***Title Page**

Accepted manuscript. DOI: 10.1016/j.exphem.2018.11.001

Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal.

Jane Jialu Xu[1,2], Monique F Smeets[1,2], Shuh Ying Tan[1,2,3], Meaghan Wall[1,2,4], Louise E Purton[1,2] & Carl R Walkley[1,2,5,*].

[1]St. Vincent's Institute, Fitzroy, VIC, 3065 Australia.
[2]Department of Medicine, St. Vincent's Hospital, University of Melbourne, Fitzroy, VIC, 3065 Australia.
[3]Department of Hematology, St. Vincent's Hospital, Fitzroy 3065, VIC, Australia.
[4]Victorian Cancer Cytogenetics Service, St. Vincent's Hospital, Fitzroy, VIC, 3065 Australia.
[5]Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, VIC, 3000 Australia.

*Correspondence should be addressed to:
Dr. Carl Walkley, B.Pharm(Hons), PhD (ORCID: 0000-0002-4784-9031)
St Vincent's Institute
9 Princes St
Fitzroy 3065 VIC
Australia
T: 61 3 9231 2480
F: 61 3 9416 2676
Email: cwalkley@svi.edu.au

Conflict of Interest Statement: All authors declare no conflicts of interest.

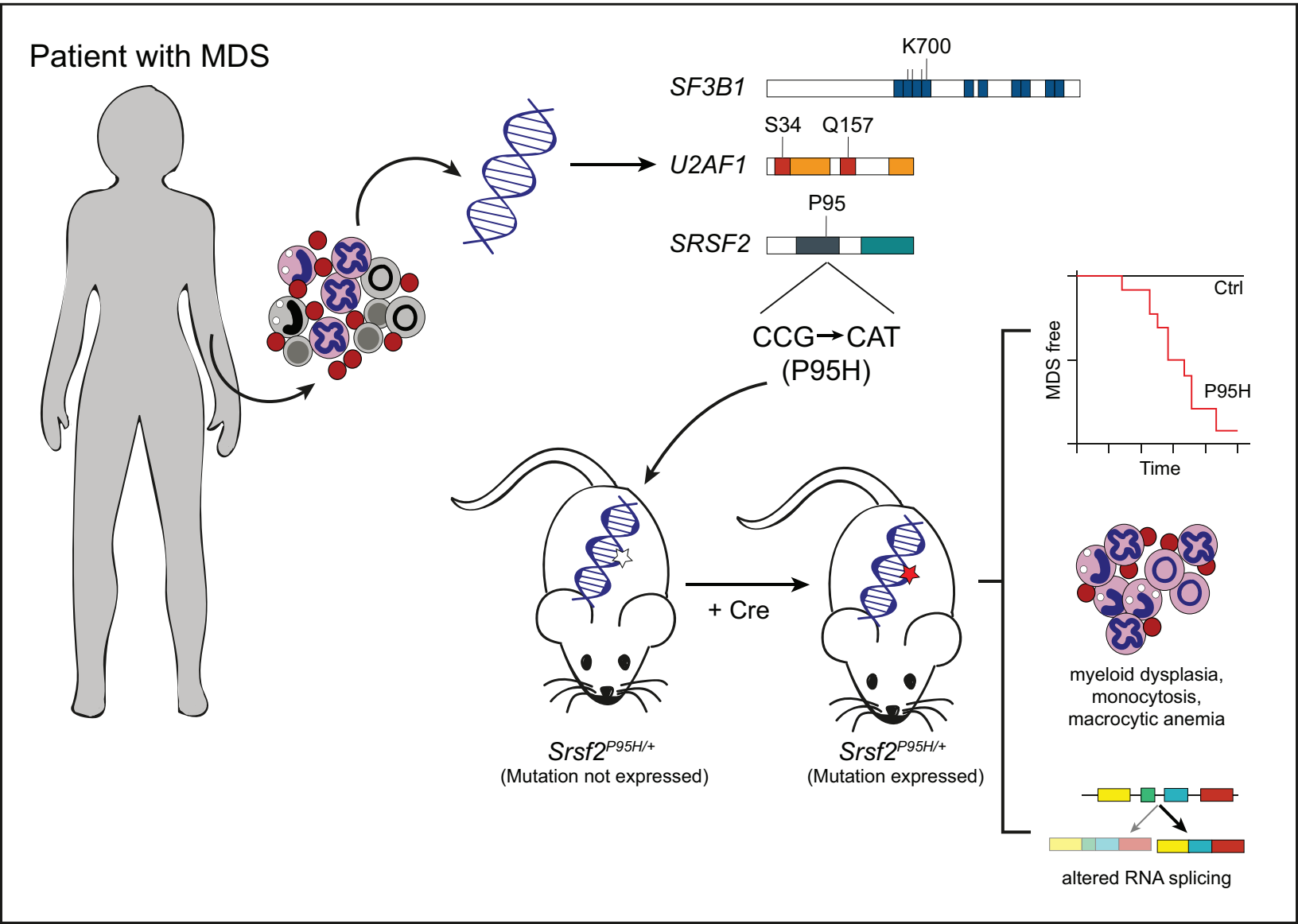
Running Title/Header: **Mouse models of RNA splicing mutations**

Category for the Table of Contents: **Malignant Hematopoiesis**

Key words: **mouse model; spliceosome; myelodysplastic syndrome; myeloproliferative disease; RNA splicing**

Metrics:
Abstract Word Count = 187
Word Count (excluding references, figure legends and tables) = 6958
Figures/Tables = 3 figures/ 1 Table

Graphical Abstract (for review)



Highlights

- Comparison of the murine knock-in models of human *SF3B1*, *U2AF1* and *SRSF2* mutations.
- The targeting strategy can have a significant impact of fidelity of the mouse model.
- Inverted exon strategies do not result in transcript heterozygosity in 3 separate genes/models.

Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal.

Jane Jialu Xu[1,2], Monique F Smeets[1,2], Shuh Ying Tan[1,2,3], Meaghan Wall[1,2,4], Louise E Purton[1,2] & Carl R Walkley[1,2,5,*].

[1]St. Vincent's Institute, Fitzroy, VIC, 3065 Australia.
[2]Department of Medicine, St. Vincent's Hospital, University of Melbourne, Fitzroy, VIC, 3065 Australia.
[3]Department of Hematology, St. Vincent's Hospital, Fitzroy 3065, VIC, Australia.
[4]Victorian Cancer Cytogenetics Service, St. Vincent's Hospital, Fitzroy, VIC, 3065 Australia.
[5]Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, VIC, 3000 Australia.

*Correspondence should be addressed to:
Dr. Carl Walkley, B.Pharm(Hons), PhD (ORCID: 0000-0002-4784-9031)
St Vincent's Institute
9 Princes St
Fitzroy 3065 VIC
Australia
T: 61 3 9231 2480
F: 61 3 9416 2676
Email: cwalkley@svi.edu.au

Conflict of Interest Statement: All authors declare no conflicts of interest.

Running Title/Header: **Mouse models of RNA splicing mutations**

Category for the Table of Contents: **Malignant Hematopoiesis**

Key words: **mouse model; spliceosome; myelodysplastic syndrome; myeloproliferative disease; RNA splicing**

Metrics:
Abstract Word Count = 187
Word Count (excluding references, figure legends and tables) = 6958
Figures/Tables = 3 figures/ 1 Table

Abstract

Myelodysplastic syndromes (MDS) and related myelodysplastic/myeloproliferative neoplasms (MDS/MPN) are clonal stem cell disorders, primarily affecting patients over 65 years of age. Mapping of the MDS and MDS/MPN genome identified recurrent heterozygous mutations in the RNA splicing machinery, with *SF3B1*, *SRSF2* and *U2AF1* frequently mutated. To better understand how spliceosomal mutations contribute to MDS pathogenesis *in vivo*, numerous groups have sought to establish conditional murine models of *SF3B1*, *SRSF2* and *U2AF1* mutations. The high degree of conservation of hematopoiesis between mice and human and the well-established phenotyping and genetic modification approaches make murine models an effective tool to study how a gene mutation contributes to disease pathogenesis. The murine models of spliceosomal mutations described to date recapitulate human MDS or MDS/MPN to varying extents. Reasons for the differences in phenotypes reported between alleles of the same mutation are varied, but the nature of the genetic modification itself and subsequent analysis methods are important to consider. In this review, we will summarize recently reported murine models of *SF3B1*, *SRSF2* and *U2AF1* mutations, with a particular focus on the genetically engineered modifications underlying the models and the experimental approaches applied.

Introduction

Myelodysplastic syndromes (MDS), myelodysplastic/myeloproliferative neoplasms (MDS/MPN) and related disorders are a heterogeneous class of blood cancers leading to ineffective hematopoiesis in the bone marrow (BM) [1, 2]. Approximately 30% of MDS patients progress to acute leukemia. Median survival ranges from 97 months for low risk categories, down to 11 months for high risk MDS [2]. The incidence of MDS in the general population is ~4-5 per 100,000 people, but this increases with age [1]. Population based studies in both Australia and the USA indicate a significant underestimation of the true burden of MDS, with frequencies estimated at 103 per 100,000 and between 75-162 per 100,000 respectively over age 65 [3-5]. A feature of MDS and MDS/MPN is the progressive establishment of clonal hematopoiesis, where mutant cells dominate the BM at the expense of normal hematopoiesis [6]. The mechanisms behind this clonal advantage are unknown, but the identification of clonal hematopoiesis of indeterminate potential (CHIP) suggests that the clonal advantage is acquired progressively over many years following the establishment of an initiating mutation [7, 8]. In general, MDS patients have limited treatment options. Current treatment strategies for MDS and MDS/MPN are mostly heterogeneous and modestly efficacious and are not associated with durable responses [9]. The dismal prognosis following hypomethylating agent failure highlights the urgent need for new treatment approaches.

Until recently, there were few described recurrent mutations or familial syndromes that could provide insight into the genetics of these cancers and be used to develop preclinical models. Now, however, the detailed mutational architecture of human MDS, chronic myelomonocytic leukemia (CMML) and other related forms of MDS/MPN has been defined. There are recurrent mutations in key pathways, including the RNA splicing machinery (e.g. *SF3B1*, *SRSF2, U2AF1*) [10-12], transcription factors (e.g. *RUNX1*, *BCOR*, *ETV6*) and epigenetic enzymes (e.g. *DNMT3A*, *TET2, EZH2*, *ASXL1*) [6]. Additional studies have deduced clonal evolution and mutational order, highlighting distinct patterns [13]. Paralleling studies of diseased samples, analysis of healthy elderly populations has demonstrated that mutations associated with MDS and leukemia accumulate and are positively selected for in the hematopoietic cells from otherwise healthy people ≥70 years of age [7, 8, 14].

The identification of recurrent mutations in the RNA splicing machinery was a breakthrough in understanding the genetics of MDS and related disorders [10-12]. RNA splicing is a highly coordinated and essential process carried out by major and minor spliceosomes to remove non-coding regions (introns) of the pre-mRNA before protein translation [15, 16]. The major spliceosome consists of five small nuclear ribonucleoprotein complexes (snRNPs), U1, U2, U4, U5 and U6. The minor spliceosome includes the U5

snRNPs and other functional snRNPs mirroring the major spliceosome. Mechanistically, RNA splicing occurs with the initial recognition of the 5' and 3' splice sites by the U1 and U2 complexes respectively, followed by the excision of the intron and exon ligation by U4, U5 and U6 complexes. *Trans*-acting Splicing factors, such as serine-arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs), bind to the regulatory elements located in the exons and introns to enhance or repress the splicing activity and contribute to alternative splicing (Figure 1). The regulatory elements include exonic splicing enhancers and silencers (ESEs and ESSs) and intronic splicing enhancers and silencers (ISEs and ISSs) [16].

In MDS, >80% of patients have a mutation in a spliceosome gene [10-12]. In phenotypically overlapping syndromes of MDS/MPN, such as CMML, spliceosome mutations are also common [10-12, 17]. Mutations occur most frequently in *SF3B1* and *SRSF2*, with a lower prevalence in other spliceosomal genes including *U2AF1*, *ZRSR2* and *U2AF2* [17, 18]. Spliceosomal mutations are thought to arise early in the course of disease, including as founder/initiating events [6]. How mutations in the spliceosome alter normal hematopoiesis and contribute to disease pathogenesis remains unclear. A key experimental system to understand the impact of the spliceosome mutations in regulating both normal and malignant hematopoiesis is the generation of high fidelity murine models. There is a very high conservation of the human and mouse core spliceosome proteins, including those mutated in MDS and CMML. For example, human SRSF2 and murine Srsf2 proteins are 100% sequence conserved.

The generation of conventional null alleles of RNA splicing genes in the mouse has been useful in identifying key RNA substrates whose splicing is perturbed in their absence and in beginning to understand their normal physiological function [19-22]. Retroviral overexpression has been used to assess how the mutations present in patients may contribute to MDS development [10, 22]. A caveat with both the loss of function alleles and retroviral overexpression of mutant cDNAs is that they fail to recapitulate the mutations as they occur in humans: heterozygous point mutations expressed from their endogenous locus, and not alleles resulting in protein deficiency. The recurrent finding that the spliceosome mutations are heterozygous indicates that retention of a wild-type copy is necessary for cells to survive, an interpretation supported by both genetic and pharmacologic evidence [23-26], and additionally that gene dosage and protein complex stoichiometry may be important in the pathogenicity of the mutations [27].

Over the last 3 years, murine models recapitulating point mutations identified in patients with MDS and CMML have been described for *Sf3b1* [28, 29], *Srsf2* [30-32] and *U2af1* [33, 34]. The initial *U2af1* mutation model reported was an inducible cDNA expressed from a heterologous locus but will be discussed alongside the true knock-in alleles [33]. The

phenotypes observed in these “humanized” models have ranged from very mild through to the development of MDS/MPN. The reasons for these variable phenotypes have been not been systematically outlined, however the targeting strategy itself and the analysis methods may contribute significantly to the understanding of these models. Here we will discuss in detail the murine models of these mutations developed to date, their phenotypes and how they compare to the phenotypes of humans with these mutations.

Murine models of RNA spliceosome mutations

U2AF1 (previously known as *U2AF35*)

U2 small nuclear RNA auxiliary factor 1 (U2AF1) binds to the AG nucleotides and polypyrimidine tract located near the 3' splice site in the intronic sequence (Figure 1). It then recruits U2AF2 and the U2 major spliceosome complex to facilitate 3' splicing site recognition [15, 35]. U2AF1 mutations have been identified in lung, breast, colon and various epithelial carcinomas [36]. In MDS, approximately 11% of patients have a *U2AF1* mutation, predominantly at serine 34 (S34F/S34Y) and glutamine 157 (Q157R/Q157R) [10, 37]. One transgenic and two knock-in models have been developed to date.

U2AF1^{S34F} transgenic model

Shirai et al., *U2AF1^{S34F}* Tet-on transgenic model

Shirai et al. [33], reported the generation of a transgenic model of *U2AF1* mutation. In this case, unlike the others that will be discussed, an inducible (tetO) human *U2AF1^{S34F}* or control *U2AF1^{WT}* cDNA were targeted to the Collagen1a1 (*Col1a1*) locus of *Rosa26*-M2rtTA ES cells [38]. Expression of the WT or S34F cDNAs was not regulated in the manner of the endogenous *U2af1* locus. To activate the expression of the *U2AF1^{S34F}* or *U2AF1^{WT}* cDNAs, animals were administered doxycycline, enabling dose dependent induction of expression of the human *U2AF1* cDNAs in the presence of two endogenous wild-type gene copies derived from the mouse (Figure 2).

No analysis was reported of the effects of *U2AF1^{S34F}* on native hematopoiesis (i.e. in animals treated with doxycycline in the absence of prior bone-marrow transplantation). The only data presented was from irradiated transplant recipients reconstituted with whole bone marrow of *U2AF1^{S34F}* M2rtTA or *U2AF1^{WT}* M2rtTA donors. Once reconstitution was established, the recipient animals were administered doxycycline [33]. A similar approach was chosen in a second report of this model [26]. The use of BM transplantation prior to gene induction restricts expression of the mutation to the reconstituted hematopoietic cells, of particular importance when using an inducer expressed from the widely expressed *Rosa26* locus or others like *Mx1* that are active in non-hematopoietic cell populations [39-41]. A caveat of analysis in post-transplant settings only is that bone marrow transplantation is known to induce hematopoietic stress, permanently modifying the bone marrow microenvironment [42] and alter the clonal dynamics of HSCs [43].

