

# *U2af1*<sup>S34F</sup> and *U2af1*<sup>Q157R</sup> myeloid neoplasm-associated hotspot mutations induce distinct hematopoietic phenotypes in mice

**Michael O. Alberti**

Department of Pathology and Immunology, Washington University, St. Louis, MO; Department of Pathology, University of Colorado Anschutz Medical Campus, Aurora, CO

**Sridhar Nonavinkere Srivatsan**

Department of Medicine, Washington University, St. Louis, MO <https://orcid.org/0000-0001-6293-4632>

**Jin Shao**

Department of Medicine, Washington University, St. Louis, MO

**Dennis L. Fei**

Department of Medicine, Meyer Cancer Center, Weill Cornell Medicine, New York, NY; Cancer Biology Section, Cancer Genetics Branch, National Human Genome Research Institute, Bethesda, MD

**Mengou Zhu**

Department of Medicine, Washington University, St. Louis, MO

**Clauida Cabrera Pastrana**

Department of Medicine, Washington University, St. Louis, MO

**Sarah Grieb**

Department of Medicine, Washington University, St. Louis, MO

**Timothy A. Graubert**

Massachusetts General Hospital Cancer Center, Harvard Medical School, Charlestown, MA

**Omar Abdel-Wahab**

Memorial Sloan Kettering Cancer Center, New York, NY

**Matthew J. Walter**

[mjwalter@wustl.edu](mailto:mjwalter@wustl.edu)

Department of Medicine, Washington University, St. Louis, MO <https://orcid.org/0000-0002-7753-1091>


---

Article

Keywords:

Posted Date: May 7th, 2025

**DOI:** <https://doi.org/10.21203/rs.3.rs-6377810/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License](#)

**Additional Declarations:** Yes there is potential conflict of interest.

---

# Abstract

Recurrent somatic mutations in the spliceosome genes *SF3B1*, *SRSF2*, and *U2AF1* are frequently identified in patients with myeloid neoplasms, such as myelodysplastic syndromes. We characterized the *in vivo* consequences of expressing two hotspot mutations in *U2AF1* that code for the S34F and Q157R substitutions. Our results indicate that the two mutations induce distinct hematopoietic phenotypes in mice, suggesting that the *U2AF1*<sup>S34F</sup> and *U2AF1*<sup>Q157R</sup> mutations should not be conflated as they may impact disease pathogenesis differently in patients. Mice expressing *U2af1*<sup>S34F</sup> have a more severe reduction in their blood and bone marrow cell counts and reduced stem cell repopulating ability, compared to mice expressing *U2af1*<sup>Q157R</sup>. The expression and splicing of target genes are largely unique between the mutations, in both mouse and human samples, potentially driving the phenotypic differences induced by either mutation. The two mutations co-occur with different gene mutations in patients and are not equally represented across myeloid neoplasms, suggesting that multiple mechanisms likely drive *U2AF1*-mutant disease pathogenesis. Collectively, our results support that *U2AF1*<sup>S34F</sup> and *U2AF1*<sup>Q157R</sup> mutations induce distinct hematopoietic, gene expression, and RNA splicing phenotypes *in vivo*. Larger population studies will be needed to determine if these phenotypic changes translate into clinico-pathologic differences in patients warranting separate classification.

## INTRODUCTION

Recurrent somatic mutations in a subset of spliceosome genes (*SF3B1*, *SRSF2*, and *U2AF1*) are frequently identified (30–60% depending on disease phenotype) in patients afflicted with myelodysplastic syndromes (MDS), myeloproliferative neoplasms (MPN) such as myelofibrosis (MF), MDS/MPN overlap disorders such as chronic myelomonocytic leukemia (CMML), and secondary acute myeloid leukemia (sAML).<sup>1–8</sup> These heterozygous and mutually exclusive mutations are enriched in hotspot codons in these 3' splicing factor proteins resulting in aberrant alternative mRNA splicing in hematopoietic cells. However, each mutant protein predominantly affects a distinct set of alternatively spliced downstream target genes suggesting that common downstream pathway alterations or cellular response to mutation expression, rather than specific shared splicing targets, may be responsible for MDS phenotypes, including dysplasia, ineffective hematopoiesis, and cytopenias.<sup>9–11</sup>

*U2AF1* provides a unique opportunity to address this question because it has two hotspot positions (serine 34 [S34] and glutamine 157 [Q157]) that are each commonly mutated in MDS and are associated with unique mRNA splicing consequences.<sup>12</sup> In addition, *U2AF1* S34 and Q157 codon mutations co-occur with mutations in different genes (e.g., *BCOR* and *ASXL1*, respectively) and patients with these mutations may have different hematopoietic phenotypes—highlighting that these mutations may induce distinct phenotypes.<sup>13–16</sup> We asked if the splicing differences resulting from S34F and Q157R mutations were thus associated with different or similar effects on hematopoiesis. To do so, we characterized and compared an established conditional S34F knock-in mouse model<sup>17</sup> to a new Cre/*lox* conditional mouse model with the Q157R mutation knocked-in to the endogenous *U2af1* locus, in order to directly study the

hematopoietic phenotype, transcriptional, and mRNA splicing consequences of individual *U2AF1* gene mutations *in vivo*.

## MATERIALS AND METHODS

### Animal models and experimental details

Experiments were performed per institutional guidelines for care and use of laboratory animals and approved by the Institutional Animal Care and Use Committee of Washington University in St. Louis (WUSTL). *U2af1*<sup>Q157R/+</sup> (MiniGene Q157R or 'MGQ157R') conditional knock-in mice were generated by Biocytogen (Waltham, MA). A full description of the targeting construct is described in the **Supplementary Methods**. *U2af1*<sup>S34F/+</sup> ('MGS34F') conditional knock-in (Jackson Laboratory [JAX] Stock #032638, Bar Harbor, ME),<sup>17</sup> *U2af1*<sup>fl/+</sup> conditional knockout (JAX Stock #037015),<sup>18</sup> and *Mx1-Cre* (JAX Stock #003556)<sup>19</sup> mice are described elsewhere. B6.SJL-*Ptprc*<sup>a</sup>*Pepc*<sup>b</sup>/BoyCrCrl or 'CD45.1' recipient mice were purchased from Charles River Laboratories (Stock #564). Heterozygous CD45.1/CD45.2 mice were bred by crossing C57BL/6J (B6; JAX Stock #000664) to B6.SJL-*Ptprc*<sup>a</sup>*Pepc*<sup>b</sup>/BoyJ (JAX Stock #002014). All mouse lines were on a B6 background. Genotyping primers are listed in **Supplementary Table 1**.

Details pertaining to bone marrow (BM) transplant, peripheral blood (PB) sampling and analysis, and flow cytometry setup and population gating are described in the **Supplementary Methods**.

### mRNA-sequencing (RNA-seq) and bioinformatics

BM myeloid progenitor (c-kit<sup>+</sup>Lineage<sup>-</sup>Sca-1<sup>-</sup>; KL) cells were sorted into FACS buffer and gDNA-depleted total RNA were purified from cell pellets using the NucleoSpin RNA Plus XS Micro Kit (Macherey-Nagel, Allentown, PA) in RNase-free water. RNA concentration and RIN were measured by Bioanalyzer (Agilent, Santa Clara, CA) and then cDNA libraries for RNA-seq were prepared by KAPA RNA HyperPrep Kit with RiboErase (Cat #KK8560/61; Roche, Indianapolis, IN). Detailed library preparation and bioinformatic analyses are described in **Supplementary Methods**.

Details pertaining to the bioinformatics and reanalysis of published MDS and AML RNA-seq datasets, analysis of *U2AF1* hotspot mutation co-occurrence in myeloid malignancies, and confirmation of splicing changes in mouse KL cell and MDS/sAML patient samples are also described in the **Supplementary Methods**. All patients provided written consent on a protocol approved by the WUSTL Human Studies Committee. Clinical characteristics of patients who donated research samples are listed in **Supplementary Table 2**.

### Statistics

Data were analyzed and visualized using GraphPad Prism 10 software (Boston, MA). Statistical tests are described in each figure legend.  $P < 0.05$  was considered statistically significant.

## RESULTS

### Establishing a mouse model with conditional knock-in of the Q157R mutation at the *U2af1* locus.

A conditional (Cre//*lox*-mediated) knock-in of the S34F mutation at the *U2af1* locus (MGS34F or *U2af1*<sup>S34F/+</sup>) was previously generated (Fig. 1A,B and **Supplementary Fig. 1A**).<sup>17</sup> To allow for direct comparison with the MGS34F mouse, a similar strategy was used to generate a conditional (Cre//*lox*-mediated) Q157R mutant allele at the endogenous *U2af1* locus (MGQ157R or *U2af1*<sup>Q157R/+</sup>) of B6 mice (Fig. 1C and **Supplementary Fig. 1B**). Successful introduction of the targeting vector at the *U2af1* locus was confirmed by Southern blot and Sanger sequencing (**Supplementary Fig. 1C**). To confirm Cre//*lox*-mediated hematopoietic expression of *U2af1*<sup>Q157R</sup> mRNA and assess the short-term effects of U2AF1<sup>Q157R</sup> in a non-transplant model (i.e., native hematopoiesis), we crossed heterozygous *U2af1*<sup>Q157R/+</sup> mice to *Mx1-Cre* transgenic mice (Fig. 1D-G and **Supplementary Fig. 1D-I**). *Mx1-Cre* is expressed in hematopoietic lineage cells following administration of polyinosinic-polycytidylic acid (plpC).<sup>19</sup>

Four weeks after plpC treatment of *U2af1*<sup>Q157R/+</sup>; *Mx1-Cre* mice, the *U2af1* wild-type (WT) and Q157R alleles were expressed at similar levels in BM myeloid progenitor (KL) cells by targeted NGS amplicon sequencing of cDNA (Fig. 1E). As expected, the WT and S34F alleles were also expressed at similar levels in BM KL cells from *U2af1*<sup>S34F/+</sup>; *Mx1-Cre* mice and only the WT allele was detected in *U2af1*<sup>+/+</sup>; *Mx1-Cre* control mice (Fig. 1E). It was previously reported that the Q157R mutation in *U2AF1* creates an alternative 5' splice site that leads to expression of a minor *U2AF1* isoform (termed 'Q157Rdel') with in-frame deletion of four amino acids immediately following the Q157R mutant codon. The *U2af1*<sup>Q157R/+</sup> mouse model recapitulates expression of the Q157Rdel isoform in BM KL cells (Fig. 1E and **Supplementary Fig. 1D**).

### U2AF1<sup>S34F</sup> and U2AF1<sup>Q157R</sup> cause different hematopoietic changes in mice.

To determine if S34F and Q157R result in similar short-term effects on native hematopoiesis, we performed complete blood counts and flow cytometric analysis on PB samples from *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/+</sup>, and *U2af1*<sup>+/+</sup> control mice (all *Mx1-Cre*<sup>+</sup>) four weeks after plpC treatment. Consistent with previous characterization,<sup>17</sup> *U2af1*<sup>S34F/+</sup> mice had no change in platelet counts, modestly reduced red blood cell (RBC) counts and hemoglobin levels (with elevated mean corpuscular volume [MCV]), and markedly reduced white blood cell counts compared to *U2af1*<sup>+/+</sup> mice (Fig. 1F). Flow cytometric analysis of *U2af1*<sup>S34F/+</sup> PB and BM demonstrated significant reductions in both myeloid and lymphoid lineages (**Supplementary Fig. 1E,F**). In contrast, *U2af1*<sup>Q157R/+</sup> mice had no significant PB or BM changes except for elevated MCV (Fig. 1F and **Supplementary Fig. 1E,F**). Assessment of BM hematopoietic stem and

progenitor cells (HSPC) four weeks after plpC treatment revealed that *U2af1*<sup>S34F/+</sup> mice had significantly reduced numbers of short-term hematopoietic stem cell (ST-HSC), KL, and common myeloid progenitor (CMP) populations with increased numbers of multipotent progenitor (MPP)2 and MPP3 populations compared with control mice (Fig. 1G). *U2af1*<sup>Q157R/+</sup> mice also had significantly reduced numbers of ST-HSC and KL populations and non-significant reductions in both CMP and megakaryocyte-erythroid progenitor (MEP) cells compared with control mice (Fig. 1G). *U2af1*<sup>S34F/+</sup> mice had a significant block in erythroid development in the BM and spleen, with an increased proportion of immunophenotypically defined nucleated erythroblasts (Ter119<sup>lo/hi</sup>CD71<sup>hi</sup>) and a decreased proportion of enucleated erythrocytes (Ter119<sup>hi</sup>CD71<sup>lo</sup>). In contrast, *U2af1*<sup>Q157R/+</sup> mice had a smaller but non-significant, increase in Ter119<sup>hi</sup>CD71<sup>hi</sup> cells in the spleen (**Supplementary Fig. 1G-I**).

