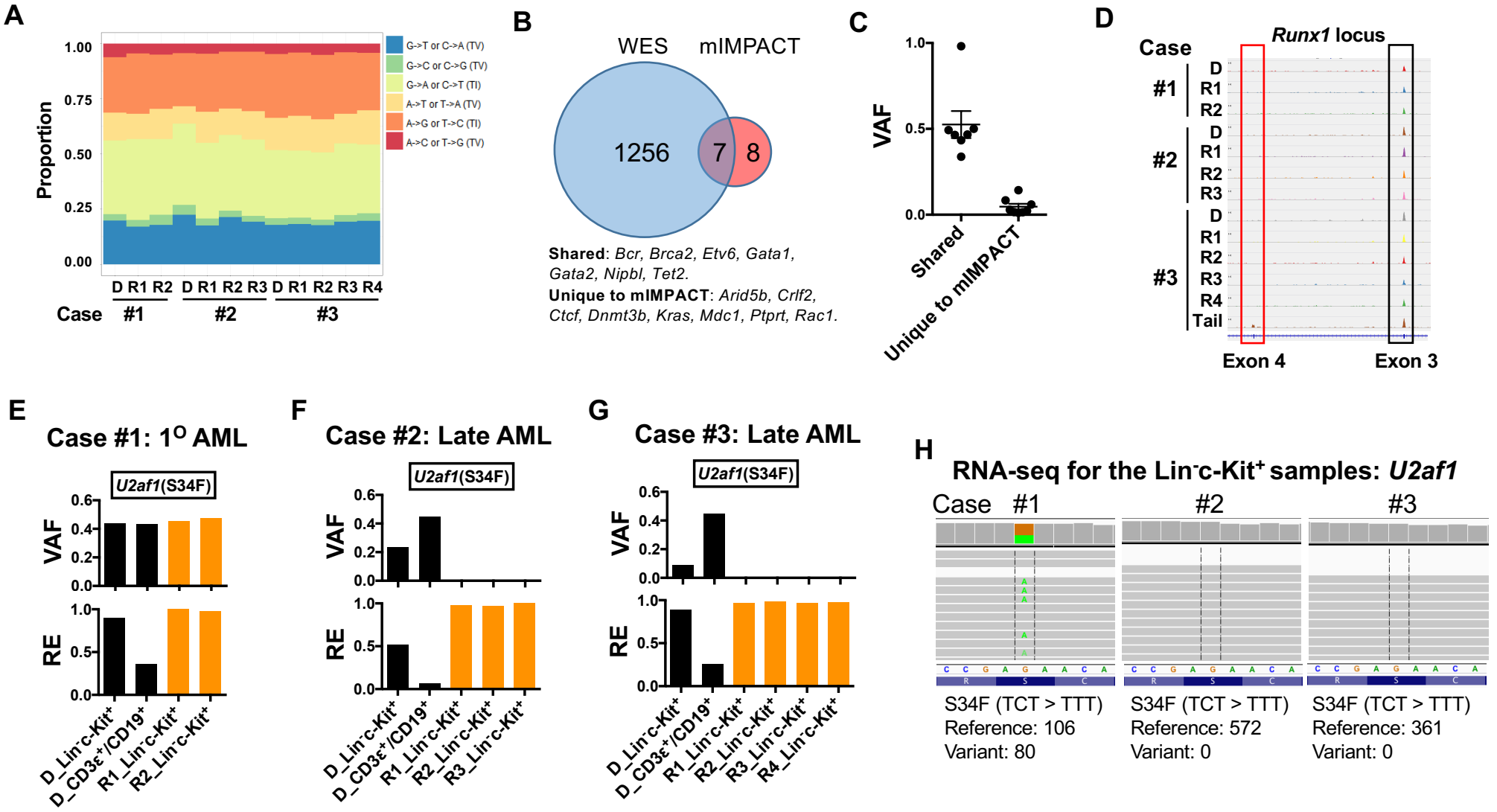




Supplemental Figure S3



Supplemental Figure S4



Supplemental Figure S5