Using this experimental approach, the authors showed that the doxycycline treated *U2AF1^{S34F}* M2rtTA recipients developed a stable peripheral blood leukopenia, without changes in red blood cell or platelet numbers, for up to 12 months. The leukopenia was the result of reduced B cells and monocytes and was dependent on the continued expression of the mutant *U2AF1*. Within the bone marrow there was evidence of increased cell death,

reduced B cells and monocytes and increased neutrophils. There was an increase in the percentages of common myeloid progenitors and a subtle increase in the percentage of LKS cells. In competitive transplants, *U2AF1*^{S34F} cells were serially transplantable through to tertiary recipients, albeit with a reduced chimerism. With extended aging (up to 500 days post bone marrow transplant), the *U2AF1*^{S34F} expressing mice were not reported to develop dysplasia, MDS or AML.

U2af1^{S34F} *Knock-in (KI) models*

Fei et al. [34], reported the generation of two conditional knock-in models of *U2af1*^{S34F} mutation targeted to the endogenous locus. The S34F mutation is encoded by exon 2 of the murine *U2af1* allele. Both models required Cre protein to induce expression of the mutant allele. Two different knock-in approaches were described, one using an inverted exon 2 (referred to in the publication as *IES34F*) and the second using a minigene (referred to as *MGS34F*) which resulted in significant differences in the expression of the mutant allele (Figure 2). Both were tested as heterozygous alleles, where one allele encoded endogenous wild-type *U2af1* and the second the mutant allele.

Fei et al., *U2af1*^{S34F} "inverted exon" model

For the inverted exon 2 model, a lox2272 flanked inverted S34F encoding exon 2 was placed in intron 2-3 and a conventional LoxP element was placed 5' of the endogenous exon 2. Cre induced recombination of the different LoxP sequences enabled removal of the endogenous exon 2 and replacement by an in-frame exon 2 encoding the S34F mutation. Analysis of mutant allele expression after Cre recombination in MEFs, demonstrated that the mutant mRNA was expressed at ~30%, despite being genomically heterozygous. It is not clear why the *IES34F* allele failed to express the transcript heterozygously, although the authors suggested it related to the efficiency of recombination of the locus. This allele was not further characterized [34].

Fei et al., *U2af1*^{S34F} "minigene" model

The minigene model was examined in more detail [34]. In this model, a LoxP flanked cDNA sequence encoding exons 2-8 followed by a transcriptional stop sequence was placed in the first intron. The S34F mutation was inserted in the endogenous exon 2. In the absence of Cre, the *U2af1* transcript from the targeted allele should be encoded by the endogenous exon 1 together with the cDNA encoding exons 2-8, with the stop cassette preventing read through and expression from the S34F encoding allele. This model was tested in the setting of native hematopoiesis using the broadly expressed *Mx1*-Cre and polyI:polyC (pIpC) induction. The analysis of recombination following pIpC administration indicated that the

minigene, unlike the inverted exon model, was efficiently recombined and truly heterozygous both genomically and transcriptionally. Analysis of the *U2af1*$^{S34F/+}$ animals post-pIpC demonstrated a mild persistent macrocytic anemia and a 2-fold reduction in leukocytes, particularly B cells, in the peripheral blood. Within the bone marrow (BM), total cellularity was preserved, but there were reductions in the percentages of LKS+ and LKS+CD48-CD150+ phenotypic hematopoietic stem cells (HSCs). Lineage distribution of more committed populations within the BM was not described. Morphological analysis of the bone marrow revealed less than 1% of cells displaying dysplastic features and no dysplastic cells were reported in the peripheral blood. No MDS or AML was observed in aged *U2af1*$^{S34F/+}$ mice. The phenotypes described for native hematopoiesis were preserved, albeit exacerbated, by bone marrow transplantation where expression of *U2af1*$^{S34F/+}$ was activated post-transplant, demonstrating a cell-autonomous effect of the mutant allele on hematopoiesis. Interestingly, and consistent across the splicing mutation models reported to date, there was a significantly reduced repopulating potential of U2af1(S34F)-expressing cells, when bone marrow cells from donors previously treated with pIpC competed against WT bone marrow cells.

SF3B1

Splicing factor 3b subunit 1 (SF3B1) is part of the SF3b protein complex that recognizes the branch point adenosine (A) base within the intron and facilitates the binding of U2 snRNP on pre-mRNA (Figure 1). *SF3B1* mutations have been identified in MDS, chronic lymphocytic leukemia (CLL) and uveal melanoma (UM). Mutations cluster between exons 14 to 16, with K700E being the most frequent mutation [10, 11]. In MDS, *SF3B1* mutations occur in >80% of refractory anemia with ring sideroblasts (RARS), a MDS subtype with a favorable outcome [11, 44]. Two independently generated *Sf3b1* conditional knock-in models have been reported and both have utilized the *Mx1*-Cre system to induce the expression of the mutant allele.

Sf3b1*^{K700E} *Knock-in (KI) models

Obeng et al., *Sf3b1*^{K700E} "inverted exon" model

A conditional *Sf3b1*^{K700E} model was reported by Obeng and colleagues [28], generated using an inverted exon strategy (Figure 2). In this model, an exon 15 with the A>G substitution and a WT exon 16 were introduced in an inverted orientation within intron 16-17. Upon Cre-induction, the inverted exons are recombined in the correct orientation in place of the endogenous WT exons through the use of different loxP sequences and, in theory, should be physiologically expressed from the endogenous locus. Whilst the *Sf3b1*^{K700E} allele was successfully inverted post-Cre induction (with pIpC treatment), it did not achieve

heterozygous expression when the RNA levels were assessed. The authors presented results of RNA-seq of lin-cKit+ cells from the bone marrow of three *Sf3b1*$^{K700E/+}$ animals, where the mutant allele frequency was between 27%-32% (see Figure S1A; Obeng et al., [28]). A similar level of sub-heterozygous mutant transcript expression was detected in a second recent report using this allele [24] and the phenotypes reported for these animals need to be considered with this caveat.

Monitoring of native hematopoiesis over 64 weeks of pIpC treated cohorts showed a progressive macrocytic anemia with decreased red blood cell counts from 20 weeks post activation of the K700E mutation. Corresponding with anemia, there was an increased level of plasma erythropoietin in *Sf3b1*$^{K700E/+}$ mice. This anemic phenotype was shown to be cell-intrinsic through use of non-competitive bone marrow transplant assays. There was no change in white blood cell or platelet counts. None of the *Sf3b1*$^{K700E/+}$ mice were reported to develop MDS during the period of monitoring. At 64 weeks post pIpC, analysis of the bone marrow showed no change in cellularity, while there was a significant increase in the number of long-term HSCs (LKS+CD150+CD48-) and a decrease in the GMPs. In the spleen, there was an accumulation of erythroblast populations (Ter119+CD71^{high}) and a decrease in the more mature erythroid population (Ter119+CD71^{low}), indicating a terminal erythroid maturation defect. Interestingly, this effect was only reproduced *in vivo* in young mice (11 weeks post pIpC) under stress (drug-induced hemolysis) or *in vitro* when bone marrow progenitor cells (24 weeks post pIpC) were treated with cytokines to induce erythroid differentiation. Morphological examination of the spleen showed an increased number of erythroid precursors and erythroid dysplasia. However, ring sideroblasts or elevated levels of iron deposition were not observed. There was a severe impairment (<20%) in competitive repopulation capacity of *Sf3b1*$^{K700E/+}$ cells (similar to Mupo et al.,[29]), even though >95% chimerism was achieved in non-competitive transplant experiments.

Mupo et al., *Sf3b1*^{K700E} "minigene" model

Mupo et al. [29], generated a conditionally activatable *Sf3b1*$^{K700E/+}$ allele using a mini-gene approach (Figure 2). Here they engineered a loxP flanked construct containing a splice acceptor site (from an Engrail-2 splice acceptor sequence), exons 12–15, intron 15, exons 15–19 (exons 12-19 were codon optimized sequence), intron 19 and the naturally occurring sequences for exons 20–25 including the 3′ UTR, all this followed by a SV40 polyadenylation signal (SV40 pA). This mini-gene was inserted between exons 11 and 12 of *Sf3b1* and the endogenous exon 15, downstream of the mini-gene, was replaced with a mutated, synthetic exon 15 harboring an A>G mutation encoding K700E. Upon expression of Cre protein from *Mx1*-Cre, induced by administration of pIpC, the mine-gene is deleted allowing the expression of the *Sf3b1* allele containing the mutated exon 15 (*Sf3b1*$^{K700E/+}$).

The allele was assessed in a heterozygous setting, with one mutant allele and one WT allele, mirroring the patient setting. RNA-seq analysis confirmed the 50% expression level of the mutated allele in whole bone marrow and lineage negative (lin-) cells from *Sf3b1*$^{K700E/+}$ mice. Four weeks post pIpC treatment native hematopoiesis was assessed and a decrease in the percentage of phenotypic HSCs (LKS+CD34-Flk2-) was observed, but no change in the percentage of lymphoid-primed multipotent progenitors (L-MPPs), common myeloid progenitors (CMPs), granulo-monocytic progenitors (GMP) or myeloid-erythroid progenitors (MEPs). Bone marrow analysis indicated a mild myeloid bias (Gr-1+Mac-1+) and impaired terminal erythroid differentiation (Ter119+CD71^{low}FSC^{low}). There was no overall survival difference between *Sf3b1*$^{K700E/+}$ mice and WT littermates followed longitudinally for up to 83 weeks. Peripheral blood analysis revealed a progressive normocytic anemia with no change in white blood cell or platelet counts. None of the *Sf3b1*$^{K700E/+}$ mice were reported to develop MDS during monitoring. Morphological analysis of the bone marrow did not find evidence of either dysplasia or ringed sideroblasts (a feature of human *SF3B1* mutant MDS), despite increased iron deposits in bone marrow macrophages. Peripheral blood morphology was not reported. Competitive bone marrow transplantation was performed using both young (2 months) and old (12 months) recipients to check the potential influence of aging recipients on donor cell engraftment. In both groups, *Sf3b1*$^{K700E/+}$ cells showed poor chimerism indicating impaired engraftment/repopulating potential. There was no difference between the young and old recipient groups with neither developing overt symptoms of MDS.

SRSF2

Serine-arginine rich factor 2 (SRSF2) binds to exonic splicing enhancers within the exonic sequence located near the 3' splice site and mediates exon inclusion and exclusion (Figure 1). Unlike *U2AF1* or *SF3B1*, *SRSF2* mutations occur near exclusively in myeloid malignancies and at or encompassing proline 95 (P95), with the most frequent mutation reported a proline to histidine substitution (P95H) [10, 12, 17]. *SRSF2* mutations occur in around 12-15% of MDS [9] and 28-52% of CMML [45-47]. Mutations are associated with poor outcomes, a shortened time to leukemic conversion and a myelo-monocytic bias [48-50]. Three independently generated models of *Srsf2*^{P95H} have now been reported and these will be discussed in greater detail than those of *U2af1* and *Sf3b1* given our own work on this gene.

Srsf2*^{P95H} *Knock-in (KI) models

Kim et al., *Srsf2*^{P95H} "gene-duplication" model

Kim et al. [31], reported the first *Srsf2*^{P95H} conditional knock-in model. The allele described approximated a gene-duplication strategy (Figure 2). The targeting construct inserted a loxP site 87bp upstream of the 5'UTR sequence of the endogenous exon 1. The sequence containing exon 3 (3'UTR) and an introduced SV40 polyadenylation sequence (SV40pA) was placed immediately after the endogenous exon 2 to generate an *Srsf2* allele that would express a wild-type protein. The second loxP was placed after the SV40pA. Downstream of the second loxP the sequence of exon 1 containing the *Srsf2*P95H mutation (CG>AC) and the sequence for exon 2 were introduced into intron 2-3 (Figure 3). Very little data was presented describing the phenotypic effects of this *Srsf2*$^{P95H/+}$ model on native hematopoiesis (i.e not from transplant recipients). The authors state that none of the *Srsf2*$^{P95H/+}$ mice developed AML out to 70 weeks of monitoring, although it is not certain if this refers to *de novo* mutated animals or transplant recipients.

The authors principally reported analysis from non-competitive bone marrow transplant assays for the remaining studies. Subsequent studies with this *Srsf2* allele from the same laboratory have also used this approach [23, 24]. At 18 weeks post pIpC, the *Srsf2*$^{P95H/+}$ recipients developed macrocytic anemia and leukopenia due to decreased B cell numbers, while the platelet numbers remained normal. There was evidence of myeloid (hypolobated and hypogranulated neutrophils) and erythroid dysplasia in the peripheral blood/bone marrow but the estimated percentage of affected cells was not reported. Within the bone marrow of the recipients, Kim et al., reported that a major effect of the *Srsf2*$^{P95H/+}$ mutation was an expansion of the HPC-2 phenotype (LKS+CD150+CD48+) within the hematopoietic stem and progenitor population. The frequencies of the HPC-2 population as reported by Kim et al,. in the control *Mx1*-Cre *Srsf2*$^{+/+}$ animals was 0.0231%, significantly higher than that reported in the index reference for this population of 0.0032±0.0014% in WT mice [51]. The frequency of this population may be impacted by both the BM transplant model and pIpC use. However, the ~10 fold difference in reported frequency is not reconciled, and the representative flow cytometry gating strategy presented by Kim et al., is significantly different to the gating strategy in the original report of this population [51]. The authors reported an increased frequency of cells in S-phase and early apoptosis within the bone marrow. Notably, in a subsequent report from the same laboratory using this *Srsf2*$^{P95H/+}$ allele and the *Mx1*-Cre model, the leukopenia and macrocytic anemia were largely absent (see Supp Fig 1E-F [23]). The inconsistency of the phenotype reported across the publications has not been addressed, but potentially indicates that the phenotype is not fully penetrant. Through analysis of RNA-seq from both human patients and murine samples the authors identified a group of mis-spliced candidate genes, with *EZH2* proposed as a key mis-spliced gene. *SRSF2*^{P95H} was proposed to lead to the inclusion of a premature stop codon in *EZH2*, resulting in a putative loss of function allele. This analysis posited a model

that a single or small number of key mis-spliced genes may account for the phenotypes in the *SRSF2*^{P95H} mutant cells.

Kon et al., *Srsf2*^{P95H} "inverted exon" model

Kon et al. [32], reported an *Srsf2*^{P95H} conditional knock-in model developed using an inverted exon strategy (termed FLEx switch inversion technique; Figure 2). For the targeted allele, the 5'UTR and exon 1 were separated. Exon 1 was flanked by loxP/VloxP sites at the 5' end and VloxP/lox2272 3' to exon 1. An inverted exon 1 containing the P95H mutation was inserted downstream of the VloxP/lox2272 in intro 1, followed by second set of loxP/lox2272 sites. After Cre activation, the inverted exon containing P95H "flips" into the correct orientation and the endogenous exon 1 is excised at the same time (Figure 3). RNA sequencing showed no expression from the modified allele in Cre-negative heterozygous animals, indicating that the floxed allele was likely null when not recombined. This was consistent with the failure to obtain animals that had the non-recombined P95H allele homozygosed even in a Cre-negative background [32]. Why this is the case is unclear at present. The targeting strategy itself may be the most significant contributor in this particular instance. The separation of the 5'UTR from the first coding exon, in effect exonization of the UTR, and the numerous modifications around the coding exon and the sequence insertion within the first intron may have had a significant effect on endogenous gene regulation.

The *Srsf2*^{P95H} conditional knock-in mice were crossed to *Vav1*-Cre mice to assess the effects of the heterozygous knock-in allele on hematopoiesis. Unlike the other described models, *Vav1*-Cre is a constitutive Cre and is active in the hematopoietic cells without the addition of an inducer such as pIpC/IFNα (*Mx1*-Cre) or tamoxifen (CreER models). At around 15 weeks of age, the *Srsf2*$^{P95H/+}$ mice developed macrocytic anemia, which remained stable. After observation out to 90 weeks of age, none of the *Srsf2*$^{P95H/+}$ mice developed MDS or leukemia. A caveat with this model, as was described for the previous models describing inverted exon based targeting methods, was that the mutant allele was expressed at 27-35% (mean 31%) based on RNA-seq analysis of the *Vav1*-Cre *Srsf2*^{P95H} mice.