To better evaluate the cell-intrinsic effects of both mutants on hematopoiesis, we transplanted BM from *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/+</sup>, or *U2af1*<sup>+/+</sup> control mice (CD45.2<sup>+</sup>; all *Mx1-Cre*<sup>+</sup>) into lethally irradiated WT congenic recipient mice (CD45.1<sup>+</sup>). Following engraftment, we treated mice (including controls) with plpC to induce expression of S34F and Q157R in donor-derived cells (Fig. 2A). Four weeks after plpC treatment, PB (Fig. 2B,C) and BM changes (**Supplementary Fig. 2A,C**) reflected similar overall trends observed in native hematopoiesis (Fig. 1F and **Supplementary Fig. 1E**) for both mutant mice. At 24 weeks, both mutant mice had significantly reduced PB RBC counts with increased MCV, as well as decreased hemoglobin in *U2af1*<sup>S34F/+</sup> mice. *U2af1*<sup>Q157R/+</sup> mice also had mildly increased platelet counts (Fig. 2B). *U2af1*<sup>S34F/+</sup> mice had significantly reduced PB and BM myeloid and lymphoid lineage cells, while *U2af1*<sup>Q157R/+</sup> mice had significantly decreased PB monocytes and a non-significant increase in BM monocytes (Fig. 2C and **Supplementary Fig. 2D**). Although myeloid and lymphoid lineages were significantly decreased in *U2af1*<sup>S34F/+</sup> mouse spleens at 4 weeks, there were no significant changes at 24 weeks (**Supplementary Fig. 2B,E**). HSPC populations reflected similar significant overall trends at 24 weeks compared to 4 weeks for *U2af1*<sup>S34F/+</sup> mice (Fig. 2D and **Supplementary Fig. 2C**). *U2af1*<sup>Q157R/+</sup> mice also showed similar, but non-significant, trends in HSPC population numbers at 24 weeks compared to 4 weeks (Fig. 2D and **Supplementary Fig. 2C**).

### ***U2af1*<sup>S34F/+</sup> HSCs are significantly more impaired than *U2af1*<sup>Q157R/+</sup> HSCs in BM repopulation assays.**

To compare the effects of S34F or Q157R expression on HSC reconstitution capacity, we performed competitive BM transplantation experiments. Lethally irradiated WT congenic recipient mice (CD45.1<sup>+</sup>) were transplanted with whole BM 'test' cells from *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/+</sup>, or *U2af1*<sup>+/+</sup> control mice (CD45.2<sup>+</sup>; all *Mx1-Cre*<sup>+</sup>) mixed with an equal number of competitor BM cells from WT congenic mice (CD45.1<sup>+</sup>/CD45.2<sup>+</sup>). Following engraftment, we treated mice (including controls) with plpC to induce expression of S34F and Q157R in donor-derived cells (Fig. 3A). Consistent with previous characterization,<sup>17</sup> we observed significant multi-lineage reductions in PB, BM, and spleen donor cell chimerism (CD45.2<sup>+</sup>) for *U2af1*<sup>S34F/+</sup> compared to *U2af1*<sup>+/+</sup> test cells (Fig. 3B-D and **Supplementary Fig. 3**). In contrast, the reduction in overall and multilineage PB donor cell chimerism for *U2af1*<sup>Q157R/+</sup>

compared to *U2af1*<sup>+/+</sup> test cells was less severe relative to *U2af1*<sup>S34F/+</sup> test cells (Fig. 3B,C). In addition, there were variable reductions in donor cell chimerism of PB, BM, and spleen myeloid lineages for *U2af1*<sup>Q157R/+</sup> compared to *U2af1*<sup>+/+</sup> test cells (Fig. 3C,D and **Supplementary Fig. 3**). Donor cell chimerism for all BM HSPC populations were significantly reduced for *U2af1*<sup>S34F/+</sup> compared to *U2af1*<sup>+/+</sup> test cells (Fig. 3E). However, reduced *U2af1*<sup>Q157R/+</sup> donor cell chimerism was restricted to the HSC and MPP2 populations, but not to the same degree as for *U2af1*<sup>S34F/+</sup> (Fig. 3E).

### **Hemizygous *U2af1*<sup>Q157R/-</sup> and *U2af1*<sup>S34F/-</sup> HSCs are both severely impaired in BM repopulation assays.**

We previously demonstrated that cell survival and reconstitution capacity are severely reduced for HSCs that express mutant U2AF1<sup>S34F</sup> without WT U2AF1 expression (hemizygous *U2af1*<sup>S34F/-</sup>).<sup>18</sup> Given the mild reconstitution defect observed for *U2af1*<sup>Q157R/+</sup> cells, we hypothesized that mutant U2AF1<sup>Q157R</sup> cells may not require the expression of WT U2AF1 for cell survival. To test this, we performed competitive BM transplantation experiments using test cells from three additional genotypes of mice: *U2af1*<sup>Q157R/-</sup>, *U2af1*<sup>S34F/-</sup>, and *U2af1*<sup>+/-</sup> mice (all *Mx1-Cre*<sup>+</sup>; Fig. 4A). Hemizygous conditional knock-in mice were generated by crossing heterozygous floxed mutant (S34F or Q157R) mice to heterozygous floxed knockout mice. Consistent with previous characterization,<sup>18</sup> we noted a rapid and significant loss in mature cell and HSPC donor cell chimerism (CD45.2<sup>+</sup>) in the PB, BM, and spleen for hemizygous *U2af1*<sup>S34F/-</sup> (but not *U2af1*<sup>+/-</sup>) compared to *U2af1*<sup>+/+</sup> test cells following administration of plpC (Fig. 4B-D and **Supplementary Fig. 4A,B**). We also observed an identical rapid loss in mature cell and HSPC chimerism for hemizygous *U2af1*<sup>Q157R/-</sup> compared to *U2af1*<sup>+/+</sup> test cells (Fig. 4B-D and **Supplementary Fig. 4A,B**). This indicates that the expression of WT U2AF1 is required for the viability of either U2AF1<sup>S34F</sup> or U2AF1<sup>Q157R</sup> mutant expressing HSCs, consistent with *U2AF1* being a haplo-essential gene,<sup>18</sup> and reinforcing that the *U2af1*<sup>Q157R</sup> allele impairs U2AF1 function despite the less severe phenotype compared to *U2af1*<sup>S34F</sup>.

### **U2AF1<sup>S34F</sup> and U2AF1<sup>Q157R</sup> induce distinct gene expression changes in mouse myeloid progenitor cells.**

To characterize the effects of mutant U2AF1 on gene expression *in vivo*, we performed RNA-seq of total RNA (rRNA-depleted) from BM myeloid progenitor (KL) cells from *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/+</sup>, or *U2af1*<sup>+/+</sup> control mice (all *Mx1-Cre*<sup>+</sup>) under native hematopoiesis conditions (as in Fig. 1D). KL cells were isolated by FACS at 4 weeks after completion of plpC injections and the variant allele frequencies of the S34F and Q157R mutations were near 50% (**Supplementary Fig. 5A**). Unsupervised principal component analysis of gene expression values (N = 19312 genes) segregated *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/+</sup>, and *U2af1*<sup>+/+</sup> KL cells (Fig. 5A and **Supplementary Table 3**). Reanalysis of *U2af1*<sup>S34F/+</sup> Native KL RNA-seq data published by Fei *et al.*<sup>17</sup> demonstrated a strong concordance in gene expression changes with our *U2af1*<sup>S34F/+</sup> KL data (**Supplementary Fig. 5B,C**). In our dataset, we identified 185 differentially expressed genes (DEGs; FDR < 0.05 and |log<sub>2</sub>FC| > 1) in *U2af1*<sup>S34F/+</sup> compared to *U2af1*<sup>+/+</sup> control mice (Fig. 5B) and 77 DEGs in *U2af1*<sup>Q157R/+</sup> KL cells (Fig. 5C). There were only 12 DEGs shared between *U2af1*<sup>S34F/+</sup> and

*U2af1*<sup>Q157R/+</sup> KL cells (4.8%; Fig. 5D) with no overlap in gene ontology (GO) analysis (**Supplementary Fig. 5D,E** and **Supplementary Table 4**). Gene set enrichment analysis (GSEA) revealed significant positive enrichment of the p53 pathway in *U2af1*<sup>S34F/+</sup> KL cells and negative enrichment of immune response related Hallmark pathways in both *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> KL cells compared to *U2af1*<sup>+/+</sup> KL cells (Fig. 5E).

### **U2AF1<sup>S34F</sup> and U2AF1<sup>Q157R</sup> induce distinct alternative pre-mRNA splicing changes in myeloid progenitor cells.**

Using the same bulk RNA-seq data, we next characterized the effects of mutant U2AF1 on alternative mRNA splicing *in vivo*. We employed replicate multivariate analysis of transcript splicing (rMATS)<sup>20</sup> to assess differential alternative pre-mRNA splicing of five different types of annotated splicing events (alternative 3' or 5' splice sites [A3SS, A5SS], mutually exclusive exons [MXE], retained introns [RI], and skipped exons [SE]) in *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/+</sup>, and *U2af1*<sup>+/+</sup> KL cells. Unsupervised principal component analysis of inclusion ratios (referred to as 'percent spliced-in' or 'PSI') for all annotated alternative splicing events (N = 11580) revealed that global alternative pre-mRNA splicing is distinct between *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/+</sup>, and *U2af1*<sup>+/+</sup> KL cells (Fig. 6A and **Supplementary Tables 5–7**).

We then applied rMATS to identify 1048 and 580 differentially spliced events (DSEs; FDR < 0.05 and | $\Delta$ PSI| > 0.05 vs *U2af1*<sup>+/+</sup>) in *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> KL cells, respectively (Fig. 6B and **Supplementary Table 8**). We also applied our rMATS analysis pipeline to the *U2af1*<sup>S34F/+</sup> Native KL RNA-seq dataset published by Fei *et al.*<sup>17</sup> (**Supplementary Fig. 6A,B**) and observed a strong concordance (i.e., unidirectional  $\Delta$ PSI values) between DSEs shared between the two *U2af1*<sup>S34F/+</sup> KL datasets (**Supplementary Fig. 6C**). Thus, rMATS analysis of independent RNA-seq data demonstrates that the *U2af1*<sup>S34F/+</sup> mouse model produces robust and reproducible gene expression and alternative pre-mRNA splicing changes in hematopoietic cells *in vivo* (**Supplementary Figs. 5B,6C**). In line with previous studies of U2AF1 mutant cell lines and patient HSPC, SE events represented the majority of DSEs identified in U2AF1 mutant mouse KL cells (Fig. 6B and **Supplementary Fig. 6A**).<sup>17,21,22</sup> *U2af1*<sup>S34F/+</sup> SE DSEs also favored exon exclusion ('skipping') over exon inclusion.<sup>23</sup> Of note, *U2af1*<sup>Q157R/+</sup> DSEs were more equally distributed between RI and SE events (Fig. 6B). The overlap of DSE shared between *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> KL cells was low (125 events or 8.3%; Fig. 6C). Conversion of DSE to differentially spliced genes (DSG) revealed 196 genes (17.5%) aberrantly spliced in common between the two mutants (Fig. 6D). GO analysis revealed that DSGs from *U2af1*<sup>S34F/+</sup> KL cells were most significantly enriched in mRNA binding and metabolism gene sets, as well as histone post-translational modification and stress granule<sup>23</sup> related gene sets (**Supplementary Fig. 6D** and **Supplementary Table 9**). DSGs from *U2af1*<sup>Q157R/+</sup> KL cells were enriched in mRNA gene sets to a weaker extent than *U2af1*<sup>S34F/+</sup> (**Supplementary Fig. 6D** and **Supplementary Table 9**).