Upon analysis of the bone marrow under native conditions, there was a reduction in the number of LT-HSC (LKS+CD150+CD48-), MPP (LKS+CD150-CD48-) and HPC-1 (LKS+CD150-CD48+) populations, associated with increased cell cycling. There was no change in the cellularity or lineage distribution of bone marrow and spleen, with the exception of a mild B cell differentiation defect in the bone marrow. There was no evidence of dysplasia reported in the bone marrow under native settings. The authors then performed non-competitive and competitive bone marrow transplant assays of the *Srsf2*$^{P95H/+}$ cells. There was engraftment in the non-competitive transplant settings, however the recipients of competitive bone marrow transplants had a significant reduction in chimerism with either

whole bone marrow or HSCs (LKS+CD34-) as the donor. The impaired competitive transplant potential is a common feature of all of the conditional splicing mutants to date. Extensive work demonstrated an engraftment defect of the *Srsf2*$^{P95H/+}$ HSCs. The *Srsf2*$^{P95H/+}$ recipients developed macrocytic anemia and leukopenia. In the bone marrow, there was a significant reduction in LT-HSC and most of the progenitor populations (MPP, HPC-1, CMP, CLP). In addition, *Srsf2*$^{P95H/+}$ recipients had a significantly higher proportion of cycling and apoptotic HSCs than controls. The lineage distribution was skewed towards myeloid at the expense of B cell differentiation and dysplastic erythroid cells as well as dys-megakaryopoiesis were evident in the bone marrow.

Analysis of RNA splicing using multiple purified hematopoietic populations from both native hematopoiesis and cells isolated from transplant recipients was described. From these analyses the authors could identify mis-splicing of a number of genes that had been identified in human *SRSF2* mutant patient samples, including *Csf3r*, *Gnas*, *Hnrnpa2b1*, and several novel differentially spliced genes such as *Atrx* and *Mllt10* both of which are implicated in hematological malignancies. Of note, despite using numerous approaches the authors could not find evidence to support *Ezh2* as a differentially spliced candidate in the mouse. It was proposed that this may be due to species differences, with human *EZH2* having two CCNG motifs whilst mouse has a single CCNG motif [32].

Smeets et al., *Srsf2*^{P95H} "gene-duplication" model

Recently, we reported a third conditional *Srsf2*P95H knock-in model [30] using a gene duplication method (Figure 2). The endogenous *Srsf2* locus including *Mfsd11* exon 1 and an introduced human growth hormone polyadenylation signal (hGHpA) was flanked by LoxP sites. The hGHpA was inserted downstream of the endogenous *Srsf2* 3'UTR to prevent/reduce transcriptional read-through into the mutant allele in the absence of Cre. A duplication of the entire *Srsf2* locus and *Mfsd11* exon 1 was inserted downstream of the 3'loxP sequence. The proline 95 (CCG) to histidine (CAT) mutation was introduced at the duplicated *Srsf2* exon 1 (Figure 3). Upon Cre-mediated excision, the endogenous *Srsf2* locus was deleted and the duplicated locus containing the P95H mutation was retained. RNA sequencing confirmed the heterozygous expression at both the genomic and transcriptional level. Analysis of Cre negative animals and animals not treated with tamoxifen (the CreER inducer used in these studies) demonstrated a low level of transcriptional read-through in the absence of activation of the P95H bearing allele. Long-term monitoring of Cre-ve and non-tamoxifen treated/recombined Cre+ve animals demonstrated that this "leaky" expression was not pathogenic.

Multiple Cre lines were applied to understand the cell population in which P95H was required to be able to impact normal hematopoiesis: whole body/broadly expressed (*R26*-

CreER^{T2} [52]), more specific to the HSC and primitive progenitor populations (h*Scl*CreER$^{T\ Tg/+}$ [53]) and a myeloid progenitor targeted constitutive Cre line (*LysM*-Cre$^{Ki/+}$ [54]). We analyzed the effect of *Srsf2*$^{P95H/+}$ on native hematopoiesis at 20 weeks post activation of the P95H mutation. In all cases, Cre+ve wild-type animals were used as controls and for CreER models these were tamoxifen treated CreER+ve Srsf2 wild-type littermate controls. At 20 weeks post *Srsf2*$^{P95H/+}$ activation, *Srsf2*$^{P95H/+}$ mice developed macrocytic anemia and increased myeloid cells in the peripheral blood. The myeloid bias was more evident in the bone marrow and was accompanied by compromised B lymphopoiesis. There was a reduction in erythropoiesis in the bone marrow accompanied by increased splenic erythropoiesis. Within the phenotypic stem and progenitor populations, there was a reduction or trend to reduction in the number of stem cell populations using two different phenotypic methods (LKS+CD34/Flt3 [55-57] or LKS+CD105/CD150 [58, 59]), while the more mature myeloid progenitor populations remained largely unchanged. The level of CD45RB, a previously characterized splicing target of *Srsf2* [21], was reduced in *Srsf2*$^{P95H/+}$ splenocytes, indicating that altered RNA splicing was being reflected in the proteome. A similar phenotype was seen in both *R26*-CreER^{T2} and h*Scl*CreER$^{T\ Tg/+}$ models but not in the *Lys*M-Cre$^{Ki/+}$ model, indicating that the *Srsf2*$^{P95H/+}$ mutation needed to arise within the primitive populations (including HSCs) to modify native hematopoiesis. RNA-seq analysis of lin-cKit+eYFP+ cells (where eYFP marked cells and their daughters/progeny in which Cre was activated) isolated 20 weeks after activation of P95H demonstrated that the *Srsf2*$^{P95H/+}$ mutation induced gene expression changes consistent with myeloid bias, loss of lymphoid potential and transcriptional signatures found in MDS. We identified similarly mis-spliced transcripts as Kon et al., and also could not define mis-splicing of *Ezh2* in the murine setting.

In agreement with the previous knock-in models of spliceosome mutations, the *Srsf2*$^{P95H/+}$ cells exhibited a poor competitive engraftment potential. This appears to be a universal feature of the models described to date, with the work of Kon et al.[32] best characterizing this phenotype and demonstrating an engraftment defect in the mutant cells. This experimental result is confounding, as *SRSF2* mutations are subjected to positive selection and are implicated in age-related clonal hematopoiesis [7, 8, 14]. However, we observed that the poor engraftment could be modified by altering the nature of the bone marrow competitor used. Using the h*Scl*CreER$^{T\ Tg/+}$ model, we found that *Srsf2*$^{P95H/+}$ cells could competitively engraft recipient animals and expanded when transplanted with the age/microenvironment matched competitor cells (i.e. cells taken from the same bone marrow but where Cre was not activated, based on *R26*-eYFP reporter marking of the cells).

Upon aging, non-transplanted *Srsf2*$^{P95H/+}$ mice developed fatal MDS by ~12 months after activating the P95H mutation in both the *R26*-CreER^{T2} and h*Scl*CreER$^{T\ Tg/+}$ model. The disease in both models was highly comparable. When moribund, the mice presented with

macrocytic anemia, myeloid bias (granulocytosis and monocytosis) and morphological dysplasia of myeloid and erythroid lineages in both the peripheral blood (>10-50% of cells) and bone marrow, all characteristics of MDS/MPN. Analysis of a small cohort by exome capture demonstrated that there was a subclonal accumulation of mutations associated with human *SRSF2* mutant disease, including *Dnmt3a*, *Tet2*, *Phf6*, and *Ras* members, in the bone marrow of the moribund animals [30, 60]. Non-competitive transplant of the moribund bone marrow recapitulated the MDS/MPN but there was no evidence of progression to acute leukemia even with long term monitoring of secondary recipients or with concurrent p53 deletion. Of the currently described *Srsf2* point mutant models, this is the only model to develop monocytosis and MDS/MPN in the setting of native hematopoiesis.

Discussion

Our understanding of the effects of spliceosome mutations comes from the detailed analysis of the mutational architecture of MDS, MDS/MPN and related disorders in humans. In these settings the mutation generally arises together with other mutations that combinatorially contribute to the disease manifestations and phenotypes [60]. In most cases, RNA splicing mutations are not the sole causes of maliganancies as other driver mutations are present, and familial monogenic examples of RNA splicing mutations have not been described. However, isolated mutations in RNA splicing components can sometimes cause disease directly, such as in ~20% of *SF3B1*-mutant MDS where other driver mutations were not reported [17]. In the setting of clonal hematopoiesis of indeterminated potential (CHIP), mutations in *SRSF2* and *SF3B1* have been identified with a range of variant allele burden, including VAFs of >10% [7, 8, 61]. These observations would suggest that when present as the only driver mutation, spliceosomal mutations can confer advantages to the host cells. The development of murine models of human spliceosome mutations, both conventional knock-out models and humanized mutant models, has improved our understanding of how these mutations perturb normal hematopoiesis and ultimately promote malignancy. The establishment of tractable, autochthonous pre-clinical models has proven more challenging. The fidelity of the "humanized" murine models must be considered against the phenotypes associated with the same spliceosomal mutations in humans.

The independently generated "humanized" spliceosomal mutant mice described to date encompass a spectrum of phenotypes from very mild to overt MDS/MPN in the setting of native hematopoiesis (Summarized in Table 1). The majority of models have also reported analysis of bone marrow transplant recipients. This is a useful strategy to limit the expression of the mutation to the transplantable hematopoietic cells but is also a source of significant cellular stress. MDS, MDS/MPN and related cancers arise in the absence of this extreme stress in the vast majority of humans. Bone marrow transplant changes HSC clonal dynamics and contribution compared to native hematopoiesis and permanently damages and remodels the bone marrow microenvironment itself [42, 43, 62-65]. The use of transplant models also limits the capacity to understand the interplay of the mutant cells and a normal immune system, now appreciated to be an important aspect in cancer evolution and escape [66].

The models described demonstrate that the targeting strategy itself can have profound impacts on the utility of the model that is generated. The three models that have utilized an inverted exon approach have all observed ~30% variant allele frequency at the RNA level after locus recombination, demonstrating that despite genomic heterozygosity the expression from these modified alleles is not heterozygous. This caveat requires acknowledgement when these models are reflected against the human phenotypes

associated with these mutations. The reasons that these alleles do not express heterozygously are not clear, but the results are consistent across three independent models and loci, indicating that this effect is most likely resultant of the targeting strategy itself. The inverted exon model has been successful in establishing murine models harboring the Jak2(V617F) mutation, where the mice develop a myeloprofilerative neoplasm resembling polycythemia vera as occurs in humans with this mutation [67]. Whilst speculative, for alleles that provide a proliferative/survival advantage or are strongly positively selected the inverted exon model may be feasible as even sub-heterozygous expression may be sufficient to afford the cell with the advantage of the mutant allele.

Another consideration is how foreign elements, such as loxP sites, are placed within the endogenous gene structure. Whilst literature describing the systematic assessment of locus modifications is limited, some principals have emerged with the evolution of the gene targeting methods and from the more recent high-throughput targeting vector generation methods [68-72]. These include the placing of loxP elements relative to the promoter and intron/exon junctions (>200bp is generally advised); avoiding disruption of evolutionarily conserved regulatory elements and CpG islands that may be important for normal gene regulation; being cognizant of the intron and exon sizes when introducing new elements with caution where there are short introns/exons as this can affect splicing of the gene; and avoidance of modification to neighboring genes if possible. These generalized observations may be relevant to the alleles described for *Srsf2*, where there are only two coding exons and the mutation occurs in the first exon. There is a CpG island that encompasses the 5' promoter region and 5'UTR/exon 1, intron and exon 2 of *Srsf2* (UCSC genome browser mm10; chr11:116,852,080-116,853,903). The *Srsf2* allele described by Kon et al. [32], for example, used an inverted exon approach and was found to be transcriptionally inactive/silent when non-recombined. This allele involved significant modifications to the gene structure, including separation of the 5'UTR and exon 1 coding sequence and insertion of multiple recombination elements in the introns [32]. The model developed by Kim et al. [31], placed the 5'LoxP element 87bp from the start of the exon 1 containing sequence and modified the intron spacing significantly in both the pre and post-recombined configuration [31]. All of these modifications can potentially impact the expression of the locus.

In humans it is thought that spliceosome mutations arise within the HSC compartment [6, 73]. It is well defined experimentally that the cell of origin can impact the phenotypes in experimental leukemia models [74, 75], so the use of different Cre drivers that target distinct and/or overlapping cell populations may be important in the presentation of the phenotypes in these models. Two of the *Srsf2* and one of the *Sf3b1* alleles have been assessed using the constitutively active *Vav*-Cre [24, 32, 76]. *Vav*-Cre is active in the hematopoietic cells including HSCs, with evidence for activity also in endothelial cells and

some activity in the reproductive tissues [41, 77, 78]. Multiple groups have used the inducible *Mx1*-Cre, a broadly active Cre that is very efficient at recombining loxP flanked alleles in hematopoietic cells including HSCs as well as other tissues in the adult animal including the cells composing the bone marrow microenvironment [41, 79, 80]. *Mx1*-Cre requires administration of pIpC to induce an innate immune mediated interferon response, a known stimulus of cycling of HSCs [81, 82]. Whereas this can be used to assess native hematopoiesis, the spliceosome mutant models described to date using *Mx1*-Cre have predominantly reported phenotypic/functional analysis from recipients of bone marrow transplants, which were treated after recovery with pIpC. This approach is useful to restrict Cre activity in non-hematopoietic tissues as only the transplanted/engrafted bone marrow will express Cre [39, 83]. Such a strategy is effective for understanding the hematopoietic intrinsic effect of the mutation, however it introduces the stress of bone marrow transplant and changes the hematopoietic dynamics and bone marrow microenvironment as a result [42, 43, 62-65]. Coupled with the administration of pIpc 4-8 weeks after transplant this elicits substantive stress on the entire hematopoietic system. An alternative is to use CreER based models, where tamoxifen is required to be administered to activate Cre activity. We utilized both the *Rosa26*-CreER ^{T2} and *hScl*-CreER^{T} lines to allow widespread and relatively HSC restricted activation respectively, of *Srsf2*^{P95H} [30]. The CreER systems and the administration of tamoxifen are not benign, with evidence for toxicity of the CreER itself and tamoxifen administered at high doses subcutaneously can have direct effects on hematopoiesis [84, 85]. These effects can be reduced by oral administration of tamoxifen, shortening the duration of exposure to tamoxifen and by the use of heterozygous CreER alleles. The contribution of the specific Cre drivers to the phenotypes reported are likely secondary to the impacts of the locus specific genetic modifications themselves. However, a direct comparison of the same splicing mutation with different Cre strains would resolve this (for example *Rosa26*-CreER vs *Mx1*-Cre). In all situations, the parallel analysis of littermate Cre+ve wild-type and Cre-ve gene modified animals is an essential control for these studies, whether using *Mx1*-Cre or CreER systems.

The models of spliceosomal mutation described to date have demonstrated a range of phenotypes. Definitive reasons for this remain uncertain, with some interpretations citing a lack of conservation of splicing between species. Whilst the spliceosomal proteins themselves are highly evolutionarily conserved, this is not an insignificant consideration [16]. By virtue of the higher level of conservation of the exons between mouse and humans (89-94%), the exonic splice enhancer sequences are more highly conserved [86, 87]. Despite this, there is still uncertainty regarding the identity of disease relevant mis-spliced candidates in human samples and murine models. For example, in humans it was reported that missplicing of *EZH2* [31], and more recently *CASP8* [24, 88], occurred in *SRSF2* mutant

cells. For *Ezh2,* at least, this observation was not confirmed in multiple murine *Srsf2*$^{P95H/+}$ mutant models where other mis-spliced candidates identified in human samples could be confirmed [30, 32]. The phenotypic recapitulation and conserved disease evolution and progression that occured in the *Srsf2*$^{P95H/+}$ model that we described, indicates that the mouse is able to reproduce the core aspects of the human phenotypes associated with *SRSF2* mutation. However, the identification of the specific mis-spliced genes that cause disease development has remained elusive. There are several possibilities that could account for this. Firstly, that the transcriptomic analysis has yet to be performed on sufficiently purified disease initiating cells, in both human and mouse, and that the heterogeneity of the samples assessed to date masks identification of the relevant mis-spliced transcipts. Compounding this is that the computational analysis of splicing is yet to reach a consensus and the different computational methods yield distinct results. An alternative possibility is that subtle changes across many genes collectively contribute to the phenotype, rather than a small number of dramatically mis-spliced candidates. In such a possibility it will require highly purified samples and sufficient sequencing depth to confidently identify these. Another possibility is that mis-spling occurs within common cellular pathways in both species, but that the underlying individual genes are more variable and not necessarily conserved. An example of species conserved transcriptional and cellular consequence despite an apparent divergence of the individual transcripts can be seen in the A-to-I RNA editing field and in the phenoytpes of *ADAR1* mutation in human and mouse [89-91]. Unlike the exonic sequences bound by SRSF2, intronic sequences are more divergent between human and mouse, with 66.1% conservation of branchpoint sequences across mammals [92]. The co-ordinated recognition of the 5' and 3' splice sites together with brachpoint and polypyrimidine tract selection are critical to splicing fidelity. The utilization of highly conserved branchpoints is likely to be conserved to a greater extent across species than those with weaker recognition motifs, less conservation or alternative branchpoint spacing. These differences in the intronic sequences and motifs between species may contribute to the phenotypic differences between mouse and humans in particular for SF3B1 which binds the branchpoint within intronic sequence of the pre-mRNA.

Murine models of human disease associated mutations allow us to understand how these mutations individually contribute to disease initiation and maintenance. To provide this understanding, they must recapitulate as closely as possible the nature of the mutation, the expression of the mutant allele and the cell of origin. A critical assessment of the genetic modification strategies applied demonstrates that there are significant differences in the approaches that have been utilized. How the different targeting strategies and experimental approaches for individual gene mutations contribute to the phenotypes observed should be

taken into consideration along with the degree to which the model presents phenotypes that are consistent with the human disease carrying the same mutation.

Acknowledgements

Research Support: Leukaemia Foundation (MW: Grant-in-Aid; ST: PhD Scholarship); Cancer Council of Victoria (CW, MW; APP1126010); Victorian Cancer Agency Research Fellowship (CW; MCRF15015); and in part by the Victorian State Government OIS (to SVI).

Author Contribution Statement

Writing – Original Draft: JX, MS, CW; Writing – Review & Editing: JX, MS, ST, MW, LP, CW.

Competing Financial Interest Statement

All authors declare no conflicts of interest.

References

1. Tefferi, A. and J.W. Vardiman, Myelodysplastic syndromes. *N Engl J Med*, 2009. **361**(19): p. 1872-85. DOI:10.1056/NEJMra0902908.
2. Tan, S.Y., M.F. Smeets, A.C. Chalk, et al., Insights into myelodysplastic syndromes from current preclinical models. *World J Hematol.*, 2016. **5**(1): p. 1-22. DOI: 10.5315/wjh.v5.i1.1.
3. McQuilten, Z.K., E.M. Wood, M.N. Polizzotto, et al., Underestimation of myelodysplastic syndrome incidence by cancer registries: Results from a population-based data linkage study. *Cancer*, 2014. **120**(11): p. 1686-94. DOI: 10.1002/cncr.28641.
4. Cogle, C.R., Incidence and Burden of the Myelodysplastic Syndromes. *Curr Hematol Malig Rep*, 2015. **10**(3): p. 272-81. DOI: 10.1007/s11899-015-0269-y.
5. Cogle, C.R., B.M. Craig, D.E. Rollison, and A.F. List, Incidence of the myelodysplastic syndromes using a novel claims-based algorithm: high number of uncaptured cases by cancer registries. *Blood*, 2011. **117**(26): p. 7121-5. DOI: 10.1182/blood-2011-02-337964.
6. Sperling, A.S., C.J. Gibson, and B.L. Ebert, The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer*, 2017. **17**(1): p. 5-19. DOI: 10.1038/nrc.2016.112.
7. Jaiswal, S., P. Fontanillas, J. Flannick, et al., Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*, 2014. **371**(26): p. 2488-98. DOI: 10.1056/NEJMoa1408617.
8. Genovese, G., A.K. Kahler, R.E. Handsaker, et al., Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*, 2014. **371**(26): p. 2477-87. DOI: 10.1056/NEJMoa1409405.
9. Gangat, N., M.M. Patnaik, and A. Tefferi, Myelodysplastic syndromes: Contemporary review and how we treat. *Am J Hematol*, 2016. **91**(1): p. 76-89. DOI: 10.1002/ajh.24253.
10. Yoshida, K., M. Sanada, Y. Shiraishi, et al., Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, 2011. **478**(7367): p. 64-9. DOI:10.1038/nature10496.
11. Papaemmanuil, E., M. Cazzola, J. Boultwood, et al., Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*, 2011. **365**(15): p. 1384-95. DOI: 10.1056/NEJMoa1103283.
12. Meggendorfer, M., A. Roller, T. Haferlach, et al., SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood*, 2012. **120**(15): p. 3080-8. DOI: 10.1182/blood-2012-01-404863.

13. Itzykson, R., O. Kosmider, A. Renneville, et al., Clonal architecture of chronic myelomonocytic leukemias. *Blood*, 2013. **121**(12): p. 2186-98. DOI: 10.1182/blood-2012-06-440347.
14. Martincorena, I., K.M. Raine, M. Gerstung, et al., Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*, 2017. **171**(5): p. 1029-1041 e21. DOI: 10.1016/j.cell.2017.09.042.
15. Wahl, M.C., C.L. Will, and R. Luhrmann, The spliceosome: design principles of a dynamic RNP machine. *Cell*, 2009. **136**(4): p. 701-18. DOI: 10.1016/j.cell.2009.02.009.
16. Lee, Y. and D.C. Rio, Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*, 2015. **84**: p. 291-323. DOI: 10.1146/annurev-biochem-060614-034316.
17. Haferlach, T., Y. Nagata, V. Grossmann, et al., Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*, 2014. **28**(2): p. 241-7. DOI: 10.1038/leu.2013.336.
18. Papaemmanuil, E., M. Gerstung, L. Malcovati, et al., Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*, 2013. **122**(22): p. 3616-27. DOI: 10.1182/blood-2013-08-518886.
19. Colla, S., D.S. Ong, Y. Ogoti, et al., Telomere dysfunction drives aberrant hematopoietic differentiation and myelodysplastic syndrome. *Cancer Cell*, 2015. **27**(5): p. 644-57. DOI: 10.1016/j.ccell.2015.04.007.
20. Visconte, V., A. Tabarroki, L. Zhang, et al., Splicing factor 3b subunit 1 (Sf3b1) haploinsufficient mice display features of low risk Myelodysplastic syndromes with ring sideroblasts. *J Hematol Oncol*, 2014. **7**: p. 89. DOI: 10.1186/s13045-014-0089-x.
21. Wang, H.Y., X. Xu, J.H. Ding, J.R. Bermingham, Jr., and X.D. Fu, SC35 plays a role in T cell development and alternative splicing of CD45. *Mol Cell*, 2001. **7**(2): p. 331-42.
22. Komeno, Y., Y.J. Huang, J. Qiu, et al., SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing. *Mol Cell Biol*, 2015. **35**(17): p. 3071-82. DOI: 10.1128/MCB.00202-15.
23. Lee, S.C., H. Dvinge, E. Kim, et al., Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nat Med*, 2016. **22**(6): p. 672-8. DOI: 10.1038/nm.4097.
24. Lee, S.C., K. North, E. Kim, et al., Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. *Cancer Cell*, 2018. **34**(2): p. 225-241 e8. DOI: 10.1016/j.ccell.2018.07.003.

25. Seiler, M., A. Yoshimi, R. Darman, et al., H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat Med*, 2018. **24**(4): p. 497-504. DOI: 10.1038/nm.4493.
26. Shirai, C.L., B.S. White, M. Tripathi, et al., Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. *Nat Commun*, 2017. **8**: p. 14060. DOI: 10.1038/ncomms14060.
27. Fei, D.L., H. Motowski, R. Chatrikhi, et al., Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival. *PLoS Genet*, 2016. **12**(10): p. e1006384. DOI: 10.1371/journal.pgen.1006384.
28. Obeng, E.A., R.J. Chappell, M. Seiler, et al., Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell*, 2016. **30**(3): p. 404-17. DOI: 10.1016/j.ccell.2016.08.006.
29. Mupo, A., M. Seiler, V. Sathiaseelan, et al., Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia*, 2017. **31**(3): p. 720-727. DOI: 10.1038/leu.2016.251.
30. Smeets, M.F., S.Y. Tan, J.J. Xu, et al., Srsf2(P95H) initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells. *Blood*, 2018. **132**(6): p. 608-621. DOI: 10.1182/blood-2018-04-845602.
31. Kim, E., J.O. Ilagan, Y. Liang, et al., SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell*, 2015. **27**(5): p. 617-30. DOI: 10.1016/j.ccell.2015.04.006.
32. Kon, A., S. Yamazaki, Y. Nannya, et al., Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood*, 2018. **131**(6): p. 621-635. DOI: 10.1182/blood-2017-01-762393.
33. Shirai, C.L., J.N. Ley, B.S. White, et al., Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*, 2015. **27**(5): p. 631-43. DOI: 10.1016/j.ccell.2015.04.008.
34. Fei, D.L., T. Zhen, B. Durham, 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. DOI: 10.1073/pnas.1812669115.
35. Wu, S., C.M. Romfo, T.W. Nilsen, and M.R. Green, Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, 1999. **402**(6763): p. 832-5. DOI: 10.1038/45590.

36. Kandoth, C., M.D. McLellan, F. Vandin, et al., Mutational landscape and significance across 12 major cancer types. *Nature*, 2013. **502**(7471): p. 333-339. DOI: 10.1038/nature12634.
37. Graubert, T.A., D. Shen, L. Ding, et al., Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet*, 2012. **44**(1): p. 53-7. DOI: 10.1038/ng.1031.
38. Hochedlinger, K., Y. Yamada, C. Beard, and R. Jaenisch, Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell*, 2005. **121**(3): p. 465-77. DOI: 10.1016/j.cell.2005.02.018.
39. Walkley, C.R. and S.H. Orkin, Rb is dispensable for self-renewal and multilineage differentiation of adult hematopoietic stem cells. *Proc Natl Acad Sci U S A*, 2006. **103**(24): p. 9057-9062. DOI: 10.1073/pnas.0603389103.
40. Purton, L.E. and D.T. Scadden, Limiting factors in murine hematopoietic stem cell assays. *Cell Stem Cell*, 2007. **1**(3): p. 263-70. DOI: 10.1016/j.stem.2007.08.016.
41. Joseph, C., J.M. Quach, C.R. Walkley, et al., Deciphering hematopoietic stem cells in their niches: a critical appraisal of genetic models, lineage tracing, and imaging strategies. *Cell Stem Cell*, 2013. **13**(5): p. 520-33. DOI: 10.1016/j.stem.2013.10.010.
42. Quach, J.M., M. Askmyr, T. Jovic, et al., Myelosuppressive therapies significantly increase pro-inflammatory cytokines and directly cause bone loss. *J Bone Miner Res*, 2015. **30**(5): p. 886-97. DOI: 10.1002/jbmr.2415.
43. Sun, J., A. Ramos, B. Chapman, et al., Clonal dynamics of native haematopoiesis. *Nature*, 2014. **514**(7522): p. 322-7. DOI: 10.1038/nature13824.
44. Patnaik, M.M. and A. Tefferi, Refractory anemia with ring sideroblasts (RARS) and RARS with thrombocytosis (RARS-T): 2017 update on diagnosis, risk-stratification, and management. *Am J Hematol*, 2017. **92**(3): p. 297-310. DOI: 10.1002/ajh.24637.
45. Mason, C.C., J.S. Khorashad, S.K. Tantravahi, et al., Age-related mutations and chronic myelomonocytic leukemia. *Leukemia*, 2015. DOI: 10.1038/leu.2015.337.
46. Patnaik, M.M., T.L. Lasho, C.M. Finke, et al., Spliceosome mutations involving SRSF2, SF3B1, and U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, and prognostic relevance. *Am J Hematol*, 2013. **88**(3): p. 201-6. DOI: 10.1002/ajh.23373.
47. Patnaik, M.M., S.A. Parikh, C.A. Hanson, and A. Tefferi, Chronic myelomonocytic leukaemia: a concise clinical and pathophysiological review. *Br J Haematol*, 2014. **165**(3): p. 273-86. DOI: 10.1111/bjh.12756.
48. Thol, F., S. Kade, C. Schlarmann, et al., Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*, 2012. **119**(15): p. 3578-84. DOI: 10.1182/blood-2011-12-399337.

49. Wu, S.J., Y.Y. Kuo, H.A. Hou, et al., The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood*, 2012. **120**(15): p. 3106-11. DOI: 10.1182/blood-2012-02-412296.
50. Mian, S.A., A.E. Smith, A.G. Kulasekararaj, et al., Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. *Haematologica*, 2013. **98**(7): p. 1058-66. DOI: 10.3324/haematol.2012.075325.
51. Oguro, H., L. Ding, and S.J. Morrison, SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*, 2013. **13**(1): p. 102-16. DOI: 10.1016/j.stem.2013.05.014.
52. Ventura, A., D.G. Kirsch, M.E. McLaughlin, et al., Restoration of p53 function leads to tumour regression in vivo. *Nature*, 2007. **445**(7128): p. 661-5. DOI: 10.1038/nature05541.
53. Gothert, J.R., S.E. Gustin, M.A. Hall, et al., In vivo fate-tracing studies using the Scl stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. *Blood*, 2005. **105**(7): p. 2724-32. DOI: 10.1182/blood-2004-08-3037.
54. Clausen, B.E., C. Burkhardt, W. Reith, R. Renkawitz, and I. Forster, Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res*, 1999. **8**(4): p. 265-77.
55. Osawa, M., K. Hanada, H. Hamada, and H. Nakauchi, Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science*, 1996. **273**(5272): p. 242-5.
56. Adolfsson, J., R. Mansson, N. Buza-Vidas, et al., Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell*, 2005. **121**(2): p. 295-306. DOI: 10.1016/j.cell.2005.02.013.
57. Yang, L., D. Bryder, J. Adolfsson, et al., Identification of Lin-Sca1+kit+CD34+Flt3-short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated recipients. *Blood*, 2005. **105**(7): p. 2717-2723. DOI: 10.1182/blood-2004-06-2159.
58. Pronk, C.J., D.J. Rossi, R. Mansson, et al., Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell*, 2007. **1**(4): p. 428-42. DOI: 10.1016/j.stem.2007.07.005.
59. Kiel, M.J., O.H. Yilmaz, T. Iwashita, et al., SLAM family receptors distinguish hematopoietic stem and progrnitor cells and reveal endothelial niches for stem cells. *Cell*, 2005. **121**: p. 1109-1121. DOI: 10.1016/j.cell.2005.05.026.

60. Makishima, H., T. Yoshizato, K. Yoshida, et al., Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet*, 2017. **49**(2): p. 204-212. DOI: 10.1038/ng.3742.
61. McKerrell, T., N. Park, T. Moreno, et al., Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep*, 2015. **10**(8): p. 1239-45. DOI: 10.1016/j.celrep.2015.02.005.
62. Rodriguez-Fraticelli, A.E., S.L. Wolock, C.S. Weinreb, et al., Clonal analysis of lineage fate in native haematopoiesis. *Nature*, 2018. **553**(7687): p. 212-216. DOI: 10.1038/nature25168.
63. Busch, K., K. Klapproth, M. Barile, et al., Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, 2015. **518**(7540): p. 542-6. DOI: 10.1038/nature14242.
64. Henninger, J., B. Santoso, S. Hans, et al., Clonal fate mapping quantifies the number of haematopoietic stem cells that arise during development. *Nat Cell Biol*, 2017. **19**(1): p. 17-27. DOI: 10.1038/ncb3444.
65. Askmyr, M., J. Quach, and L.E. Purton, Effects of the bone marrow microenvironment on hematopoietic malignancy. *Bone*, 2011. **48**(1): p. 115-20. DOI: 10.1016/j.bone.2010.06.003.
66. Barreyro, L., T.M. Chlon, and D.T. Starczynowski, Chronic immune response dysregulation in MDS pathogenesis. *Blood*, 2018. DOI: 10.1182/blood-2018-03-784116.
67. Mullally, A., S.W. Lane, B. Ball, et al., Physiological Jak2V617F expression causes a lethal myeloproliferative neoplasm with differential effects on hematopoietic stem and progenitor cells. *Cancer Cell*, 2010. **17**(6): p. 584-96. DOI: 10.1016/j.ccr.2010.05.015.
68. Economides, A.N., D. Frendewey, P. Yang, et al., Conditionals by inversion provide a universal method for the generation of conditional alleles. *Proc Natl Acad Sci U S A*, 2013. **110**(34): p. E3179-88. DOI: 10.1073/pnas.1217812110.
69. Wassef, M., A. Luscan, A. Battistella, et al., Versatile and precise gene-targeting strategies for functional studies in mammalian cell lines. *Methods*, 2017. **121-122**: p. 45-54. DOI: 10.1016/j.ymeth.2017.05.003.
70. Lewandoski, M., Conditional control of gene expression in the mouse. *Nat Rev Genet*, 2001. **2**(10): p. 743-55. DOI: 10.1038/35093537.
71. Hall, B., A. Limaye, and A.B. Kulkarni, Overview: generation of gene knockout mice. *Curr Protoc Cell Biol*, 2009. **Chapter 19**: p. Unit 19 12 19 12 1-17. DOI: 10.1002/0471143030.cb1912s44.