Analysis of consensus 3' splice site (3'SS) sequences from differentially spliced SE events in *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> KL cells confirmed previous dependencies identified in U2AF1 mutant cell lines and patient HSPC. Specifically, exon inclusion favored a C and exon exclusion favored a T at the - 3 position of the 3'SS in *U2af1*<sup>S34F/+</sup> cells (Fig. 6E, middle). In contrast, exon inclusion favored a G and exon exclusion favored an A at the + 1 position of the 3'SS in *U2af1*<sup>Q157R/+</sup> cells (Fig. 6E, right). Overall, these findings highlight that the U2AF1<sup>S34F</sup> and U2AF1<sup>Q157R</sup> mutants induce significant but distinct changes to alternative mRNA splicing *in vivo*.

### **U2af1<sup>S34F/+</sup> and U2af1<sup>Q157R/+</sup> mouse models recapitulate alternative pre-mRNA splicing changes found in MDS and AML patients.**

To assess how well alternative splicing changes in mouse KL cells recapitulate changes seen in MDS and AML patient hematopoietic cells, we performed a meta-analysis using publicly available RNA-seq data from three published studies.<sup>9,11,24</sup> Each study included 2–10 U2AF1<sup>S34F</sup> and only 1–2 U2AF1<sup>Q157R</sup> patients. Therefore, U2AF1<sup>R156H</sup> and U2AF1<sup>Q157(P/R)</sup> patient samples were grouped together (N = 4–5 *U2AF1*<sup>R156H/Q157(P/R)</sup> patients per study; Fig. 7A) consistent with previous studies demonstrating similar 3'SS sequence dependencies.<sup>21,22</sup> In each study, samples from MDS/AML patients who did not have identifiable mutations in *SF3B1* or *SRSF2* were used as a comparator (Splicing Factor [SF]<sup>WT</sup>). To allow for a more rigorous analysis of differential splicing, we reanalyzed the FASTQ files for each study using the same analysis workflows and applied the same significance thresholds (FDR < 0.05 and |ΔPSI| > 0.05 vs SF<sup>WT</sup>) as used for the analysis of mouse KL cells. Using this approach, we credentialed each of the three datasets (referred to as Madan,<sup>11</sup> Pellagatti,<sup>9</sup> and Beat AML<sup>24</sup>) (Fig. 7A and **Supplementary Fig. 7A-I** and **Supplementary Tables 10–15**). Specifically, SE events were the most frequent DSE type identified in each study for S34F and R156/Q157 (**Supplementary Fig. 7A-C**) and these events favored the characteristic consensus 3'SS sequence dependencies identified previously (**Supplementary Fig. 7G-I**).<sup>9,17,21,22,25,26</sup> To increase rigor of our meta-analysis we prioritized only the DSEs that were shared between at least two of the three MDS/AML datasets for either *U2AF1*<sup>S34F</sup> or *U2AF1*<sup>R156/Q157</sup> (Fig. 7B and **Supplementary Table 16**). The distribution of these DSEs was similar to each individual dataset with SE events still representing the majority event type in U2AF1 mutant MDS/AML cells (Fig. 7C). As in the mice, the overlap of DSE shared between *U2AF1*<sup>S34F</sup> and *U2AF1*<sup>R156/Q157</sup> MDS/AML cells was low (144 of 1978 events or 7.3%; Fig. 7D). Conversion of DSE to DSG revealed a total of 284 of 1305 genes (21.8%) aberrantly spliced in common between the two mutants (Fig. 7E).

The overlap of DSG identified in human and mouse cells revealed that approximately 20% of aberrantly spliced genes in KL (mouse) cells were also mis-spliced in MDS/AML (human) cells for both U2AF1<sup>S34F</sup> (17.6% shared) and U2AF1<sup>Q157R</sup> (19.7% shared) mutants (Fig. 7F,G). GO analysis revealed that shared S34F DSGs were most significantly enriched in mRNA binding and metabolism gene sets as well as stress granule<sup>23</sup> and mRNA translation related gene sets (Fig. 7H and **Supplementary Table 17**). Shared Q157R DSGs were less significantly enriched in mRNA gene sets than S34F. Histone binding and DNA

damage response gene sets were among some of the significantly enriched gene sets for Q157R DSGs (Fig. 7H and **Supplementary Table 17**).

We validated several of these putatively shared aberrant splicing changes identified by rMATS analysis by performing RT-PCR followed by gel electrophoresis of RNA isolated from additional mouse KL cell samples (N = 4 per genotype) and MDS patient samples (N = 4–9 per genotype). Consistent with previous observations,<sup>17,25–28</sup> we confirmed aberrant splicing of functionally relevant transcripts (*H2AFY* and *GNAS*) in *U2AF1*<sup>S34F</sup> mutant mouse KL and MDS cells (Fig. 7I,J). We also demonstrate that aberrantly spliced transcripts (*MPHOSPH9*, *SETD5*, *ATP6V0A1*, and *CLIP1*) in *U2AF1*<sup>Q157R</sup> mutant MDS patient cells are similarly mis-spliced in KL cells from *U2af1*<sup>Q157R/+</sup> mice (Fig. 7I,K-L). Aberrant splicing of *CLIP1* is one example of an SE event that is differentially spliced in opposite directions by *U2AF1*<sup>S34F</sup> (increased exon inclusion) and *U2AF1*<sup>Q157R</sup> (increased exon skipping/exclusion) in mouse and human cells (Fig. 7L), further highlighting the distinct splicing differences induced by these two *U2AF1* mutants.

### ***U2AF1*<sup>R156/Q157</sup> mutations are enriched in patients with CMML and MPN compared to *U2AF1*<sup>S34F</sup> mutations.**

Given the differences in gene expression, splicing, and hematopoietic phenotypes induced by *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> mutations in mice, we asked if the two hotspot mutations were differentially enriched in various myeloid neoplasms. We identified 487 patients with a diagnosis of AML, sAML (from MDS), MDS, CMML, or MPN who had a *U2AF1* mutation based on available sequencing data and calculated the proportion of patients with *U2AF1*<sup>R156/Q157</sup> or *U2AF1*<sup>S34</sup> mutations (see **Supplementary Methods**). We observed that *U2AF1*<sup>R156/Q157</sup> mutations were more common in CMML and MPN patients, *U2AF1*<sup>S34</sup> mutations more common in sAML and AML patients, and a similar proportion of both mutations occurred in MDS (Fig. 8A and **Supplementary Table 18**). The co-occurrence of *U2AF1* and signaling gene mutations also differed across myeloid neoplasms, with *NRAS* and *FLT3* mutations being more common with S34 mutations and *CBL*, *PTPN11* and *CSF3R* mutations more common with R156/Q157 mutations. Similar to previous reports by our group and others, we also observed preferential co-occurrence of other gene mutations with *U2AF1*<sup>R156/Q157</sup> (e.g., *ASXL1*) or *U2AF1*<sup>S34F</sup> (e.g., *BCOR*) mutations in MDS patients (Fig. 8B and **Supplementary Tables 19–20**).<sup>15,16</sup>

## **DISCUSSION**

In this study, we characterized the *in vivo* consequences of expressing two myeloid neoplasm-associated hotspot mutations in *U2AF1* that code for S34F and Q157R substitutions. Our results indicate that the two mutations induce distinct hematopoietic phenotypes in mice, suggesting that the *U2af1*<sup>S34F</sup> and *U2af1*<sup>Q157R</sup> mutations should not be conflated as they may impact disease pathogenesis differently in patients. Mice expressing *U2af1*<sup>S34F</sup> have a more severe reduction in their PB and BM cell counts, and reduced HSPCs repopulating ability, compared to mice expressing *U2af1*<sup>Q157R</sup>. The expression and splicing of the majority of target genes are unique between the mutations, in both mouse and human

samples, potentially driving the phenotypic differences induced by the two mutations. The two mutations co-occur with different gene mutations and are not equally represented in various myeloid neoplasms, suggesting that multiple mechanisms are likely to drive the pathogenesis of *U2AF1* mutant myeloid diseases.

Our results add to the growing body of literature highlighting the paradigm that different hotspot mutations in a specific cancer gene can lead to distinct functional consequences and should, therefore, not necessarily be conflated. In one of the more well studied examples, different *KRAS* hotspot mutations (e.g., G12, G13, Q61) lead to varying levels of *KRAS* activation through modulation of distinct biochemical properties of *KRAS*.<sup>29</sup> In turn, different *KRAS* hotspot mutations confer different prognostic value in various cancers (e.g., colorectal cancer) and are predictive of response to chemotherapy and/or targeted therapies.<sup>29</sup> Similarly, the prognostic significance of distinct *SF3B1* mutations can be different in MDS, including their impact on overall survival.<sup>30,31</sup> With respect to *U2AF1*, a prognostic scoring model for MF (MIPSS70 + v2.0) now incorporates the negative impact of Q157 (but not S34) codon mutations.<sup>32</sup> *U2AF1* S34 and Q157 codon-specific clinical characteristics have also been reported in MDS.<sup>33,34</sup> Further studies are needed to fully understand the functional impact of each *U2AF1* hotspot mutation in patients and determine whether these differences confer consistent prognostic or therapeutic value across the spectrum of myeloid malignancies.

Enrichment of *U2AF1*<sup>S34</sup> vs *U2AF1*<sup>Q157</sup> mutations in different myeloid diseases suggest that mutations may contribute to the disease phenotype by differences in the target genes that they dysregulate and/or cooperating gene mutations. Identifying and validating the key target genes that are dysregulated and confer mutation-specific cellular phenotypes will require future *in vivo* functional studies. Additionally, based on differences in hotspot mutations in other cancers, the cellular ‘soil’ that a S34 or Q157 mutation occurs in likely also matters. In *U2AF1*-mutated solid tumors, particularly lung adenocarcinomas and endometrial cancers, S34 codon mutations are highly enriched compared to Q157 codon mutations.<sup>35</sup> This observation is not specific to *U2AF1*, as *SF3B1* R625 codon mutations are enriched in uveal and cutaneous melanomas, whereas K700 mutations are more common in breast cancer and chronic lymphoid leukemia specimens.<sup>35</sup> The subtle phenotype in the *U2af1*<sup>Q157R/+</sup> mouse, including lack of severe cytopenias and possibly a slight increase in platelets, could make Q157R cells more permissive to transformation with a MPN-associated cooperating mutation (e.g., *CSF3R*) resulting in higher blood counts, something that will require future studies. In contrast, S34F induces cytopenias in mice and may contribute to cytopenias seen in MDS. In addition, S34 and Q157 do not cooperate with the same mutations in MDS (e.g., *BCOR* with S34 > Q157 and *ASXL1* with Q157 > S34),<sup>36</sup> and this could impact mutation-associated phenotypes in patients.<sup>15,37</sup> Finally, *U2AF1* hotspot mutation phenotypes could be influenced by the order of cooperating gene mutation acquisition (i.e., *U2AF1* mutation occurring before or after a cooperating gene mutation) or the presence of hematopoietic stressors, requiring future experiments.