72. Friedel, R.H., W. Wurst, B. Wefers, and R. Kuhn, Generating conditional knockout mice. *Methods Mol Biol*, 2011. **693**: p. 205-31. DOI: 10.1007/978-1-60761-974-1_12.
73. Woll, P.S., U. Kjallquist, O. Chowdhury, et al., Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell*, 2014. **25**(6): p. 794-808. DOI: 10.1016/j.ccr.2014.03.036.
74. Huntly, B.J., H. Shigematsu, K. Deguchi, et al., MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell*, 2004. **6**(6): p. 587-96. DOI: 10.1016/j.ccr.2004.10.015.
75. Krivtsov, A.V., D. Twomey, Z. Feng, et al., Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature*, 2006. **442**(7104): p. 818-22. DOI: 10.1038/nature04980.
76. de Boer, J., A. Williams, G. Skavdis, et al., Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur J Immunol*, 2003. **33**(2): p. 314-25. DOI: 10.1002/immu.200310005.
77. Georgiades, P., S. Ogilvy, H. Duval, et al., VavCre transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis*, 2002. **34**(4): p. 251-6. DOI: 10.1002/gene.10161.
78. Croker, B.A., D. Metcalf, L. Robb, et al., SOCS3 is a critical physiological negative regulator of G-CSF signaling and emergency granulopoiesis. *Immunity*, 2004. **20**(2): p. 153-65.
79. Kuhn, R., F. Schwenk, M. Aguet, and K. Rajewsky, Inducible gene targeting in mice. *Science*, 1995. **269**(5229): p. 1427-9.
80. Velasco-Hernandez, T., P. Sawen, D. Bryder, and J. Cammenga, Potential Pitfalls of the Mx1-Cre System: Implications for Experimental Modeling of Normal and Malignant Hematopoiesis. *Stem Cell Reports*, 2016. **7**(1): p. 11-8. DOI: 10.1016/j.stemcr.2016.06.002.
81. Essers, M.A., S. Offner, W.E. Blanco-Bose, et al., IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature*, 2009. **458**(7240): p. 904-8. DOI: 10.1038/nature07815.
82. Baldridge, M.T., K.Y. King, N.C. Boles, D.C. Weksberg, and M.A. Goodell, Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. *Nature*, 2010. **465**(7299): p. 793-7. DOI: 10.1038/nature09135.
83. Park, D., J.A. Spencer, B.I. Koh, et al., Endogenous bone marrow MSCs are dynamic, fate-restricted participants in bone maintenance and regeneration. *Cell Stem Cell*, 2012. **10**(3): p. 259-72. DOI: 10.1016/j.stem.2012.02.003.

84. Higashi, A.Y., T. Ikawa, M. Muramatsu, et al., Direct hematological toxicity and illegitimate chromosomal recombination caused by the systemic activation of CreERT2. *J Immunol*, 2009. **182**(9): p. 5633-40. DOI: 10.4049/jimmunol.0802413.
85. Sanchez-Aguilera, A., L. Arranz, D. Martin-Perez, et al., Estrogen signaling selectively induces apoptosis of hematopoietic progenitors and myeloid neoplasms without harming steady-state hematopoiesis. *Cell Stem Cell*, 2014. **15**(6): p. 791-804. DOI: 10.1016/j.stem.2014.11.002.
86. Caceres, E.F. and L.D. Hurst, The evolution, impact and properties of exonic splice enhancers. *Genome Biol*, 2013. **14**(12): p. R143. DOI: 10.1186/gb-2013-14-12-r143.
87. Goren, A., O. Ram, M. Amit, et al., Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell*, 2006. **22**(6): p. 769-81. DOI: 10.1016/j.molcel.2006.05.008.
88. Shiozawa, Y., L. Malcovati, A. Galli, et al., Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat Commun*, 2018. **9**(1): p. 3649. DOI: 10.1038/s41467-018-06063-x.
89. Liddicoat, B.J., R. Piskol, A.M. Chalk, et al., RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science*, 2015. **349**(6252): p. 1115-20. DOI: 10.1126/science.aac7049.
90. Heraud-Farlow, J.E. and C.R. Walkley, The role of RNA editing by ADAR1 in prevention of innate immune sensing of self-RNA. *J Mol Med (Berl)*, 2016. **94**(10): p. 1095-1102. DOI: 10.1007/s00109-016-1416-1.
91. Rice, G.I., P.R. Kasher, G.M. Forte, et al., Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nat Genet*, 2012. **44**(11): p. 1243-8. DOI: 10.1038/ng.2414.
92. Mercer, T.R., M.B. Clark, S.B. Andersen, et al., Genome-wide discovery of human splicing branchpoints. *Genome Res*, 2015. **25**(2): p. 290-303. DOI: 10.1101/gr.182899.114.

Table 1. Summary of the phenotypes of the "humanized" spliceosomal mutant mice described

Gene	Model	rtTA/Cre	VAF	Native hematopoiesis			Non-competitive transplant			Competitive transplant
				Peripheral Blood/Spleen	BM	Dysplasia/ Disease	Peripheral Blood/Spleen	BM	Dysplasia/ Disease	
U2af1	Tet-On (Shirai et al.[33])	*R26*-M2rtTA	NA[1]	Not described			Leukopenia ↓B cell ↓monocytes	↑CMP[2]% ↑LKS+%	No dysplasia, MDS or AML up to 500 days post Tx	N/A
	Inversion (Fei et al.[34])	*UBC*-CreER^{T2}	~30%	MEFs only: Not further characterized						
	Mini-gene (Fei et al.[34])	*Mx1*-Cre	~50%	Macrocytic anemia; Leukopenia: ↓B cells	↓%LKS+ ↓%LTHSC	< 1% in BM, 0% in PB. No MDS or AML	Similar to native hematopoiesis			↓↓Engraftment
Sf3b1	Inversion (Obeng et al.[28])	*Mx1*-Cre	~30%	Macrocytic anemia ↓RBC ↑Plasma EPO ↑Erythroblast in spleen	↑LTHSC ↓GMP	Erythroid dysplasia in spleen, no ring sideroblasts /No MDS (64 wks post pIpC)	Similar to native hematopoiesis			↓↓Engraftment
	Mini-gene (Mupo et al.[29])	*Mx1*-Cre	~50%	Normocytic anemia	↓HSC% ↑Grans ↓Mature Eryth	No dysplasia / ring sideroblasts; No MDS (83 wks post pIpC)	N/A			↓↓Engraftment
Srsf2	Duplication (Kim et al.[31])	*Mx1*-Cre	~50%	Not described			Macrocytic anemia Leukopenia: ↓B cells	↑LKS; ↑LKS in S phase; ↑HPC2 ↓B cells	Neutrophil and erythroid dysplasia in PB; No AML (70wks)	↓↓Engraftment ↑LKS ↑HPC-2

	Inversion (Kon et al.[32])	*Vav1*-Cre	~31%	Macrocytic anemia	↓LTHSC ↓MPP ↓HPC-1	No dysplasia /No MDS or AML (90 weeks)	Macrocytic anemia Leukopenia ↑Grans ↓B cell	↓LTHSC; ↓MPP; ↓HPC1; ↓CMP; ↓CLP;	Dysplastic erythroid and ↑megakaryo-cytes in BM /No MDS or AML	↓↓Engraftment ; defect in homing
	Duplication (Smeets et al.[30])	*R26*-CreER^{T2}; h*Scl*-CreER^{T} $^{Tg/+}$; *Lys*M-Cre	~50%	Macrocytic anemia; ↑Grans; ↑Monocytes.	↓LTHSC; ↑%Grans; ↓B cells; ↓erythroid.	MDS/MPN by ~12 mths; no AML; ↑ and dysplastic myeloid cells (10-50% PB), monocytosis in PB / transplantable to 2^{ndry} recipients	Macrocytic anemia Leukopenia ↑%Grans & monocytes ↓B cell; Spleen - ↑% monocytes	↑%Grans; ↑%monoc ytes.	No AML	↓↓Engraftment in *R26*-CreER^{T2} model; myeloid bias; engraftment and ↑ mutant cells in h*Scl*CreER^{T} $^{Tg/+}$ model.

1. The allele frequency is doxycycline dose-dependent.
2. The immunophenotypic population listed in this table is defined as following: LKS+, linage-/c-Kit+/Sca-1+; LKS-, linage-/c-Kit+/Sca-1-; LT-HSC, LKS+/CD150+/CD48-; MPP, LKS+/CD150-/CD48-; HPC-1, LKS+/CD150-/CD48+; HPC-2, LKS+/CD150+/CD48+; HSC (Mupo et al.), LKS+/CD34-/flk2-; pre-MegE, LKS-/CD41-/FcγR-/CD150+/CD105-; pre-CFUE, LKS-/CD41-/FcγR-/CD15-+/CD105+. % = the percentage of this population.

Figure legends

Figure 1. A schematic outline of RNA splicing and the role of those proteins mutated in MDS. Within the intron of the pre-messenger RNA the U2 snRNP complex, containing SF3B1, binds the branchpoint site adenosine, U2AF2 binds to the polypyrimidine tract sequence and U2AF1 binds the 3' splice site; SRSF2 binds to the exonic splice enhancer sequence within the exonic sequence. *Cis*- and *trans*-acting factors can both positively or negative regulate splicing. Proteins where murine models have been generated are in bold font.

Figure 2. Simplified schematic outline of the currently described spliceosome mutant mouse models. LoxP elements are indicated as grey or green triangles indicating different types of LoxP elements; Expressed exons are in blue, exons not expressed until after locus recombination are in teal; Mutation bearing exons are orange before recombination (silent) and red when recombined and expressed.

Figure 3. Detailed genomic architecture of the *Srsf2*(P95H) knock-in mutant models. (**A**) the endogenous *Srsf2* locus. Figure to scale provided. (**B**) The *Srsf2* allele described by Smeets el al. This allele is drawn to scale with the spacing of the inserted sequences as shown. (**C**) An approximation of the allele generaterd by Kim et al., based on the description within the methods. The 5'loxP is stated as being 87bp upstream of exon 1. The spacing and position of the remaining sequences is approximate. (**D**) An approximation of the allele generated by Kon et al., using an inverted exon approach. The spacing and positions of the inserted sequences are approximations based on the description of the targeted allele. For panel (**C**) and (**D**) the modified locus depiction is approximate and based on the descriptions available in the publications reporting these alleles.

Response to Reviewers comments:

Ref.: MS. No. 18-183

Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal.

Please see a full version of the revised text with all changes highlighted in blue font at the end of the response to reviewers for review of the text changes (note references and table not included in this version).

Reviewer 1

This is an excellent review of mouse models of splicing defects in MDS and myeloid malignancies. My comments are minor and listed below.

1. Although found in other reviews, it would be useful to have a cartoon outlining the components of the splicing machinery as well as the cis elements (e.g. ESE, ESA) listed in Introduction.

Response:

As requested we have included a cartoon depicting in the splicing machinery highlighting the members relevant to the murine models discussed.

2. In the first paragraph, there are more recent references for estimates of MDS, which are higher than listed e.g. Cogle Cr et al, Blood 2015. Although the local data in Victoria, Australia is of interest, a population based study is of more relevance.

Response:

We thank the review for highlighting these additional references and have included these and text related to broadening the message of a significant underestimation of the rates of MDS in those >65 yrs old.

3. In the third paragraph on p. 4 "..indicates that retention of a wild-type copy is necessary for cells to survive, an interpretation supported by genetic evidence." , I would included "….supported by genetic and pharmacologic evidence", with the same and other references.

Response:

We have amended the text and references as suggested.

4. In all headings of the transgenic model, the actual reference number should be included e.g. on p.6 rather than Shirai et al, should be Shirai et al [29], so that it is clear which paper is being discussed.

Response:

We have modified the referencing to include the citation number as suggested by the reviewer.

5. The Discussion revisits some of the themes that are already discussed in the body of the review. Some further future looking thoughts would be of value here; e.g. how well can the mouse models actually identify specific genes involved in the mechanism of action of human disease. For instance, although EZH2 is not identified as a target, the phenotype of mouse models are somewhat similar to human disease. But where are the differences? And how can mouse models be improved to better understand molecular mechanisms of human disease? Can they be used to look for drug targets?

Response:
We thank the reviewer for their constructive comments. We have amended the discussion text to included a greater and more in depth discussion of the potential of the murine models to be used to identify mis-spliced transcripts.

Reviewer 2

This is a well-written review that does a good job of outlining the dissimilarities between different mouse models of splicing gene mutations associated with human MDS. It would also serve as a good guide for those designing complex mutant alleles for other genetic mutations, particularly for models of gain-of-function mutations (vs knock-out alleles).
The literature is covered fairly and cited aptly in the main. However, there are two important inaccuracies both on Page 16, Para 1 of discussion:
i) the premise that individual cases of splicing factor-mutant MDS always carry additional mutations. This is not the case for approximately 1 in 5 cases of SF3B1-mutant MDS, which appear not to carry other known driver mutations (e.g. Haferlach et al, Leukemia, 2014) and
ii) related to this, clonal hematopoiesis driven by splicing gene mutations is not always present at lower allele burdens (e.g. in ref 58 several cases had VAFs of >10%). The importance of getting this right is that otherwise readers will assume that splicing gene mutations cannot, in themselves, impart a significant clonal growth advantage on host cells.

Response:

We thank the reviewer for their constructive comments which are relevant and required clarification. We have amended the text and included text to reflect the reviewers comments for both points.

There are also some minor points and typos:
P3 Para 1 Line 2: "arising" should be "leading to"
P3 Para1 Line 15: "responses" is repeated"
P3 Para 2 Line 6: U2AF1 also deserves a mention here
P3 Para 3 Line 6: "part of the U5 snRNPs" should be "the U5 snRNP"
P3 Para 3 Line 8: add "respectively"
P4 Para 3 Line 7: "(not alleles resulting in protein deficiency)": move to end of sentence
P4 Para 3: evidence from mouse models (ref 41) should be cited with ref 21]
P4 Para 5 Line *: "to the usefulness" should change - e.g. "to the understanding of"
P6 Para 3: mention if mice irradiated
P6 Para 3 Line 11:change "modify" to "modifying"
P7: LoxP(2272) should change to Lox2272
P11 last line: "model" could change to "strategy"
P12 Para 1 lines 1-2:there are some inconsistencies here vs figure 1
P13 Para 3 line 7: exon1 Mfsd11 should be indicated in Figure 2
P14 Para 1 Line 15:not sure what is meant by "human MDS"
P14 Para 2 Line 9:"successfully competitive" needs to be explained
P14 Para 3 Line 7-8:some examples of mutations could help

Response:

We have amended the text as suggested for all of the points raised above and thank the reviewer for their careful reading of the text and suggested improvements.

P18 Para 2 Lines 9-10: the >65% conservation is accurate for dominant branchpoints, but not of weaker branchpoints which can be selected preferentially by mutant proteins such as SF3B1. This should be made clearer.

Response:

We have expanded our discussion of this point (and as it relates to Srsf2) in the discussion to reflect the comments of the reviewer.

NOTE: Text here with changes highlighted in BLUE so changes can be identified

Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal.

Jane Jialu Xu[1,2], Monique F Smeets[1,2], Shuh Ying Tan[1,2,3], Meaghan Wall[1,2,4], Louise E Purton[1,2] & Carl R Walkley[1,2,5,*].

[1]St. Vincent's Institute, Fitzroy, VIC, 3065 Australia.
[2]Department of Medicine, St. Vincent's Hospital, University of Melbourne, Fitzroy, VIC, 3065 Australia.
[3]Department of Hematology, St. Vincent's Hospital, Fitzroy 3065, VIC, Australia.
[4]Victorian Cancer Cytogenetics Service, St. Vincent's Hospital, Fitzroy, VIC, 3065 Australia.
[5]Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, VIC, 3000 Australia.

*Correspondence should be addressed to:
Carl Walkley (ORCID: 0000-0002-4784-9031)
St Vincent's Institute
9 Princes St
Fitzroy 3065 VIC
Australia
T: 61 3 9231 2480
F: 61 3 9416 2676
Email: cwalkley@svi.edu.au

Conflict of Interest Statement: All authors declare no conflicts of interest.

Running Title/Header: **Mouse models of RNA splicing mutations**

Category for the Table of Contents: **Malignant Hematopoiesis**

Key words: **mouse model; spliceosome; myelodysplastic syndrome; myeloproliferative disease; RNA splicing**

Metrics:
Abstract Word Count = 187
Word Count (excluding references, figure legends and tables) = 6958
Figures/Tables = 3 figures/ 1 Table

Abstract

Myelodysplastic syndromes (MDS) and related myelodysplastic/myeloproliferative neoplasms (MDS/MPN) are clonal stem cell disorders, primarily affecting patients over 65 years of age. Mapping of the MDS and MDS/MPN genome identified recurrent heterozygous mutations in the RNA splicing machinery, with *SF3B1*, *SRSF2* and *U2AF1* frequently mutated. To better understand how spliceosomal mutations contribute to MDS pathogenesis *in vivo*, numerous groups have sought to establish conditional murine models of *SF3B1*, *SRSF2* and *U2AF1* mutations. The high degree of conservation of hematopoiesis between mice and human and the well-established phenotyping and genetic modification approaches make murine models an effective tool to study how a gene mutation contributes to disease pathogenesis. The murine models of spliceosomal mutations described to date recapitulate human MDS or MDS/MPN to varying extents. Reasons for the differences in phenotypes reported between alleles of the same mutation are varied, but the nature of the genetic modification itself and subsequent analysis methods are important to consider. In this review, we will summarize recently reported murine models of *SF3B1*, *SRSF2* and *U2AF1* mutations, with a particular focus on the genetically engineered modifications underlying the models and the experimental approaches applied.

Introduction

Myelodysplastic syndromes (MDS), myelodysplastic/myeloproliferative neoplasms (MDS/MPN) and related disorders are a heterogeneous class of blood cancers leading to ineffective hematopoiesis in the bone marrow (BM) [1, 2]. Approximately 30% of MDS patients progress to acute leukemia. Median survival ranges from 97 months for low risk categories, down to 11 months for high risk MDS [2]. The incidence of MDS in the general population is ~4-5 per 100,000 people, but this increases with age [1]. Population based studies in both Australia and the USA indicate a significant underestimation of the true burden of MDS, with frequencies estimated at 103 per 100,000 and between 75-162 per 100,000 respectively over age 65 [3-5]. A feature of MDS and MDS/MPN is the progressive establishment of clonal hematopoiesis, where mutant cells dominate the BM at the expense of normal hematopoiesis [6]. The mechanisms behind this clonal advantage are unknown, but the identification of clonal hematopoiesis of indeterminate potential (CHIP) suggests that the clonal advantage is acquired progressively over many years following the establishment of an initiating mutation [7, 8]. In general, MDS patients have limited treatment options. Current treatment strategies for MDS and MDS/MPN are mostly heterogeneous and modestly efficacious and are not associated with durable responses [9]. The dismal prognosis following hypomethylating agent failure highlights the urgent need for new treatment approaches.

Until recently, there were few described recurrent mutations or familial syndromes that could provide insight into the genetics of these cancers and be used to develop preclinical models. Now, however, the detailed mutational architecture of human MDS, chronic myelomonocytic leukemia (CMML) and other related forms of MDS/MPN has been defined. There are recurrent mutations in key pathways, including the RNA splicing machinery (e.g. *SF3B1*, *SRSF2*, *U2AF1*) [10-12], transcription factors (e.g. *RUNX1*, *BCOR*, *ETV6*) and epigenetic enzymes (e.g. *DNMT3A*, *TET2, EZH2*, *ASXL1*) [6]. Additional studies have deduced clonal evolution and mutational order, highlighting distinct patterns [13]. Paralleling studies of diseased samples, analysis of healthy elderly populations has demonstrated that mutations associated with MDS and leukemia accumulate and are positively selected for in the hematopoietic cells from otherwise healthy people ≥70 years of age [7, 8, 14].

The identification of recurrent mutations in the RNA splicing machinery was a breakthrough in understanding the genetics of MDS and related disorders [10-12]. RNA splicing is a highly coordinated and essential process carried out by major and minor spliceosomes to remove non-coding regions (introns) of the pre-mRNA before protein translation [15, 16]. The major spliceosome consists of five small nuclear ribonucleoprotein complexes (snRNPs), U1, U2, U4, U5 and U6. The minor spliceosome includes the U5

snRNPs and other functional snRNPs mirroring the major spliceosome. Mechanistically, RNA splicing occurs with the initial recognition of the 5' and 3' splice sites by the U1 and U2 complexes respectively, followed by the excision of the intron and exon ligation by U4, U5 and U6 complexes. *Trans*-acting Splicing factors, such as serine-arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs), bind to the regulatory elements located in the exons and introns to enhance or repress the splicing activity and contribute to alternative splicing (Figure 1). The regulatory elements include exonic splicing enhancers and silencers (ESEs and ESSs) and intronic splicing enhancers and silencers (ISEs and ISSs) [16].

In MDS, >80% of patients have a mutation in a spliceosome gene [10-12]. In phenotypically overlapping syndromes of MDS/MPN, such as CMML, spliceosome mutations are also common [10-12, 17]. Mutations occur most frequently in *SF3B1* and *SRSF2*, with a lower prevalence in other spliceosomal genes including *U2AF1*, *ZRSR2* and *U2AF2* [17, 18]. Spliceosomal mutations are thought to arise early in the course of disease, including as founder/initiating events [6]. How mutations in the spliceosome alter normal hematopoiesis and contribute to disease pathogenesis remains unclear. A key experimental system to understand the impact of the spliceosome mutations in regulating both normal and malignant hematopoiesis is the generation of high fidelity murine models. There is a very high conservation of the human and mouse core spliceosome proteins, including those mutated in MDS and CMML. For example, human SRSF2 and murine Srsf2 proteins are 100% sequence conserved.

The generation of conventional null alleles of RNA splicing genes in the mouse has been useful in identifying key RNA substrates whose splicing is perturbed in their absence and in beginning to understand their normal physiological function [19-22]. Retroviral overexpression has been used to assess how the mutations present in patients may contribute to MDS development [10, 22]. A caveat with both the loss of function alleles and retroviral overexpression of mutant cDNAs is that they fail to recapitulate the mutations as they occur in humans: heterozygous point mutations expressed from their endogenous locus, and not alleles resulting in protein deficiency. The recurrent finding that the spliceosome mutations are heterozygous indicates that retention of a wild-type copy is necessary for cells to survive, an interpretation supported by both genetic and pharmacologic evidence [23-26], and additionally that gene dosage and protein complex stoichiometry may be important in the pathogenicity of the mutations [27].

Over the last 3 years, murine models recapitulating point mutations identified in patients with MDS and CMML have been described for *Sf3b1* [28, 29], *Srsf2* [30-32] and *U2af1* [33, 34]. The initial *U2af1* mutation model reported was an inducible cDNA expressed from a heterologous locus but will be discussed alongside the true knock-in alleles [33]. The

phenotypes observed in these “humanized” models have ranged from very mild through to the development of MDS/MPN. The reasons for these variable phenotypes have been not been systematically outlined, however the targeting strategy itself and the analysis methods may contribute significantly to the understanding of these models. Here we will discuss in detail the murine models of these mutations developed to date, their phenotypes and how they compare to the phenotypes of humans with these mutations.

Murine models of RNA spliceosome mutations

U2AF1 (previously known as *U2AF35*)

U2 small nuclear RNA auxiliary factor 1 (U2AF1) binds to the AG nucleotides and polypyrimidine tract located near the 3' splice site in the intronic sequence (Figure 1). It then recruits U2AF2 and the U2 major spliceosome complex to facilitate 3' splicing site recognition [15, 35]. U2AF1 mutations have been identified in lung, breast, colon and various epithelial carcinomas [36]. In MDS, approximately 11% of patients have a *U2AF1* mutation, predominantly at serine 34 (S34F/S34Y) and glutamine 157 (Q157R/Q157R) [10, 37]. One transgenic and two knock-in models have been developed to date.

U2AF1^{S34F} transgenic model

Shirai et al., *U2AF1^{S34F}* Tet-on transgenic model

Shirai et al. [33], reported the generation of a transgenic model of *U2AF1* mutation. In this case, unlike the others that will be discussed, an inducible (tetO) human *U2AF1^{S34F}* or control *U2AF1^{WT}* cDNA were targeted to the Collagen1a1 (*Col1a1*) locus of *Rosa26*-M2rtTA ES cells [38]. Expression of the WT or S34F cDNAs was not regulated in the manner of the endogenous *U2af1* locus. To activate the expression of the *U2AF1^{S34F}* or *U2AF1^{WT}* cDNAs, animals were administered doxycycline, enabling dose dependent induction of expression of the human *U2AF1* cDNAs in the presence of two endogenous wild-type gene copies derived from the mouse (Figure 2).

No analysis was reported of the effects of *U2AF1^{S34F}* on native hematopoiesis (i.e. in animals treated with doxycycline in the absence of prior bone-marrow transplantation). The only data presented was from irradiated transplant recipients reconstituted with whole bone marrow of *U2AF1^{S34F}* M2rtTA or *U2AF1^{WT}* M2rtTA donors. Once reconstitution was established, the recipient animals were administered doxycycline [33]. A similar approach was chosen in a second report of this model [26]. The use of BM transplantation prior to gene induction restricts expression of the mutation to the reconstituted hematopoietic cells, of particular importance when using an inducer expressed from the widely expressed *Rosa26* locus or others like *Mx1* that are active in non-hematopoietic cell populations [39-41]. A caveat of analysis in post-transplant settings only is that bone marrow transplantation is known to induce hematopoietic stress, permanently modifying the bone marrow microenvironment [42] and alter the clonal dynamics of HSCs [43].

Using this experimental approach, the authors showed that the doxycycline treated *U2AF1^{S34F}* M2rtTA recipients developed a stable peripheral blood leukopenia, without changes in red blood cell or platelet numbers, for up to 12 months. The leukopenia was the result of reduced B cells and monocytes and was dependent on the continued expression of the mutant *U2AF1*. Within the bone marrow there was evidence of increased cell death,

reduced B cells and monocytes and increased neutrophils. There was an increase in the percentages of common myeloid progenitors and a subtle increase in the percentage of LKS cells. In competitive transplants, *U2AF1*^{S34F} cells were serially transplantable through to tertiary recipients, albeit with a reduced chimerism. With extended aging (up to 500 days post bone marrow transplant), the *U2AF1*^{S34F} expressing mice were not reported to develop dysplasia, MDS or AML.

U2af1^{S34F} *Knock-in (KI) models*

Fei et al. [34], reported the generation of two conditional knock-in models of *U2af1*^{S34F} mutation targeted to the endogenous locus. The S34F mutation is encoded by exon 2 of the murine *U2af1* allele. Both models required Cre protein to induce expression of the mutant allele. Two different knock-in approaches were described, one using an inverted exon 2 (referred to in the publication as *IES34F*) and the second using a minigene (referred to as *MGS34F*) which resulted in significant differences in the expression of the mutant allele (Figure 2). Both were tested as heterozygous alleles, where one allele encoded endogenous wild-type *U2af1* and the second the mutant allele.

Fei et al., *U2af1*^{S34F} "inverted exon" model

For the inverted exon 2 model, a lox2272 flanked inverted S34F encoding exon 2 was placed in intron 2-3 and a conventional LoxP element was placed 5' of the endogenous exon 2. Cre induced recombination of the different LoxP sequences enabled removal of the endogenous exon 2 and replacement by an in-frame exon 2 encoding the S34F mutation. Analysis of mutant allele expression after Cre recombination in MEFs, demonstrated that the mutant mRNA was expressed at ~30%, despite being genomically heterozygous. It is not clear why the *IES34F* allele failed to express the transcript heterozygously, although the authors suggested it related to the efficiency of recombination of the locus. This allele was not further characterized [34].

Fei et al., *U2af1*^{S34F} "minigene" model

The minigene model was examined in more detail [34]. In this model, a LoxP flanked cDNA sequence encoding exons 2-8 followed by a transcriptional stop sequence was placed in the first intron. The S34F mutation was inserted in the endogenous exon 2. In the absence of Cre, the *U2af1* transcript from the targeted allele should be encoded by the endogenous exon 1 together with the cDNA encoding exons 2-8, with the stop cassette preventing read through and expression from the S34F encoding allele. This model was tested in the setting of native hematopoiesis using the broadly expressed *Mx1*-Cre and polyI:polyC (pIpC) induction. The analysis of recombination following pIpC administration indicated that the

minigene, unlike the inverted exon model, was efficiently recombined and truly heterozygous both genomically and transcriptionally. Analysis of the *U2af1*$^{S34F/+}$ animals post-pIpC demonstrated a mild persistent macrocytic anemia and a 2-fold reduction in leukocytes, particularly B cells, in the peripheral blood. Within the bone marrow (BM), total cellularity was preserved, but there were reductions in the percentages of LKS+ and LKS+CD48-CD150+ phenotypic hematopoietic stem cells (HSCs). Lineage distribution of more committed populations within the BM was not described. Morphological analysis of the bone marrow revelaed less than 1% of cells displaying dysplastic features and no dysplastic cells were reported in the peripheral blood. No MDS or AML was observed in aged *U2af1*$^{S34F/+}$ mice. The phenotypes described for native hematopoiesis were preserved, albeit exacerbated, by bone marrow transplantation where expression of *U2af1*$^{S34F/+}$ was activated post-transplant, demonstrating a cell-autonomous effect of the mutant allele on hematopoiesis. Interestingly, and consistent across the splicing mutation models reported to date, there was a significantly reduced repopulating potential of U2af1(S34F)-expressing cells, when bone marrow cells from donors previously treated with pIpC competed against WT bone marrow cells.

SF3B1

Splicing factor 3b subunit 1 (SF3B1) is part of the SF3b protein complex that recognizes the branch point adenosine (A) base within the intron and facilitates the binding of U2 snRNP on pre-mRNA (Figure 1). *SF3B1* mutations have been identified in MDS, chronic lymphocytic leukemia (CLL) and uveal melanoma (UM). Mutations cluster between exons 14 to 16, with K700E being the most frequent mutation [10, 11]. In MDS, *SF3B1* mutations occur in >80% of refractory anemia with ring sideroblasts (RARS), a MDS subtype with a favorable outcome [11, 44]. Two independently generated *Sf3b1* conditional knock-in models have been reported and both have utilized the *Mx1*-Cre system to induce the expression of the mutant allele.

Sf3b1*^{K700E} *Knock-in (KI) models

Obeng et al., *Sf3b1*^{K700E} "inverted exon" model

A conditional *Sf3b1*^{K700E} model was reported by Obeng and colleagues [28], generated using an inverted exon strategy (Figure 2). In this model, an exon 15 with the A>G substitution and a WT exon 16 were introduced in an inverted orientation within intron 16-17. Upon Cre-induction, the inverted exons are recombined in the correct orientation in place of the endogenous WT exons through the use of different loxP sequences and, in theory, should be physiologically expressed from the endogenous locus. Whilst the *Sf3b1*^{K700E} allele was successfully inverted post-Cre induction (with pIpC treatment), it did not achieve

heterozygous expression when the RNA levels were assessed. The authors presented results of RNA-seq of lin-cKit+ cells from the bone marrow of three *Sf3b1*$^{K700E/+}$ animals, where the mutant allele frequency was between 27%-32% (see Figure S1A; Obeng et al., [28]). A similar level of sub-heterozygous mutant transcript expression was detected in a second recent report using this allele [24] and the phenotypes reported for these animals need to be considered with this caveat.

Monitoring of native hematopoiesis over 64 weeks of pIpC treated cohorts showed a progressive macrocytic anemia with decreased red blood cell counts from 20 weeks post activation of the K700E mutation. Corresponding with anemia, there was an increased level of plasma erythropoietin in *Sf3b1*$^{K700E/+}$ mice. This anemic phenotype was shown to be cell-intrinsic through use of non-competitive bone marrow transplant assays. There was no change in white blood cell or platelet counts. None of the *Sf3b1*$^{K700E/+}$ mice were reported to develop MDS during the period of monitoring. At 64 weeks post pIpC, analysis of the bone marrow showed no change in cellularity, while there was a significant increase in the number of long-term HSCs (LKS+CD150+CD48-) and a decrease in the GMPs. In the spleen, there was an accumulation of erythroblast populations (Ter119+CD71^{high}) and a decrease in the more mature erythroid population (Ter119+CD71^{low}), indicating a terminal erythroid maturation defect. Interestingly, this effect was only reproduced *in vivo* in young mice (11 weeks post pIpC) under stress (drug-induced hemolysis) or *in vitro* when bone marrow progenitor cells (24 weeks post pIpC) were treated with cytokines to induce erythroid differentiation. Morphological examination of the spleen showed an increased number of erythroid precursors and erythroid dysplasia. However, ring sideroblasts or elevated levels of iron deposition were not observed. There was a severe impairment (<20%) in competitive repopulation capacity of *Sf3b1*$^{K700E/+}$ cells (similar to Mupo et al.,[29]), even though >95% chimerism was achieved in non-competitive transplant experiments.

Mupo et al., *Sf3b1*^{K700E} "minigene" model

Mupo et al. [29], generated a conditionally activatable *Sf3b1*$^{K700E/+}$ allele using a mini-gene approach (Figure 2). Here they engineered a loxP flanked construct containing a splice acceptor site (from an Engrail-2 splice acceptor sequence), exons 12–15, intron 15, exons 15–19 (exons 12-19 were codon optimized sequence), intron 19 and the naturally occurring sequences for exons 20–25 including the 3′ UTR, all this followed by a SV40 polyadenylation signal (SV40 pA). This mini-gene was inserted between exons 11 and 12 of *Sf3b1* and the endogenous exon 15, downstream of the mini-gene, was replaced with a mutated, synthetic exon 15 harboring an A>G mutation encoding K700E. Upon expression of Cre protein from *Mx1*-Cre, induced by administration of pIpC, the mine-gene is deleted allowing the expression of the *Sf3b1* allele containing the mutated exon 15 (*Sf3b1*$^{K700E/+}$).