Collectively, our results support that *U2AF1*<sup>S34F</sup> and *U2AF1*<sup>Q157R</sup> mutations induce distinct hematopoietic, gene expression, and RNA splicing phenotypes *in vivo*. Larger population studies will be needed to determine if these phenotypic changes translate into clinico-pathologic differences in patients warranting separate classification.

## Declarations

### ACKNOWLEDGEMENTS

The authors thank Harold E. Varmus for the gift of the MGQ157R mice; William C. Eades, Daniel Schweppe, Michael Savio, and Matt Patana of the the Alvin J. Siteman Cancer Center at Washington University School of Medicine (WUSM) and Barnes-Jewish Hospital (St. Louis, MO) for the use of the Siteman Flow Cytometry Core and sorting assistance; Zev J. Greenberg for sorting assistance; Nichole M. Helton for help with RNA-seq preparation; the McDonnell Genome Institute (MGI) for RNA-seq; Jessica Hoisington-Lopez and MariaLynn Crosby of the DNA Sequencing Innovation Lab (DSIL) at WUSM Center for Genome Sciences & Systems Biology for help with amplicon sequencing; Eric J. Duncavage and Kiran Vij for hematopathology expertise; John D. Pfeifer for critical support; and Timothy J. Ley, Daniel C. Link, and members of the Walter Lab for useful discussions.

This work was supported by grants to M.J.W. from the Edward P. Evans Foundation, Taub Foundation, Lottie Caroline Hardy Trust, Foundation for Barnes-Jewish Hospital Cancer Frontier Fund, the National Cancer Institute (NCI) of the National Institutes of Health (NIH) (P01 CA101937, Timothy J. Ley, PI; P50 CA171963, Daniel C. Link, PI), and the Leukemia & Lymphoma Society (7024-21). M.O.A. was supported by career development awards from the Edward P. Evans Foundation and NCI and National Heart, Lung, and Blood Institutes (NHLBI) of the NIH under the award numbers K12 CA167540 (Washington University Paul Calabresi) and K08 HL159354. M.Z. was supported by an American Society of Hematology (ASH) Physician-Scientist Career Development Award. C.P. was supported by an ASH Minority Hematology Graduate Award (MHGA) and a NIH/NCI award (F31 CA284751). O.A.-W. is supported by the Neil S. Hirsch Foundation, Edward P. Evans Foundation, Break Through Cancer, NIH/NCI (R01 CA251138, R01 CA242020, R01 CA283364, and P50 CA254838), and NIH/NHLBI (R01 HL128239), and the Leukemia & Lymphoma Society. T.A.G. is supported by the Edward P. Evans Foundation, NIH/NCI (P50 CA171963), and the Leukemia & Lymphoma Society (7024-21). Sample banking was provided by the Genomics of Acute Myeloid Leukemia Program Project Grant (P01 CA101937, Timothy J. Ley, PI) and Specialized Program of Research Excellence in Acute Myeloid Leukemia (P50 CA171963, Daniel C. Link, PI). We thank the Alvin J. Siteman Cancer Center at WUSM and Barnes-Jewish Hospital, as well as the Institute of Clinical and Translational Sciences (ICTS) at Washington University in St. Louis, for the use of the Siteman Flow Cytometry Core, the Tissue Procurement Core, and the Genome Technology Access Center. The Siteman Cancer Center is supported in part by an NCI Cancer Center Support Grant (P30 CA091842) and the ICTS is funded by the NIH NCATS Clinical and Translational Science Award (CTSA) program (UL1 TR002345). The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## AUTHOR CONTRIBUTIONS

M.O.A., D.L.F., T.A.G., O.A.-W., and M.J.W. were responsible for conceptualization; M.O.A., S.N.S., D.L.F., and M.J.W. were responsible for methodology; M.O.A., S.N.S., J.S., M.Z., C.C., and S.G. performed investigation; M.O.A., S.N.S., and M.J.W. wrote the original draft manuscript; M.O.A., S.N.S., and M.J.W. reviewed and edited the manuscript; M.J.W. was responsible for funding acquisition; and M.J.W. provided supervision.

## COMPETING INTERESTS

O.A.-W. is a founder and scientific advisor of Codify Therapeutics, holds equity, and receives research funding from this company. O.A.-W. has served as a consultant for Amphista Therapeutics, and MagnetBio, and is on scientific advisory boards of Envisagenics Inc. and Harmonic Discovery Inc. O.A.-W. has received research funding from Astra Zeneca, Nurix Therapeutics, and Minovia Therapeutics, unrelated to this study. The remaining authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material.

## DATA AVAILABILITY

Data are available on request to the corresponding author. RNA-seq data generated from this study have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (accession # GSE282060).

## References

1. Damm F, Kosmider O, Gelsi-Boyer V, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. 2012;119(14):3211–3218.
2. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28(2):241–247.
3. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616–3627; quiz 3699.
4. Barraco D, Elala YC, Lasho TL, et al. Molecular correlates of anemia in primary myelofibrosis: a significant and independent association with U2AF1 mutations. *Blood Cancer J*. 2016;6(4):e415.
5. Kandarpa M, Robinson D, Wu Y-M, et al. Broad next generation integrated sequencing of myelofibrosis identifies disease-specific and age-related genomic alterations. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res*. 2024;
6. Palomo L, Meggendorfer M, Hutter S, et al. Molecular landscape and clonal architecture of adult myelodysplastic/myeloproliferative neoplasms. *Blood*. 2020;136(16):1851–1862.

7. Patel BJ, Przychodzen B, Thota S, et al. Genomic determinants of chronic myelomonocytic leukemia. *Leukemia*. 2017;31(12):2815–2823.
8. Lindsley RC, Mar BG, Mazzola E, et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood*. 2015;125(9):1367–1376.
9. Pellagatti A, Armstrong RN, Steeples V, et al. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood*. 2018;132(12):1225–1240.
10. Hershberger CE, Moyer DC, Adema V, et al. Complex landscape of alternative splicing in myeloid neoplasms. *Leukemia*. 2021;35(4):1108–1120.
11. Madan V, Li J, Zhou S, et al. Distinct and convergent consequences of splice factor mutations in myelodysplastic syndromes. *Am. J. Hematol*. 2020;95(2):133–143.
12. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64–69.
13. Tefferi A, Mudireddy M, Finke CM, et al. U2AF1 mutation variants in myelodysplastic syndromes and their clinical correlates. *Am. J. Hematol*. 2018;93(6):E146–E148.
14. Tefferi A, Finke CM, Lasho TL, et al. U2AF1 mutation types in primary myelofibrosis: phenotypic and prognostic distinctions. *Leukemia*. 2018;32(10):2274–2278.
15. Badar T, Vanegas YAM, Nanaa A, et al. U2AF1 pathogenic variants in myeloid neoplasms and precursor states: distribution of co-mutations and prognostic heterogeneity. *Blood Cancer J*. 2023;13(1):149.
16. Alberti MO, Srivatsan SN, Shao J, et al. Discriminating a common somatic ASXL1 mutation (c.1934dup; p.G646Wfs\*12) from artifact in myeloid malignancies using NGS. *Leukemia*. 2018;32(8):1874–1878.
17. Fei DL, Zhen T, Durham B, et al. Impaired hematopoiesis and leukemia development in mice with a conditional knock-in allele of a mutant splicing factor gene U2af1. *Proc. Natl. Acad. Sci. U. S. A*. 2018;115(44):E10437–E10446.
18. Wadugu BA, Nonavinkere Srivatsan S, Heard A, et al. U2af1 is a haplo-essential gene required for hematopoietic cancer cell survival in mice. *J. Clin. Invest*. 2021;131(21):e141401.
19. Kühn R, Schwenk F, Aguet M, Rajewsky K. Inducible gene targeting in mice. *Science*. 1995;269(5229):1427–1429.
20. Shen S, Park JW, Lu Z, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A*. 2014;111(51):E5593-5601.
21. Pangallo J, Kiladjian J-J, Cassinat B, et al. Rare and private spliceosomal gene mutations drive partial, complete, and dual phenocopies of hotspot alterations. *Blood*. 2020;135(13):1032–1043.
22. Ilagan JO, Ramakrishnan A, Hayes B, et al. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res*. 2015;25(1):14–26.

23. Biancon G, Joshi P, Zimmer JT, et al. Precision analysis of mutant U2AF1 activity reveals deployment of stress granules in myeloid malignancies. *Mol. Cell*. 2022;82(6):1107-1122.e7.
24. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526–531.
25. Shirai CL, Ley JN, White BS, et al. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*. 2015;27(5):631–643.
26. Yip BH, Steeples V, Repapi E, et al. The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. *J. Clin. Invest*. 2017;127(6):2206–2221.
27. Kim SP, Srivatsan SN, Chavez M, et al. Mutant U2AF1-induced alternative splicing of H2afy (macroH2A1) regulates B-lymphopoiesis in mice. *Cell Rep*. 2021;36(9):109626.
28. Wheeler EC, Vora S, Mayer D, et al. Integrative RNA-omics Discovers GNAS Alternative Splicing as a Phenotypic Driver of Splicing Factor-Mutant Neoplasms. *Cancer Discov*. 2022;12(3):836–855.
29. Haigis KM. KRAS Alleles: The Devil Is in the Detail. *Trends Cancer*. 2017;3(10):686–697.
30. Dalton WB, Helmenstine E, Pieterse L, et al. The K666N mutation in SF3B1 is associated with increased progression of MDS and distinct RNA splicing. *Blood Adv*. 2020;4(7):1192–1196.
31. Kanagal-Shamanna R, Montalban-Bravo G, Sasaki K, et al. Only SF3B1 mutation involving K700E independently predicts overall survival in myelodysplastic syndromes. *Cancer*. 2021;127(19):3552–3565.
32. Tefferi A, Guglielmelli P, Lasho TL, et al. MIPSS70+ Version 2.0: Mutation and Karyotype-Enhanced International Prognostic Scoring System for Primary Myelofibrosis. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol*. 2018;36(17):1769–1770.
33. Li B, Liu J, Jia Y, et al. Clinical features and biological implications of different U2AF1 mutation types in myelodysplastic syndromes. *Genes. Chromosomes Cancer*. 2018;57(2):80–88.
34. Wang H, Guo Y, Dong Z, et al. Differential U2AF1 mutation sites, burden and co-mutation genes can predict prognosis in patients with myelodysplastic syndrome. *Sci. Rep*. 2020;10(1):18622.
35. Seiler M, Peng S, Agrawal AA, et al. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep*. 2018;23(1):282-296.e4.
36. Bernard E, Hasserjian RP, Greenberg PL, et al. Molecular taxonomy of myelodysplastic syndromes and its clinical implications. *Blood*. 2024;144(15):1617–1632.
37. Pritzl SL, Gurney M, Badar T, et al. Clinical and molecular spectrum and prognostic outcomes of U2AF1 mutant clonal hematopoiesis- a prospective mayo clinic cohort study. *Leuk. Res*. 2023;125:107007.
38. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*. 2011;6(7):e21800.

## Figures

Figure 1

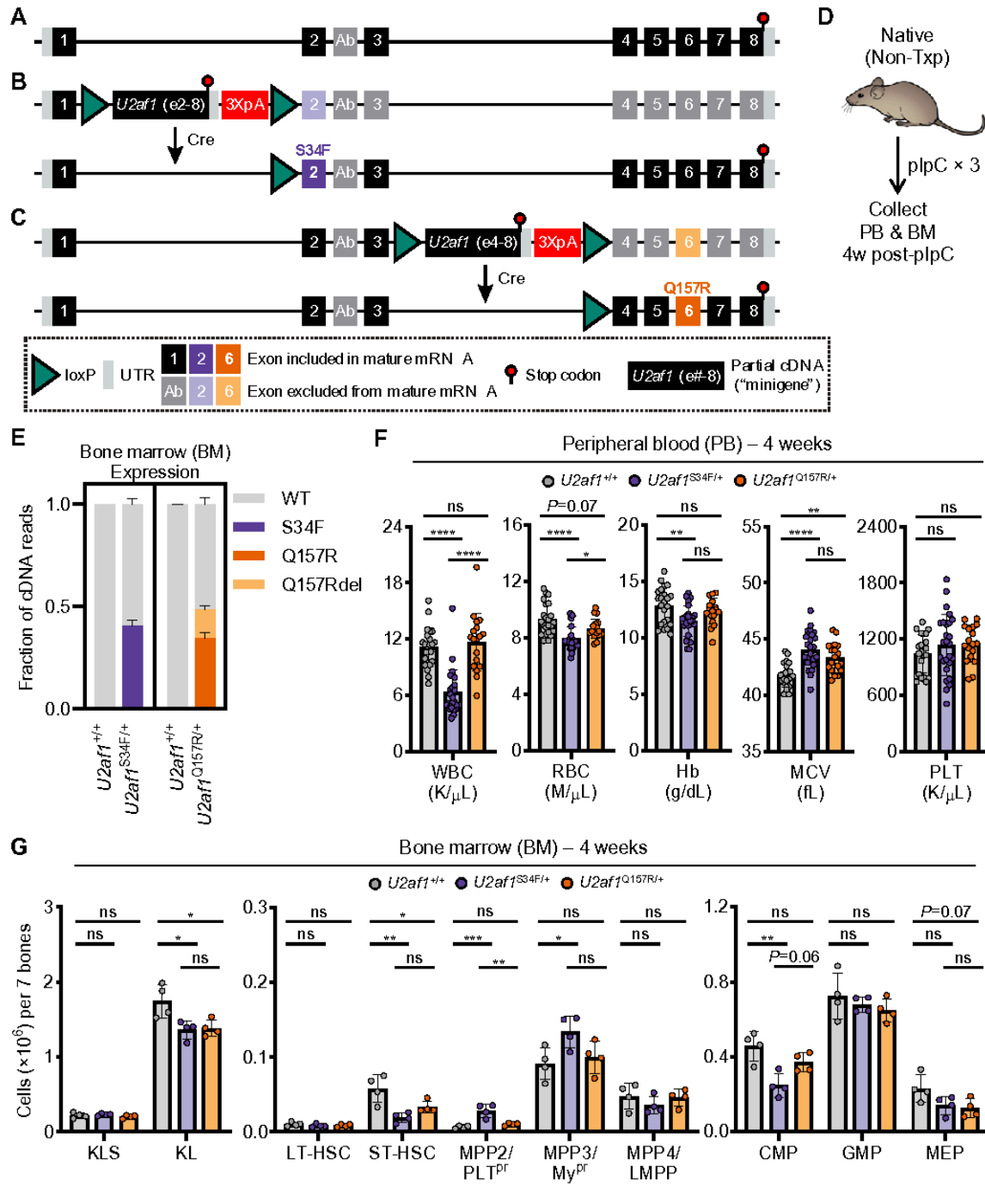


Figure 1

**Characterization of native hematopoiesis in *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> conditional knock-in mice. (A-C)** Diagrams of the wild-type (WT) mouse endogenous *U2af1* locus (**A**) and endogenous *U2af1* locus with conditional knock-in of either the S34F mutation in exon 2 (TCT>TIT) and upstream loxP flanked (floxed) *MiniGene* (MG; encoding WT *U2af1* exons 2-8) in intron 1 ('MGS34F'; **B**) or Q157R mutation in exon 6 (CAG>CGG) and upstream floxed MG (encoding WT *U2af1* exons 4-8) in intron 3 ('MGQ157R'; **C**).

Cre-mediated recombination of the floxed MGS34F or MGQ157R alleles results in removal of the WT MG cassette and conditional expression of *U2af1*<sup>S34F</sup> or *U2af1*<sup>Q157R</sup>, respectively, from the mouse endogenous locus. 3XpA, three repeats of the SV40 late polyadenylation signal. See **Supplementary Fig. 1A-C** for targeting vectors and additional locus detail. **(D)** Non-transplant (native hematopoiesis) assay design. *U2af1*<sup>+/+</sup>, *U2af1*<sup>S34F/+</sup>, or *U2af1*<sup>Q157R/+</sup> mice (all *Mx1-Cre*<sup>+</sup>) were treated with three doses of plpC at 6-12 weeks of age. **(E)** Assessment of S34F and Q157R mRNA expression levels in BM KL cells at 4 weeks post-plpC treatment. cDNA was prepared from KL cells for targeted NGS amplicon sequencing of the S34 (left) and Q157 (right) codons. The fraction of reads matching either WT or mutated alleles is plotted. *U2af1*<sup>+/+</sup> mice were assessed for both S34F and Q157R/Q157Rdel alleles. The Q157R mutation in *U2af1* creates an alternative 5' splice site that leads to expression of a minor *U2af1* isoform (termed "Q157Rdel") with in-frame deletion of four amino acids immediately following the Q157R mutant codon. See also **Supplementary Fig. 1D**. N=3 mice per genotype. **(F)** Complete blood count analysis (white blood cell [WBC], red blood cell [RBC], and platelet [PLT] counts, Hb [hemoglobin], and RBC mean corpuscular volume [MCV]) of PB samples from mice at 4 weeks post-plpC. N=18-26 mice per genotype pooled from five independent experiments. **(G)** Absolute cell counts of BM HSPC populations (KLS [c-kit<sup>+</sup>Lineage<sup>-</sup>Sca-1<sup>+</sup>], KL [c-kit<sup>+</sup>Lineage<sup>-</sup>Sca-1<sup>-</sup>], long- and short-term HSC [LT-HSC and ST-HSC], multipotent progenitors [MPP2, MPP3, and MPP4], common myeloid progenitors [CMP], granulocyte-macrophage progenitors [GMP], and megakaryocyte-erythrocyte progenitors [MEP]) were determined by flow cytometric analysis at 4 weeks post-plpC. N=4 mice per genotype. See also **Supplementary Fig. 1E-G**. Results represent the mean ± standard deviation (SD) (**E-G**). One-way analysis of variance (ANOVA) with Tukey multiple comparison correction (**F-G**) was used for the comparison of groups. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001. ns, not significant (or labeled if *P* < 0.10).

Figure 2

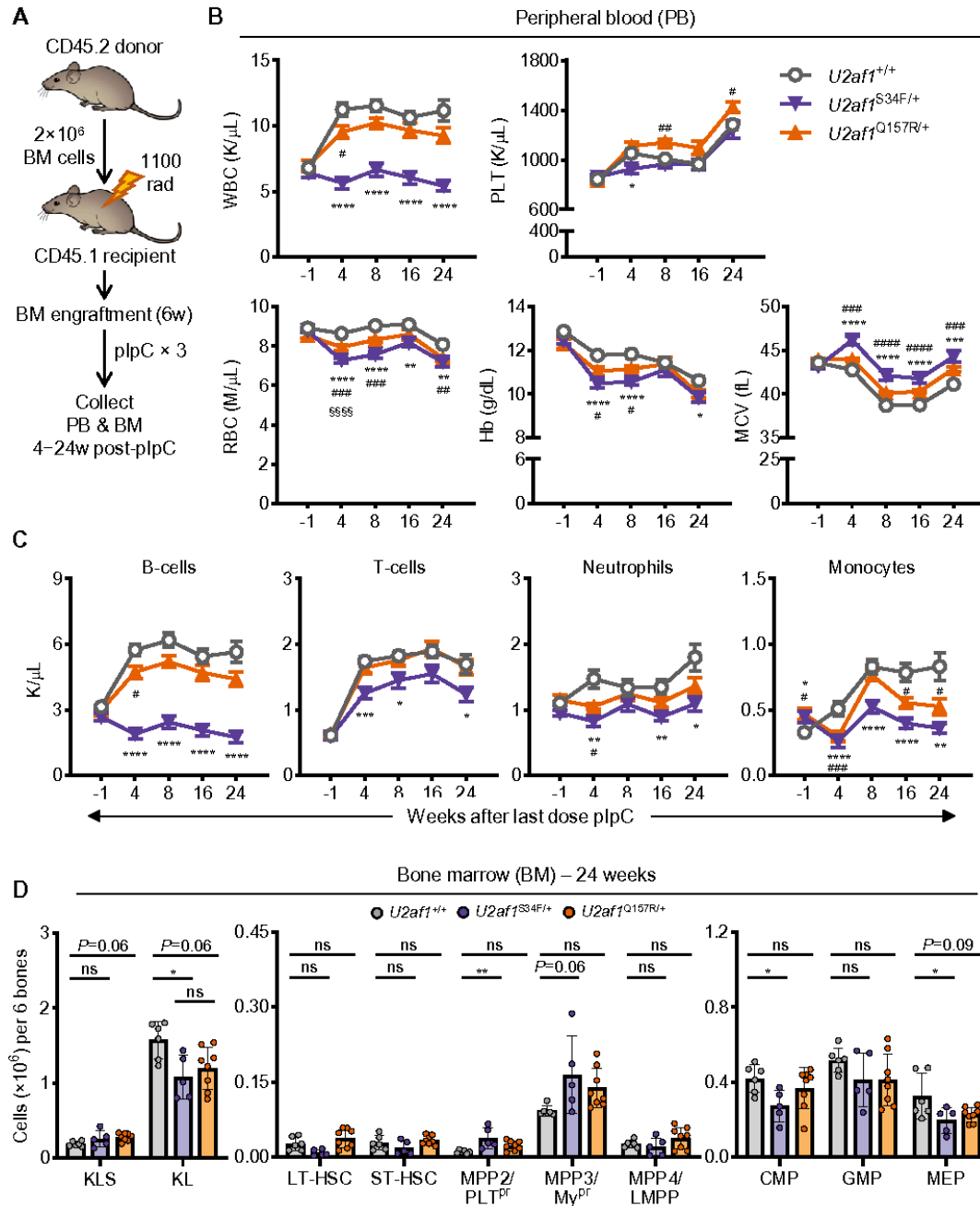


Figure 2

**U2AF1<sup>S34F</sup> and U2AF1<sup>Q157R</sup> cause different cell-intrinsic effects on hematopoiesis.** (A) Transplant assay design. CD45.2<sup>+</sup> donor BM cells from  $U2af1^{+/+}$ ,  $U2af1^{S34F/+}$ , or  $U2af1^{Q157R/+}$  mice (all  $Mx1-Cre^{+}$ ) were transplanted into lethally irradiated WT congenic (CD45.1<sup>+</sup>) recipient mice. Recipient mice were treated with plpC at 6 weeks post-transplant. (B) Complete blood counts of PB samples from recipient mice

before (–1 week) and up to 24 weeks post-plpC. **(C)** Flow cytometric analysis of PB samples was performed before and after plpC to determine absolute counts of lymphoid (B-cells or T-cells) and myeloid (Neutrophils or Monocytes) cells. For **B-C**, N=28-30 recipient mice per genotype pooled from two independent experiments. **(D)** Absolute cell counts of BM HSPC populations in recipient mice were determined by flow cytometric analysis at 24 weeks post-plpC. N=5-8 recipient mice per genotype pooled from two independent experiments. See also **Supplementary Fig. 2**. Results represent the mean  $\pm$  standard error of the mean (SEM) (**B-C**) or mean  $\pm$  SD (**D**). A mixed effects analysis with repeated measures and Tukey multiple comparison correction (**B-C**) or one-way ANOVA with Tukey multiple comparison correction (**D**) were used for the comparison of groups. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ . ns, not significant (or labeled if  $P < 0.10$ ). Symbols ( $U2af1^{+/+}$  vs  $U2af1^{S34F/+}$  [\*];  $U2af1^{+/+}$  vs  $U2af1^{Q157R/+}$  [#]) are used to differentiate comparisons in **B-C**.