The allele was assessed in a heterozygous setting, with one mutant allele and one WT allele, mirroring the patient setting. RNA-seq analysis confirmed the 50% expression level of the mutated allele in whole bone marrow and lineage negative (lin-) cells from *Sf3b1*$^{K700E/+}$ mice. Four weeks post pIpC treatment native hematopoiesis was assessed and a decrease in the percentage of phenotypic HSCs (LKS+CD34-Flk2-) was observed, but no change in the percentage of lymphoid-primed multipotent progenitors (L-MPPs), common myeloid progenitors (CMPs), granulo-monocytic progenitors (GMP) or myeloid-erythroid progenitors (MEPs). Bone marrow analysis indicated a mild myeloid bias (Gr-1+Mac-1+) and impaired terminal erythroid differentiation (Ter119+CD71^{low}FSC^{low}). There was no overall survival difference between *Sf3b1*$^{K700E/+}$ mice and WT littermates followed longitudinally for up to 83 weeks. Peripheral blood analysis revealed a progressive normocytic anemia with no change in white blood cell or platelet counts. None of the *Sf3b1*$^{K700E/+}$ mice were reported to develop MDS during monitoring. Morphological analysis of the bone marrow did not find evidence of either dysplasia or ringed sideroblasts (a feature of human *SF3B1* mutant MDS), despite increased iron deposits in bone marrow macrophages. Peripheral blood morphology was not reported. Competitive bone marrow transplantation was performed using both young (2 months) and old (12 months) recipients to check the potential influence of aging recipients on donor cell engraftment. In both groups, *Sf3b1*$^{K700E/+}$ cells showed poor chimerism indicating impaired engraftment/repopulating potential. There was no difference between the young and old recipient groups with neither developing overt symptoms of MDS.

SRSF2

Serine-arginine rich factor 2 (SRSF2) binds to exonic splicing enhancers within the exonic sequence located near the 3' splice site and mediates exon inclusion and exclusion (Figure 1). Unlike *U2AF1* or *SF3B1*, *SRSF2* mutations occur near exclusively in myeloid malignancies and at or encompassing proline 95 (P95), with the most frequent mutation reported a proline to histidine substitution (P95H) [10, 12, 17]. *SRSF2* mutations occur in around 12-15% of MDS [9] and 28-52% of CMML [45-47]. Mutations are associated with poor outcomes, a shortened time to leukemic conversion and a myelo-monocytic bias [48-50]. Three independently generated models of *Srsf2*^{P95H} have now been reported and these will be discussed in greater detail than those of *U2af1* and *Sf3b1* given our own work on this gene.

Srsf2*^{P95H} *Knock-in (KI) models

Kim et al., *Srsf2*^{P95H} "gene-duplication" model

Kim et al. [31], reported the first *Srsf2*^{P95H} conditional knock-in model. The allele described approximated a gene-duplication strategy (Figure 2). The targeting construct inserted a loxP site 87bp upstream of the 5'UTR sequence of the endogenous exon 1. The sequence containing exon 3 (3'UTR) and an introduced SV40 polyadenylation sequence (SV40pA) was placed immediately after the endogenous exon 2 to generate an *Srsf2* allele that would express a wild-type protein. The second loxP was placed after the SV40pA. Downstream of the second loxP the sequence of exon 1 containing the *Srsf2*P95H mutation (CG>AC) and the sequence for exon 2 were introduced into intron 2-3 (Figure 3). Very little data was presented describing the phenotypic effects of this *Srsf2*$^{P95H/+}$ model on native hematopoiesis (i.e not from transplant recipients). The authors state that none of the *Srsf2*$^{P95H/+}$ mice developed AML out to 70 weeks of monitoring, although it is not certain if this refers to *de novo* mutated animals or transplant recipients.

The authors principally reported analysis from non-competitive bone marrow transplant assays for the remaining studies. Subsequent studies with this *Srsf2* allele from the same laboratory have also used this approach [23, 24]. At 18 weeks post pIpC, the *Srsf2*$^{P95H/+}$ recipients developed macrocytic anemia and leukopenia due to decreased B cell numbers, while the platelet numbers remained normal. There was evidence of myeloid (hypolobated and hypogranulated neutrophils) and erythroid dysplasia in the peripheral blood/bone marrow but the estimated percentage of affected cells was not reported. Within the bone marrow of the recipients, Kim et al., reported that a major effect of the *Srsf2*$^{P95H/+}$ mutation was an expansion of the HPC-2 phenotype (LKS+CD150+CD48+) within the hematopoietic stem and progenitor population. The frequencies of the HPC-2 population as reported by Kim et al,. in the control *Mx1*-Cre *Srsf2*$^{+/+}$ animals was 0.0231%, significantly higher than that reported in the index reference for this population of 0.0032±0.0014% in WT mice [51]. The frequency of this population may be impacted by both the BM transplant model and pIpC use. However, the ~10 fold difference in reported frequency is not reconciled, and the representative flow cytometry gating strategy presented by Kim et al., is significantly different to the gating strategy in the original report of this population [51]. The authors reported an increased frequency of cells in S-phase and early apoptosis within the bone marrow. Notably, in a subsequent report from the same laboratory using this *Srsf2*$^{P95H/+}$ allele and the *Mx1*-Cre model, the leukopenia and macrocytic anemia were largely absent (see Supp Fig 1E-F [23]). The inconsistency of the phenotype reported across the publications has not been addressed, but potentially indicates that the phenotype is not fully penetrant. Through analysis of RNA-seq from both human patients and murine samples the authors identified a group of mis-spliced candidate genes, with *EZH2* proposed as a key mis-spliced gene. *SRSF2*^{P95H} was proposed to lead to the inclusion of a premature stop codon in *EZH2*, resulting in a putative loss of function allele. This analysis posited a model

that a single or small number of key mis-spliced genes may account for the phenotypes in the *SRSF2*^{P95H} mutant cells.

Kon et al., *Srsf2*^{P95H} "inverted exon" model

Kon et al. [32], reported an *Srsf2*^{P95H} conditional knock-in model developed using an inverted exon strategy (termed FLEx switch inversion technique; Figure 2). For the targeted allele, the 5'UTR and exon 1 were separated. Exon 1 was flanked by loxP/VloxP sites at the 5' end and VloxP/lox2272 3' to exon 1. An inverted exon 1 containing the P95H mutation was inserted downstream of the VloxP/lox2272 in intro 1, followed by second set of loxP/lox2272 sites. After Cre activation, the inverted exon containing P95H "flips" into the correct orientation and the endogenous exon 1 is excised at the same time (Figure 3). RNA sequencing showed no expression from the modified allele in Cre-negative heterozygous animals, indicating that the floxed allele was likely null when not recombined. This was consistent with the failure to obtain animals that had the non-recombined P95H allele homozygosed even in a Cre-negative background [32]. Why this is the case is unclear at present. The targeting strategy itself may be the most significant contributor in this particular instance. The separation of the 5'UTR from the first coding exon, in effect exonization of the UTR, and the numerous modifications around the coding exon and the sequence insertion within the first intron may have had a significant effect on endogenous gene regulation.

The *Srsf2*^{P95H} conditional knock-in mice were crossed to *Vav1*-Cre mice to assess the effects of the heterozygous knock-in allele on hematopoiesis. Unlike the other described models, *Vav1*-Cre is a constitutive Cre and is active in the hematopoietic cells without the addition of an inducer such as pIpC/IFNα (*Mx1*-Cre) or tamoxifen (CreER models). At around 15 weeks of age, the *Srsf2*$^{P95H/+}$ mice developed macrocytic anemia, which remained stable. After observation out to 90 weeks of age, none of the *Srsf2*$^{P95H/+}$ mice developed MDS or leukemia. A caveat with this model, as was described for the previous models describing inverted exon based targeting methods, was that the mutant allele was expressed at 27-35% (mean 31%) based on RNA-seq analysis of the *Vav1*-Cre *Srsf2*^{P95H} mice.

Upon analysis of the bone marrow under native conditions, there was a reduction in the number of LT-HSC (LKS+CD150+CD48-), MPP (LKS+CD150-CD48-) and HPC-1 (LKS+CD150-CD48+) populations, associated with increased cell cycling. There was no change in the cellularity or lineage distribution of bone marrow and spleen, with the exception of a mild B cell differentiation defect in the bone marrow. There was no evidence of dysplasia reported in the bone marrow under native settings. The authors then performed non-competitive and competitive bone marrow transplant assays of the *Srsf2*$^{P95H/+}$ cells. There was engraftment in the non-competitive transplant settings, however the recipients of competitive bone marrow transplants had a significant reduction in chimerism with either

whole bone marrow or HSCs (LKS+CD34-) as the donor. The impaired competitive transplant potential is a common feature of all of the conditional splicing mutants to date. Extensive work demonstrated an engraftment defect of the *Srsf2*$^{P95H/+}$ HSCs. The *Srsf2*$^{P95H/+}$ recipients developed macrocytic anemia and leukopenia. In the bone marrow, there was a significant reduction in LT-HSC and most of the progenitor populations (MPP, HPC-1, CMP, CLP). In addition, *Srsf2*$^{P95H/+}$ recipients had a significantly higher proportion of cycling and apoptotic HSCs than controls. The lineage distribution was skewed towards myeloid at the expense of B cell differentiation and dysplastic erythroid cells as well as dys-megakaryopoiesis were evident in the bone marrow.

Analysis of RNA splicing using multiple purified hematopoietic populations from both native hematopoiesis and cells isolated from transplant recipients was described. From these analyses the authors could identify mis-splicing of a number of genes that had been identified in human *SRSF2* mutant patient samples, including *Csf3r*, *Gnas*, *Hnrnpa2b1*, and several novel differentially spliced genes such as *Atrx* and *Mllt10* both of which are implicated in hematological malignancies. Of note, despite using numerous approaches the authors could not find evidence to support *Ezh2* as a differentially spliced candidate in the mouse. It was proposed that this may be due to species differences, with human *EZH2* having two CCNG motifs whilst mouse has a single CCNG motif [32].

Smeets et al., *Srsf2*^{P95H} "gene-duplication" model

Recently, we reported a third conditional *Srsf2*P95H knock-in model [30] using a gene duplication method (Figure 2). The endogenous *Srsf2* locus including *Mfsd11* exon 1 and an introduced human growth hormone polyadenylation signal (hGHpA) was flanked by LoxP sites. The hGHpA was inserted downstream of the endogenous *Srsf2* 3'UTR to prevent/reduce transcriptional read-through into the mutant allele in the absence of Cre. A duplication of the entire *Srsf2* locus and *Mfsd11* exon 1 was inserted downstream of the 3'loxP sequence. The proline 95 (CCG) to histidine (CAT) mutation was introduced at the duplicated *Srsf2* exon 1 (Figure 3). Upon Cre-mediated excision, the endogenous *Srsf2* locus was deleted and the duplicated locus containing the P95H mutation was retained. RNA sequencing confirmed the heterozygous expression at both the genomic and transcriptional level. Analysis of Cre negative animals and animals not treated with tamoxifen (the CreER inducer used in these studies) demonstrated a low level of transcriptional read-through in the absence of activation of the P95H bearing allele. Long-term monitoring of Cre-ve and non-tamoxifen treated/recombined Cre+ve animals demonstrated that this "leaky" expression was not pathogenic.

Multiple Cre lines were applied to understand the cell population in which P95H was required to be able to impact normal hematopoiesis: whole body/broadly expressed (*R26*-

CreER^{T2} [52]), more specific to the HSC and primitive progenitor populations (h*Scl*CreER$^{T\ Tg/+}$ [53]) and a myeloid progenitor targeted constitutive Cre line (*LysM*-Cre$^{Ki/+}$ [54]). We analyzed the effect of *Srsf2*$^{P95H/+}$ on native hematopoiesis at 20 weeks post activation of the P95H mutation. In all cases, Cre+ve wild-type animals were used as controls and for CreER models these were tamoxifen treated CreER+ve Srsf2 wild-type littermate controls. At 20 weeks post *Srsf2*$^{P95H/+}$ activation, *Srsf2*$^{P95H/+}$ mice developed macrocytic anemia and increased myeloid cells in the peripheral blood. The myeloid bias was more evident in the bone marrow and was accompanied by compromised B lymphopoiesis. There was a reduction in erythropoiesis in the bone marrow accompanied by increased splenic erythropoiesis. Within the phenotypic stem and progenitor populations, there was a reduction or trend to reduction in the number of stem cell populations using two different phenotypic methods (LKS+CD34/Flt3 [55-57] or LKS+CD105/CD150 [58, 59]), while the more mature myeloid progenitor populations remained largely unchanged. The level of CD45RB, a previously characterized splicing target of *Srsf2* [21], was reduced in *Srsf2*$^{P95H/+}$ splenocytes, indicating that altered RNA splicing was being reflected in the proteome. A similar phenotype was seen in both *R26*-CreER^{T2} and h*Scl*CreER$^{T\ Tg/+}$ models but not in the *Lys*M-Cre$^{Ki/+}$ model, indicating that the *Srsf2*$^{P95H/+}$ mutation needed to arise within the primitive populations (including HSCs) to modify native hematopoiesis. RNA-seq analysis of lin-cKit+eYFP+ cells (where eYFP marked cells and their daughters/progeny in which Cre was activated) isolated 20 weeks after activation of P95H demonstrated that the *Srsf2*$^{P95H/+}$ mutation induced gene expression changes consistent with myeloid bias, loss of lymphoid potential and transcriptional signatures found in MDS. We identified similarly mis-spliced transcripts as Kon et al., and also could not define mis-splicing of *Ezh2* in the murine setting.

In agreement with the previous knock-in models of spliceosome mutations, the *Srsf2*$^{P95H/+}$ cells exhibited a poor competitive engraftment potential. This appears to be a universal feature of the models described to date, with the work of Kon et al.[32] best characterizing this phenotype and demonstrating an engraftment defect in the mutant cells. This experimental result is confounding, as *SRSF2* mutations are subjected to positive selection and are implicated in age-related clonal hematopoiesis [7, 8, 14]. However, we observed that the poor engraftment could be modified by altering the nature of the bone marrow competitor used. Using the h*Scl*CreER$^{T\ Tg/+}$ model, we found that *Srsf2*$^{P95H/+}$ cells could competitively engraft recipient animals and expanded when transplanted with the age/microenvironment matched competitor cells (i.e. cells taken from the same bone marrow but where Cre was not activated, based on *R26*-eYFP reporter marking of the cells).

Upon aging, non-transplanted *Srsf2*$^{P95H/+}$ mice developed fatal MDS by ~12 months after activating the P95H mutation in both the *R26*-CreER^{T2} and h*Scl*CreER$^{T\ Tg/+}$ model. The disease in both models was highly comparable. When moribund, the mice presented with

macrocytic anemia, myeloid bias (granulocytosis and monocytosis) and morphological dysplasia of myeloid and erythroid lineages in both the peripheral blood (>10-50% of cells) and bone marrow, all characteristics of MDS/MPN. Analysis of a small cohort by exome capture demonstrated that there was a subclonal accumulation of mutations associated with human *SRSF2* mutant disease, including *Dnmt3a*, *Tet2*, *Phf6*, and *Ras* members, in the bone marrow of the moribund animals [30, 60]. Non-competitive transplant of the moribund bone marrow recapitulated the MDS/MPN but there was no evidence of progression to acute leukemia even with long term monitoring of secondary recipients or with concurrent p53 deletion. Of the currently described *Srsf2* point mutant models, this is the only model to develop monocytosis and MDS/MPN in the setting of native hematopoiesis.

Discussion

Our understanding of the effects of spliceosome mutations comes from the detailed analysis of the mutational architecture of MDS, MDS/MPN and related disorders in humans. In these settings the mutation generally arises together with other mutations that combinatorially contribute to the disease manifestations and phenotypes [60]. In most cases, RNA splicing mutations are not the sole causes of maliganancies as other driver mutations are present, and familial monogenic examples of RNA splicing mutations have not been described. However, isolated mutations in RNA splicing components can sometimes cause disease directly, such as in ~20% of *SF3B1*-mutant MDS where other driver mutations were not reported [17]. In the setting of clonal hematopoiesis of indeterminated potential (CHIP), mutations in *SRSF2* and *SF3B1* have been identified with a range of variant allele burden, including VAFs of >10% [7, 8, 61]. These observations would suggest that when present as the only driver mutation, spliceosomal mutations can confer advantages to the host cells. The development of murine models of human spliceosome mutations, both conventional knock-out models and humanized mutant models, has improved our understanding of how these mutations perturb normal hematopoiesis and ultimately promote malignancy. The establishment of tractable, autochthonous pre-clinical models has proven more challenging. The fidelity of the "humanized" murine models must be considered against the phenotypes associated with the same spliceosomal mutations in humans.