Figure 3

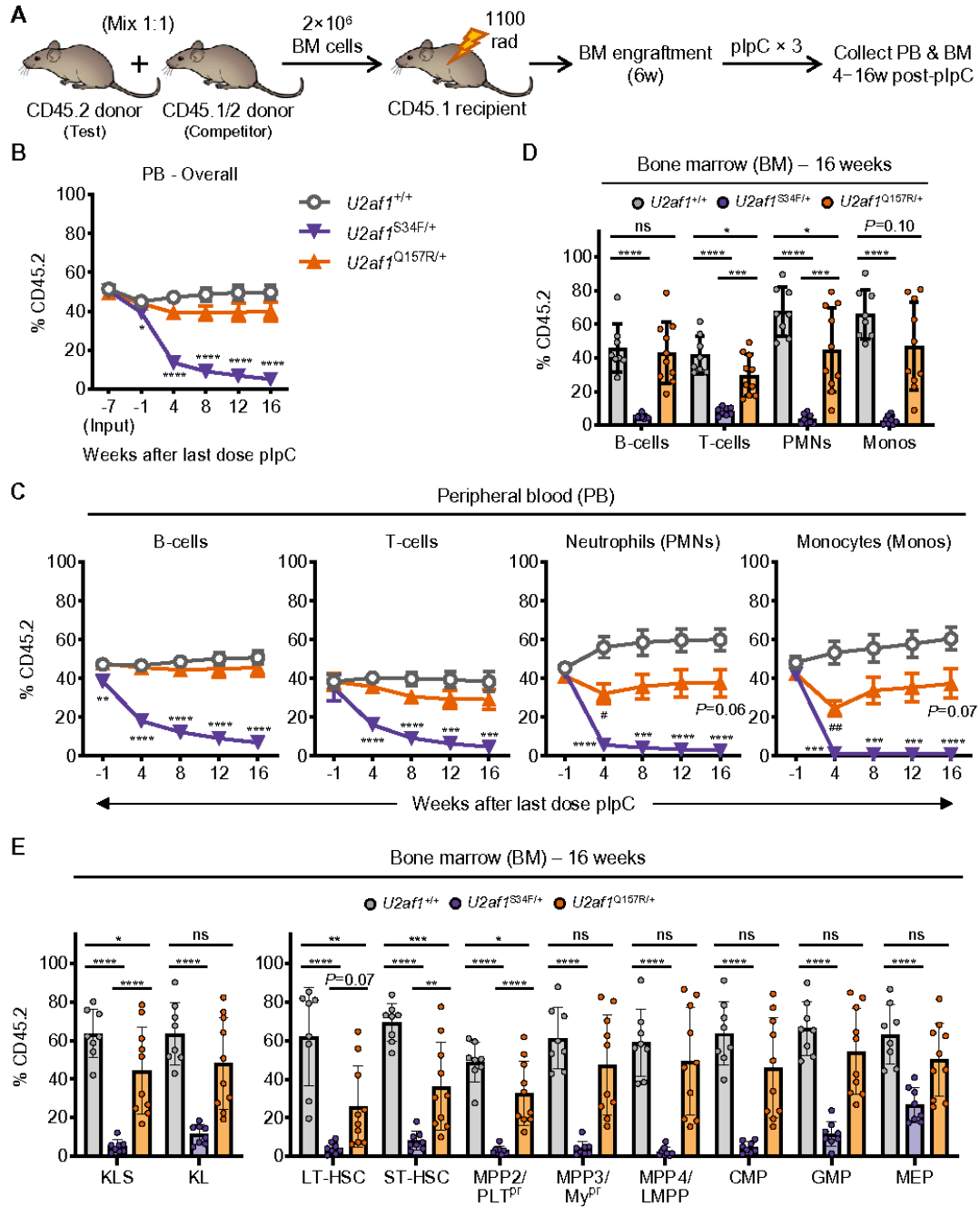


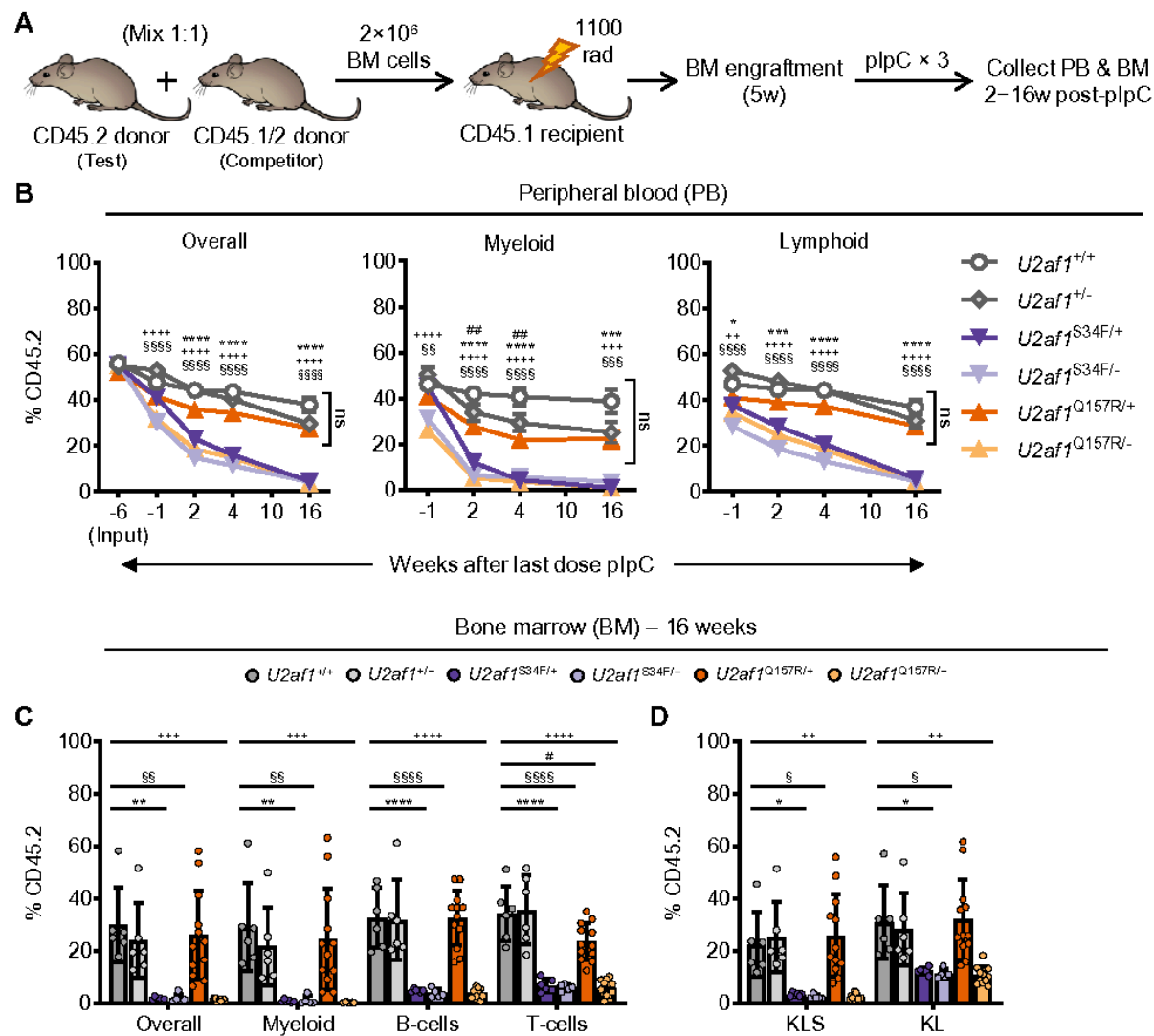
Figure 3

***U2af1*<sup>S34F/+</sup> HSCs are significantly more impaired than *U2af1*<sup>Q157R/+</sup> HSCs in BM repopulation assays.**

(A) Competitive transplant assay design. CD45.2<sup>+</sup> (test) donor BM cells from *U2af1*<sup>+/+</sup>, *U2af1*<sup>S34F/+</sup>, or *U2af1*<sup>Q157R/+</sup> mice (all *Mx1-Cre*<sup>+</sup>) were each mixed 1:1 with CD45.1<sup>+</sup>/CD45.2<sup>+</sup> competitor BM cells and transplanted into lethally irradiated WT congenic (CD45.1<sup>+</sup>) recipient mice. Recipient mice were treated

with plpC at 6 weeks post-transplant. **(B-C)** Donor cell chimerism (CD45.2<sup>+</sup>) was assessed on PB from recipient mice before (-1 week) and up to 16 weeks post-plpC. Input (-7 weeks) refers to the 1:1 BM cell mixtures transplanted into recipient mice. **(B)** Overall chimerism of PB leukocytes. **(C)** Chimerism of lymphoid (B-cells or T-cells) and myeloid (Neutrophils or Monocytes) cell populations. **(D-E)** Donor cell chimerism (CD45.2<sup>+</sup>) was assessed on BM from recipient mice at 16 weeks post-plpC. **(D)** Chimerism of lymphoid (B-cells or T-cells) and myeloid (PMNs or Monos) cell populations. **(E)** Chimerism of HSPC populations. N=8-10 recipient mice per genotype pooled from two independent experiments **(B-E)**. See also **Supplementary Fig. 3**. Results represent the mean  $\pm$  SEM **(B-C)** or mean  $\pm$  SD **(D-E)**. A two-way ANOVA with repeated measures and Tukey multiple comparison correction **(B-C)** or one-way ANOVA with Tukey multiple comparison correction **(D-E)** were used for the comparison of groups. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ . ns, not significant (or labeled if  $P < 0.10$ ). Symbols (*U2af1*<sup>+/+</sup> vs *U2af1*<sup>S34F/+</sup> [\*]; *U2af1*<sup>+/+</sup> vs *U2af1*<sup>Q157R/+</sup> [#]) are used to differentiate comparisons in **B-C**.

Figure 4



## Figure 4

**Hemizygous *U2af1*<sup>Q157R/-</sup> and *U2af1*<sup>S34F/-</sup> HSCs are severely impaired in BM repopulation assays. (A)** Competitive transplant assay design. CD45.2<sup>+</sup> (test) donor BM cells from *U2af1*<sup>+/+</sup>, *U2af1*<sup>+/-</sup>, *U2af1*<sup>S34F/+</sup>, *U2af1*<sup>Q157R/+</sup>, *U2af1*<sup>S34F/-</sup>, or *U2af1*<sup>Q157R/-</sup> mice (all *Mx1-Cre*<sup>+</sup>) were each mixed 1:1 with CD45.1<sup>+</sup>/CD45.2<sup>+</sup> competitor BM cells and transplanted into lethally irradiated WT congenic (CD45.1<sup>+</sup>) recipient mice. Recipient mice were treated with plpC at 5 weeks post-transplant. **(B)** Donor cell chimerism (CD45.2<sup>+</sup>) was assessed on PB from recipient mice before (-1 week) and up to 16 weeks post-plpC. Input (-6 weeks) refers to the 1:1 BM cell mixtures transplanted into recipient mice. Overall, Myeloid (CD11b<sup>+</sup> cells), and Lymphoid (B-cells and T-cells) PB chimerism are shown. **(C-D)** Donor cell chimerism (CD45.2<sup>+</sup>) was assessed on BM from recipient mice at 16 weeks post-plpC. **(C)** Chimerism of Myeloid (CD11b<sup>+</sup> cells) and lymphoid (B-cells or T-cells) cell populations. **(D)** Chimerism of HSPC populations. For **B-D**, Data are from a single experiment in which a pool of competitor BM cells (N=3 donors) was individually mixed with test BM cells from N=15 different donors (N=2-4 per genotype) prior to transplant into N=80 recipients (N=5-8 recipient mice per BM cell mixture and N=10-20 total recipient mice per genotype group). BM analysis was performed on a subset (N=6-12 randomized mice) of each genotype group. Data from one *U2af1*<sup>S34F/-</sup> mouse was identified as a significant outlier (Grubb's test, *P* < 0.05) and removed from final analysis. See also **Supplementary Fig. 4**. Results represent the mean ± SEM **(B)** or mean ± SD **(C)**. A two-way ANOVA with repeated measures and Tukey multiple comparison correction **(B)** or one-way ANOVA with Tukey multiple comparison correction **(C-D)** were used for the comparison of groups. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001. ns, not significant (or labeled if *P* < 0.10). Symbols (*U2af1*<sup>+/+</sup> vs *U2af1*<sup>S34F/+</sup> [\*]; *U2af1*<sup>+/+</sup> vs *U2af1*<sup>Q157R/+</sup> [#]; *U2af1*<sup>+/+</sup> vs *U2af1*<sup>S34F/-</sup> [§]; *U2af1*<sup>+/+</sup> vs *U2af1*<sup>Q157R/-</sup> [+]) are used to differentiate comparisons in **B-D**.

Figure 5

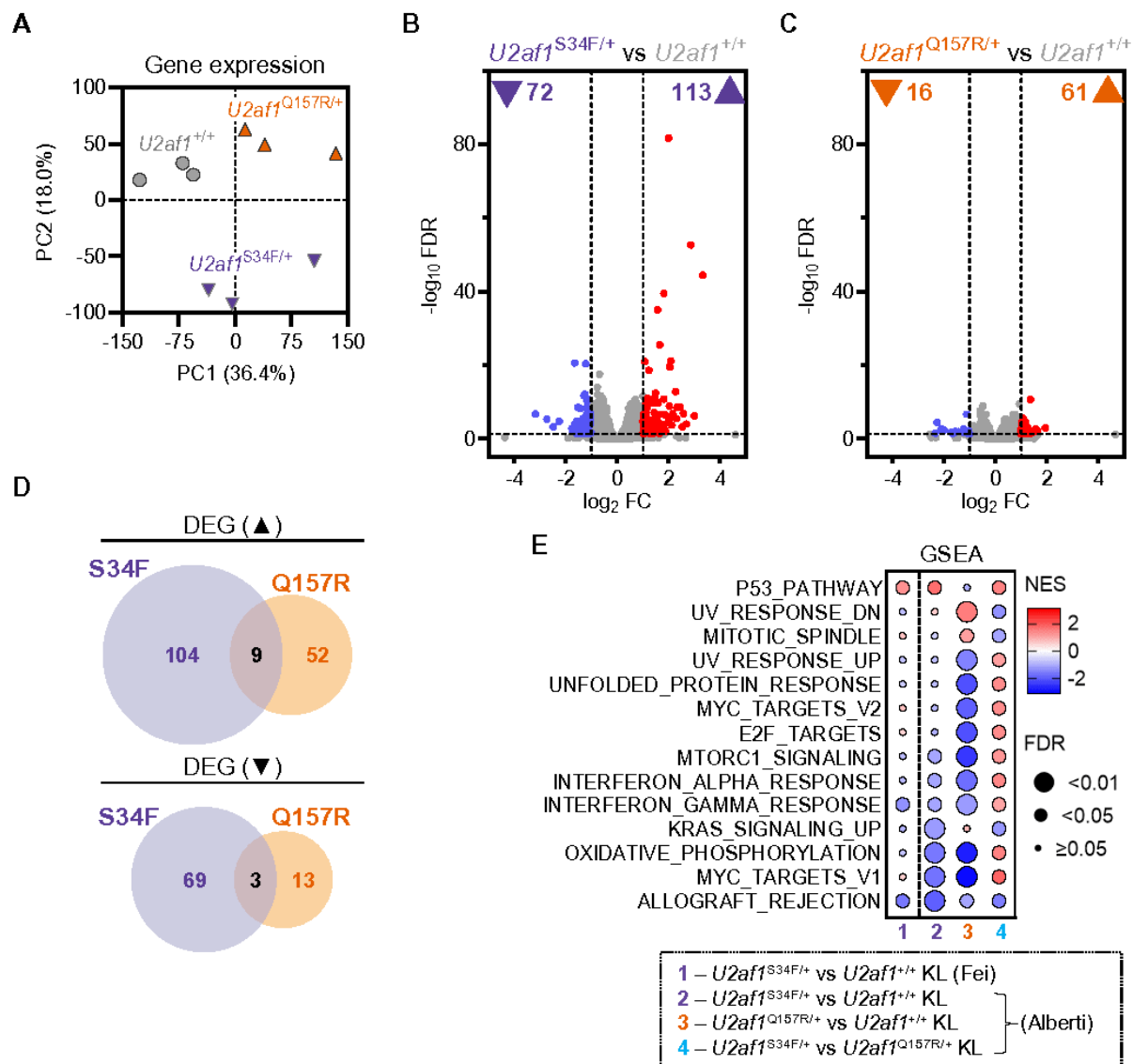


Figure 5

**U2AF1<sup>S34F</sup> and U2AF1<sup>Q157R</sup> induce distinct gene expression changes in myeloid progenitor cells. (A-E)** Assessment of differential gene expression by RNA-seq in BM KL cells from *U2af1*<sup>+/+</sup>, *U2af1*<sup>S34F/+</sup>, and *U2af1*<sup>Q157R/+</sup> mice under native hematopoiesis conditions (as in **Figure 1D**). N=3 KL cell samples per genotype. **(A)** Unsupervised principal component (PC) analysis of gene expression levels in KL cells. **(B-C)** Volcano plot of differentially expressed genes (DEG; FDR<0.05 and |log<sub>2</sub> FC|>1 vs *U2af1*<sup>+/+</sup>) in KL cells from *U2af1*<sup>S34F/+</sup> **(B)** or *U2af1*<sup>Q157R/+</sup> **(C)** mice. The numbers of up- (▲) and down- (▼) regulated DEG are listed. **(D)** Overlap of upregulated (top) and downregulated (bottom) DEG in *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> KL cells. **(E)** Gene set enrichment analysis (GSEA) for Hallmark gene sets that were significantly enriched (FDR<0.05) in *U2af1*<sup>S34F/+</sup> vs *U2af1*<sup>Q157R/+</sup> KL cells (column 4). Normalized enrichment scores (NES) for

*U2af1*<sup>S34F/+</sup> vs *U2af1*<sup>+/+</sup> (column 2) and *U2af1*<sup>Q157R/+</sup> vs *U2af1*<sup>+/+</sup> (column 3) KL cells are also shown. Reanalyzed RNA-seq data (GSE112174) from *U2af1*<sup>+/+</sup> and *U2af1*<sup>S34F/+</sup> KL cells under native hematopoiesis conditions in Fei *et al.*<sup>17</sup> (column 1) is also included. Circle color indicates the NES score for each term and size is proportional to the magnitude of the FDR (q-value).

Figure 6

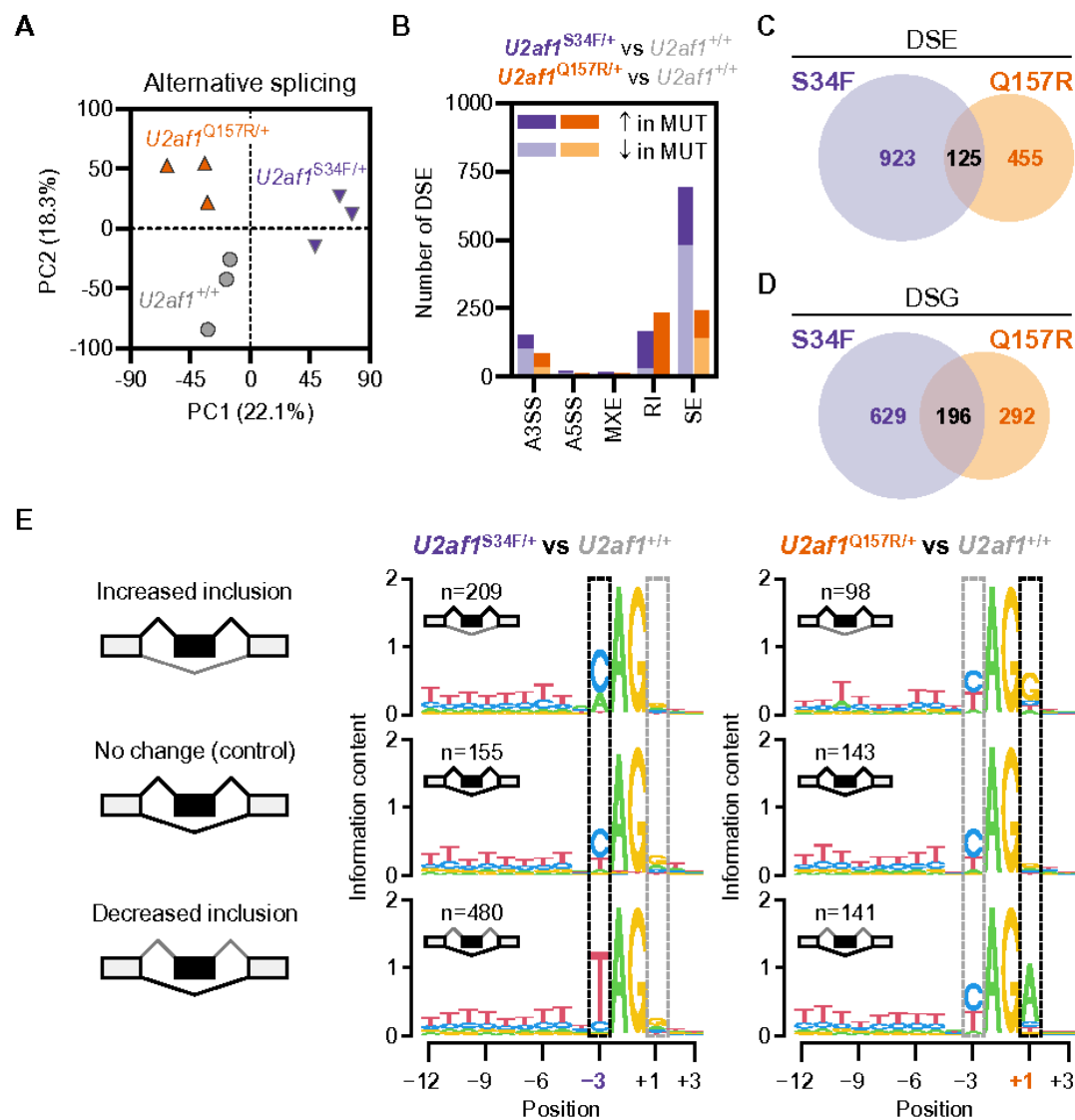


Figure 6

**U2AF1<sup>S34F</sup> and U2AF1<sup>Q157R</sup> induce distinct alternative pre-mRNA splicing changes in myeloid progenitor cells.** (A-E) Assessment of differential alternative pre-mRNA splicing by RNA-seq in BM KL cells from *U2af1*<sup>+/+</sup>, *U2af1*<sup>S34F/+</sup>, and *U2af1*<sup>Q157R/+</sup> mice under native hematopoiesis conditions (Fig. 1D). N=3 KL

cell samples per genotype. **(A)** Unsupervised principal component (PC) analysis of exon-inclusion ratios (referred to as 'percent spliced-in' or 'PSI') for all annotated alternative splicing events in KL cells. **(B)** Number and type (alternative 3' or 5' splice sites [A3SS, A5SS], mutually exclusive exons [MXE], retained introns [RI], and skipped exons [SE]) of differentially spliced events (DSE; FDR<0.05 and  $|\Delta\text{PSI}|>0.05$  vs *U2af1*<sup>+/+</sup>) in KL cells from *U2af1*<sup>S34F/+</sup> (left bars) or *U2af1*<sup>Q157R/+</sup> (right bars) mice. **(C)** Overlap of DSE in *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> KL cells. **(D)** Overlap of differentially spliced genes (DSG) in *U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> KL cells. DSE from **C** were converted to DSG for analysis. **(E)** Analysis of consensus 3' splice site (3'SS) sequences from control (i.e., no change in mutant vs *U2af1*<sup>+/+</sup>) and differentially spliced SE events in *U2af1*<sup>S34F/+</sup> (middle) or *U2af1*<sup>Q157R/+</sup> (right) KL cells. The highlighted -3 and +1 positions of the 3'SS recapitulate the aberrant consensus 3'SS sequence dependencies identified previously in *U2AF1*<sup>S34F</sup> and *U2AF1*<sup>Q157R</sup> MDS patients. See also **Supplementary Fig. 6**.

Figure 7

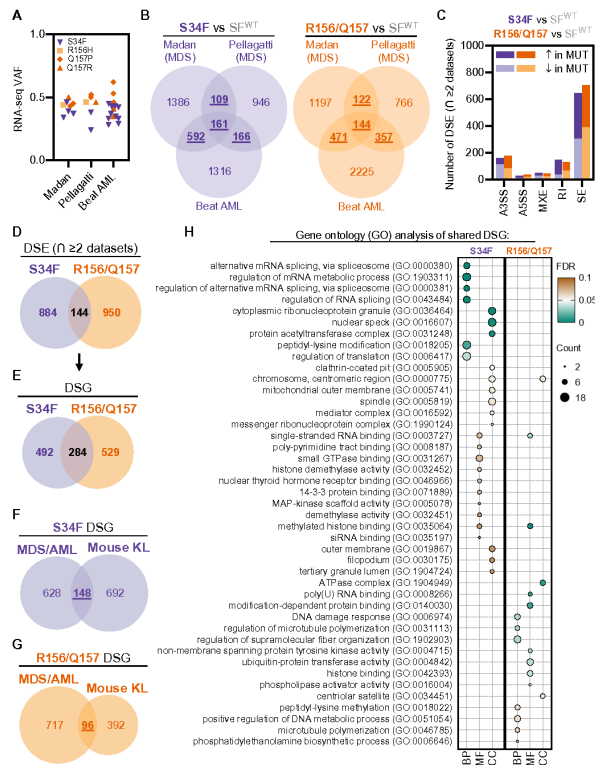


Figure 7 cont.

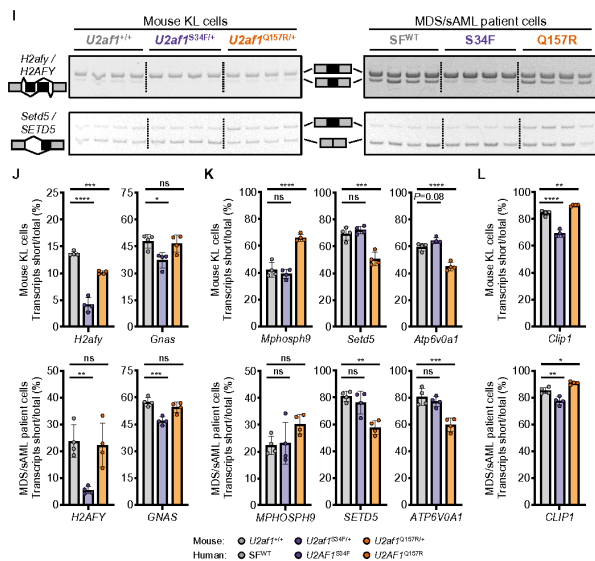


Figure 7

***U2af1*<sup>S34F/+</sup> and *U2af1*<sup>Q157R/+</sup> mouse models recapitulate alternative pre-mRNA splicing changes found in MDS and AML patients.** (A-E) Assessment of differential alternative pre-mRNA splicing in BM cells from splicing factor WT [SF<sup>WT</sup>] MDS and AML patients and those harboring *U2AF1*<sup>S34F</sup> (S34F) or *U2AF1*<sup>R156H/Q157(P/R)</sup> (R156/Q157) mutations in three publicly available RNA-seq datasets (Madan *et*

*al.*,<sup>11</sup> Pellagatti *et al.*,<sup>9</sup> and Beat AML<sup>24</sup>). RNA-seq data (GSE128429, GSE114922, and phs001657.v1.p1) were reanalyzed for this study. N=2-10 samples per mutant genotype per study. N=8 (Madan), 40 (Pellagatti), or 279 (Beat AML) SF<sup>WT</sup> samples. **(A)** BM cell variant allele frequencies (VAF) of S34F, R156H, and Q157(P/R) mutations in *U2AF1* mRNA from MDS and AML patients harboring *U2AF1* mutations in Madan, Pellagatti, and Beat AML. **(B)** Total number and intersection of differentially spliced events (DSE; FDR<0.05 and |ΔPSI|>0.05 vs SF<sup>WT</sup> patients) in BM cells from MDS and AML patients harboring S34F or R156/Q157 mutations in Madan, Pellagatti, and Beat AML. See also **Supplementary Fig. 7**. DSE shared (∩) between at least two datasets are underlined and bolded. **(C)** Number and type (alternative 3' or 5' splice sites [A3SS, A5SS], mutually exclusive exons [MXE], retained introns [RI], and skipped exons [SE]) of DSE (∩≥2 MDS/AML datasets) in BM cells from patients harboring S34F (left bars) or R156/Q157 (right bars) mutations. **(D)** Overlap of DSE (∩≥2 MDS/AML datasets) in BM cells from MDS and AML patients harboring S34F or R156/Q157 mutations. **(E)** Overlap of differentially spliced genes (DSG) in BM cells from MDS and AML patients harboring S34F or R156/Q157 mutations. DSE from **D** were converted to DSG for analysis. **(F-G)** Overlap of DSG from **E** (MDS-AML) with DSG from **Fig. 6D** (Mouse KL) for S34F (**F**) or R156/Q157 (**G**) mutations. **(H)** GO analysis of S34F (left) and R156/Q157 (right) shared DSGs from **F-G**. Circle size is proportional to the gene count for each term and the color indicates the magnitude of the FDR (q-value). REVIGO was used to consolidate 51 (S34F) or 40 (R156/Q157) gene sets into a representative subset of GO terms.<sup>38</sup> All significant GO terms are listed in **Supplementary Table 17**. **(I-L)** RT-PCR orthogonal confirmation of S34F or Q157R aberrantly spliced transcripts in mutant mouse KL and MDS/s-AML patient cells. **(I)** Representative RT-PCR/polyacrylamide gel results for *H2afy/H2AFY* (aberrantly spliced by S34F, left) and *Setd5/SETD5* (aberrantly spliced by Q157R, right) prior to gel densitometry quantification. N=4 samples per genotype. **(J-L)** Quantification of aberrantly spliced transcripts in S34F (**J**), Q157P/R (**K**), or both (**L**). Results represent the mean ± SD (**J-L**). A one-way ANOVA with Tukey multiple comparison correction (**J-L**) was used for the comparison of groups. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001. ns, not significant (or labeled if *P* < 0.10).

Figure 8

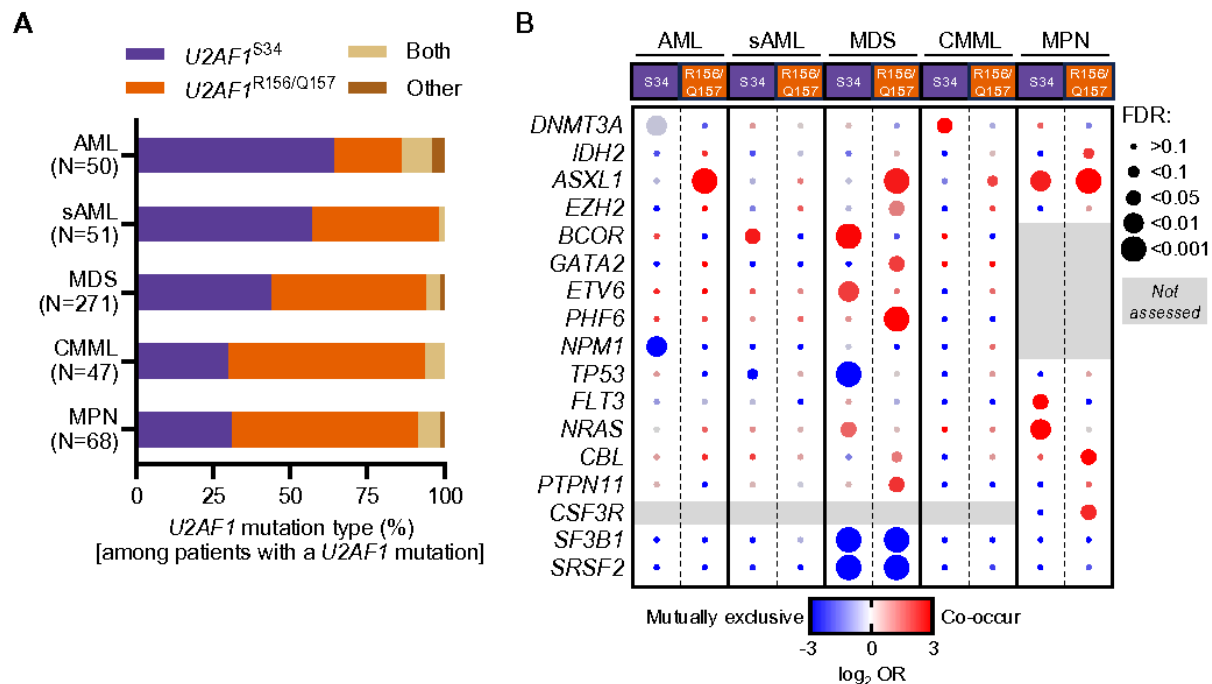


Figure 8

**Frequency of  $U2AF1^{S34}$  and  $U2AF1^{R156/Q157}$  hotspot mutations and co-occurrence with other gene mutations differ in myeloid malignancies.** (A) Frequency of  $U2AF1$  hotspot mutations in myeloid malignancy patients. Patients with a  $U2AF1$  mutation(s) (i.e., S34[F/Y], R156H/Q157[P/R], both S34 and R156/Q157, or 'other' rare variants) and a diagnosis of AML (N= 50 patients), sAML (from MDS; N=51 patients), MDS (N=271 patients), CMML (N=47 patients), and MPN (N=68 patients), were identified from 21 published studies (see **Supplementary Methods** and **Supplementary Table 18**). (B) Analysis of  $U2AF1$  hotspot mutation co-occurrence and mutual exclusivity in myeloid malignancies. Mutation data for patients with a diagnosis of AML (N=1857 patients), sAML (from MDS; N=458 patients), MDS (N=3159 patients), CMML (N=430 patients), and MPN (N=1551 patients) were included from 20 published studies that performed  $U2AF1$  sequencing and had patient-level mutation data available for a common set of 23 (MPN) or 31 (AML, sAML, MDS, and CMML) genes sequenced across all studies (see **Supplementary Methods** and **Supplementary Table 19**). cBioPortal was used for the co-occurrence and mutual exclusivity of genomic alteration analysis within each disease group using the default settings. Genes with significant interactions (FDR<0.1) with  $U2AF1$  are shown (for complete analysis see **Supplementary Fig. 8** and **Supplementary Table 20**). Circle color indicates the  $\log_2$  odds ratio (OR) for each gene pair and size is proportional to the magnitude of the FDR (q-value).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationv02.pdf](#)
- [SupplementaryTables120.zip](#)