The independently generated "humanized" spliceosomal mutant mice described to date encompass a spectrum of phenotypes from very mild to overt MDS/MPN in the setting of native hematopoiesis (Summarized in Table 1). The majority of models have also reported analysis of bone marrow transplant recipients. This is a useful strategy to limit the expression of the mutation to the transplantable hematopoietic cells but is also a source of significant cellular stress. MDS, MDS/MPN and related cancers arise in the absence of this extreme stress in the vast majority of humans. Bone marrow transplant changes HSC clonal dynamics and contribution compared to native hematopoiesis and permanently damages and remodels the bone marrow microenvironment itself [42, 43, 62-65]. The use of transplant models also limits the capacity to understand the interplay of the mutant cells and a normal immune system, now appreciated to be an important aspect in cancer evolution and escape [66].

The models described demonstrate that the targeting strategy itself can have profound impacts on the utility of the model that is generated. The three models that have utilized an inverted exon approach have all observed ~30% variant allele frequency at the RNA level after locus recombination, demonstrating that despite genomic heterozygosity the expression from these modified alleles is not heterozygous. This caveat requires acknowledgement when these models are reflected against the human phenotypes

associated with these mutations. The reasons that these alleles do not express heterozygously are not clear, but the results are consistent across three independent models and loci, indicating that this effect is most likely resultant of the targeting strategy itself. The inverted exon model has been successful in establishing murine models harboring the Jak2(V617F) mutation, where the mice develop a myeloprofilerative neoplasm resembling polycythemia vera as occurs in humans with this mutation [67]. Whilst speculative, for alleles that provide a proliferative/survival advantage or are strongly positively selected the inverted exon model may be feasible as even sub-heterozygous expression may be sufficient to afford the cell with the advantage of the mutant allele.

Another consideration is how foreign elements, such as loxP sites, are placed within the endogenous gene structure. Whilst literature describing the systematic assessment of locus modifications is limited, some principals have emerged with the evolution of the gene targeting methods and from the more recent high-throughput targeting vector generation methods [68-72]. These include the placing of loxP elements relative to the promoter and intron/exon junctions (>200bp is generally advised); avoiding disruption of evolutionarily conserved regulatory elements and CpG islands that may be important for normal gene regulation; being cognizant of the intron and exon sizes when introducing new elements with caution where there are short introns/exons as this can affect splicing of the gene; and avoidance of modification to neighboring genes if possible. These generalized observations may be relevant to the alleles described for *Srsf2*, where there are only two coding exons and the mutation occurs in the first exon. There is a CpG island that encompasses the 5' promoter region and 5'UTR/exon 1, intron and exon 2 of *Srsf2* (UCSC genome browser mm10; chr11:116,852,080-116,853,903). The *Srsf2* allele described by Kon et al. [32], for example, used an inverted exon approach and was found to be transcriptionally inactive/silent when non-recombined. This allele involved significant modifications to the gene structure, including separation of the 5'UTR and exon 1 coding sequence and insertion of multiple recombination elements in the introns [32]. The model developed by Kim et al. [31], placed the 5'LoxP element 87bp from the start of the exon 1 containing sequence and modified the intron spacing significantly in both the pre and post-recombined configuration [31]. All of these modifications can potentially impact the expression of the locus.

In humans it is thought that spliceosome mutations arise within the HSC compartment [6, 73]. It is well defined experimentally that the cell of origin can impact the phenotypes in experimental leukemia models [74, 75], so the use of different Cre drivers that target distinct and/or overlapping cell populations may be important in the presentation of the phenotypes in these models. Two of the *Srsf2* and one of the *Sf3b1* alleles have been assessed using the constitutively active *Vav*-Cre [24, 32, 76]. *Vav*-Cre is active in the hematopoietic cells including HSCs, with evidence for activity also in endothelial cells and

some activity in the reproductive tissues [41, 77, 78]. Multiple groups have used the inducible *Mx1*-Cre, a broadly active Cre that is very efficient at recombining loxP flanked alleles in hematopoietic cells including HSCs as well as other tissues in the adult animal including the cells composing the bone marrow microenvironment [41, 79, 80]. *Mx1*-Cre requires administration of pIpC to induce an innate immune mediated interferon response, a known stimulus of cycling of HSCs [81, 82]. Whereas this can be used to assess native hematopoiesis, the spliceosome mutant models described to date using *Mx1*-Cre have predominantly reported phenotypic/functional analysis from recipients of bone marrow transplants, which were treated after recovery with pIpC. This approach is useful to restrict Cre activity in non-hematopoietic tissues as only the transplanted/engrafted bone marrow will express Cre [39, 83]. Such a strategy is effective for understanding the hematopoietic intrinsic effect of the mutation, however it introduces the stress of bone marrow transplant and changes the hematopoietic dynamics and bone marrow microenvironment as a result [42, 43, 62-65]. Coupled with the administration of pIpc 4-8 weeks after transplant this elicits substantive stress on the entire hematopoietic system. An alternative is to use CreER based models, where tamoxifen is required to be administered to activate Cre activity. We utilized both the *Rosa26*-CreER ^{T2} and *hScl*-CreER^{T} lines to allow widespread and relatively HSC restricted activation respectively, of *Srsf2*^{P95H} [30]. The CreER systems and the administration of tamoxifen are not benign, with evidence for toxicity of the CreER itself and tamoxifen administered at high doses subcutaneously can have direct effects on hematopoiesis [84, 85]. These effects can be reduced by oral administration of tamoxifen, shortening the duration of exposure to tamoxifen and by the use of heterozygous CreER alleles. The contribution of the specific Cre drivers to the phenotypes reported are likely secondary to the impacts of the locus specific genetic modifications themselves. However, a direct comparison of the same splicing mutation with different Cre strains would resolve this (for example *Rosa26*-CreER vs *Mx1*-Cre). In all situations, the parallel analysis of littermate Cre+ve wild-type and Cre-ve gene modified animals is an essential control for these studies, whether using *Mx1*-Cre or CreER systems.

The models of spliceosomal mutation described to date have demonstrated a range of phenotypes. Definitive reasons for this remain uncertain, with some interpretations citing a lack of conservation of splicing between species. Whilst the spliceosomal proteins themselves are highly evolutionarily conserved, this is not an insignificant consideration [16]. By virtue of the higher level of conservation of the exons between mouse and humans (89-94%), the exonic splice enhancer sequences are more highly conserved [86, 87]. Despite this, there is still uncertainty regarding the identity of disease relevant mis-spliced candidates in human samples and murine models. For example, in humans it was reported that missplicing of *EZH2* [31], and more recently *CASP8* [24, 88], occurred in *SRSF2* mutant

cells. For *Ezh2,* at least, this observation was not confirmed in multiple murine *Srsf2*$^{P95H/+}$ mutant models where other mis-spliced candidates identified in human samples could be confirmed [30, 32]. The phenotypic recapitulation and conserved disease evolution and progression that occured in the *Srsf2*$^{P95H/+}$ model that we described, indicates that the mouse is able to reproduce the core aspects of the human phenotypes associated with *SRSF2* mutation. However, the identification of the specific mis-spliced genes that cause disease development has remained elusive. There are several possibilities that could account for this. Firstly, that the transcriptomic analysis has yet to be performed on sufficiently purified disease initiating cells, in both human and mouse, and that the heterogeneity of the samples assessed to date masks identification of the relevant mis-spliced transcipts. Compounding this is that the computational analysis of splicing is yet to reach a consensus and the different computational methods yield distinct results. An alternative possibility is that subtle changes across many genes collectively contribute to the phenotype, rather than a small number of dramatically mis-spliced candidates. In such a possibility it will require highly purified samples and sufficient sequencing depth to confidently identify these. Another possibility is that mis-spling occurs within common cellular pathways in both species, but that the underlying individual genes are more variable and not necessarily conserved. An example of species conserved transcriptional and cellular consequence despite an apparent divergence of the individual transcripts can be seen in the A-to-I RNA editing field and in the phenoytpes of *ADAR1* mutation in human and mouse [89-91]. Unlike the exonic sequences bound by SRSF2, intronic sequences are more divergent between human and mouse, with 66.1% conservation of branchpoint sequences across mammals [92]. The co-ordinated recognition of the 5' and 3' splice sites together with brachpoint and polypyrimidine tract selection are critical to splicing fidelity. The utilization of highly conserved branchpoints is likely to be conserved to a greater extent across species than those with weaker recognition motifs, less conservation or alternative branchpoint spacing. These differences in the intronic sequences and motifs between species may contribute to the phenotypic differences between mouse and humans in particular for SF3B1 which binds the branchpoint within intronic sequence of the pre-mRNA.

Murine models of human disease associated mutations allow us to understand how these mutations individually contribute to disease initiation and maintenance. To provide this understanding, they must recapitulate as closely as possible the nature of the mutation, the expression of the mutant allele and the cell of origin. A critical assessment of the genetic modification strategies applied demonstrates that there are significant differences in the approaches that have been utilized. How the different targeting strategies and experimental approaches for individual gene mutations contribute to the phenotypes observed should be

taken into consideration along with the degree to which the model presents phenotypes that are consistent with the human disease carrying the same mutation.

Figure legends

Figure 1. A schematic outline of RNA splicing and the role of those proteins mutated in MDS. Within the intron of the pre-messenger RNA the U2 snRNP complex, containing SF3B1, binds the branchpoint site adenosine, U2AF2 binds to the polypyrimidine tract sequence and U2AF1 binds the 3' splice site; SRSF2 binds to the exonic splice enhancer sequence within the exonic sequence. *Cis*- and *trans*-acting factors can both positively or negative regulate splicing. Proteins where murine models have been generated are in bold font.

Figure 2 and 3 legend unchanged

Figure 1

Figure 1. Xu et al

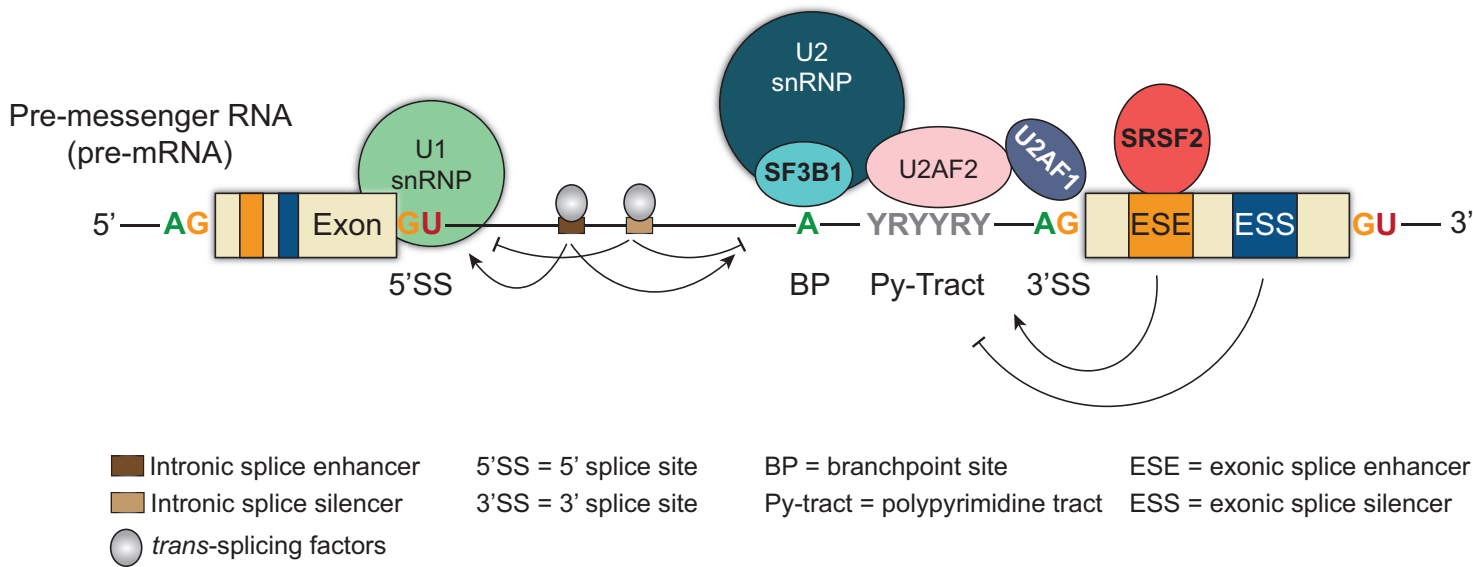


Figure 2

Figure 2. Xu et al

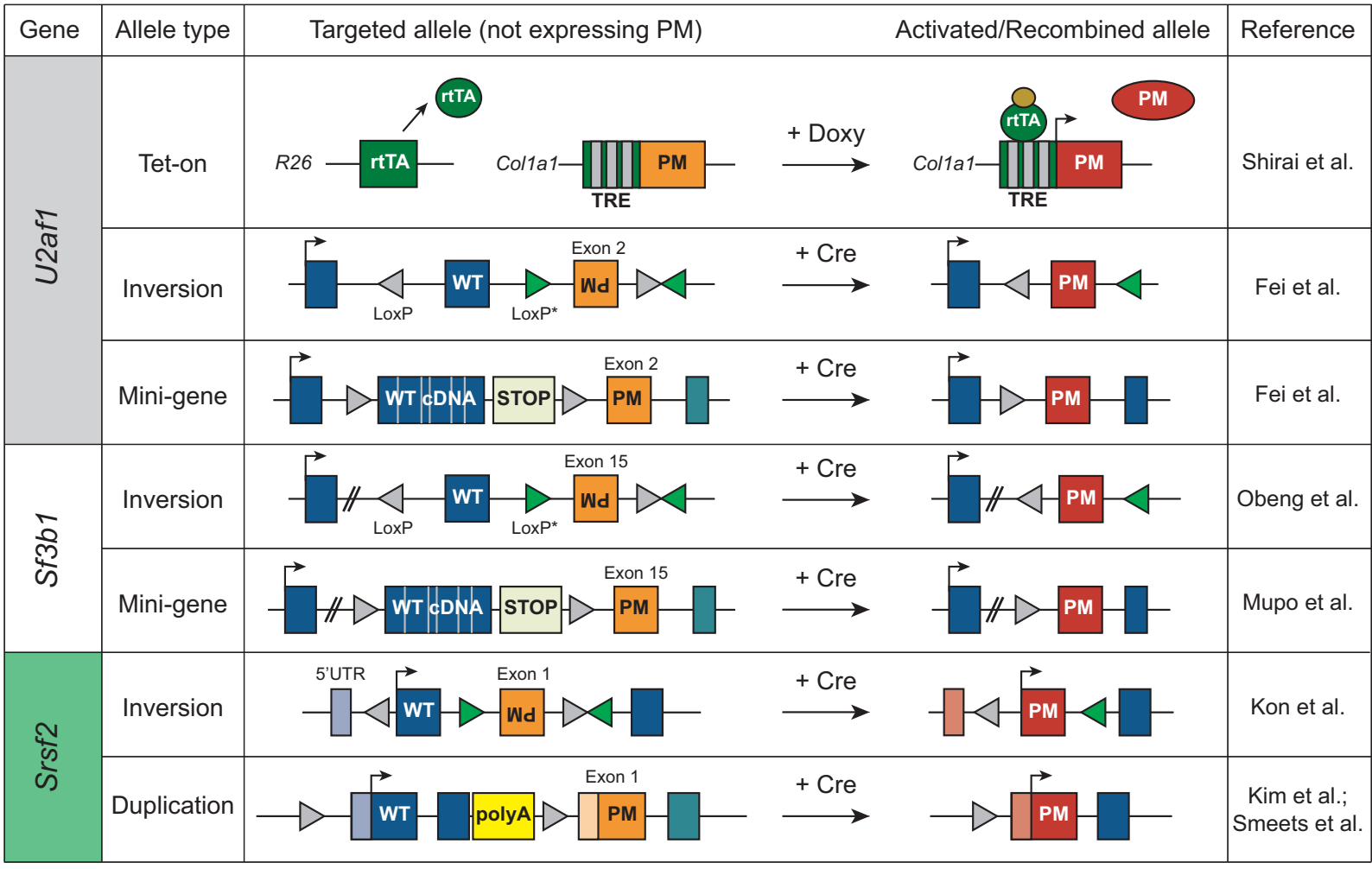
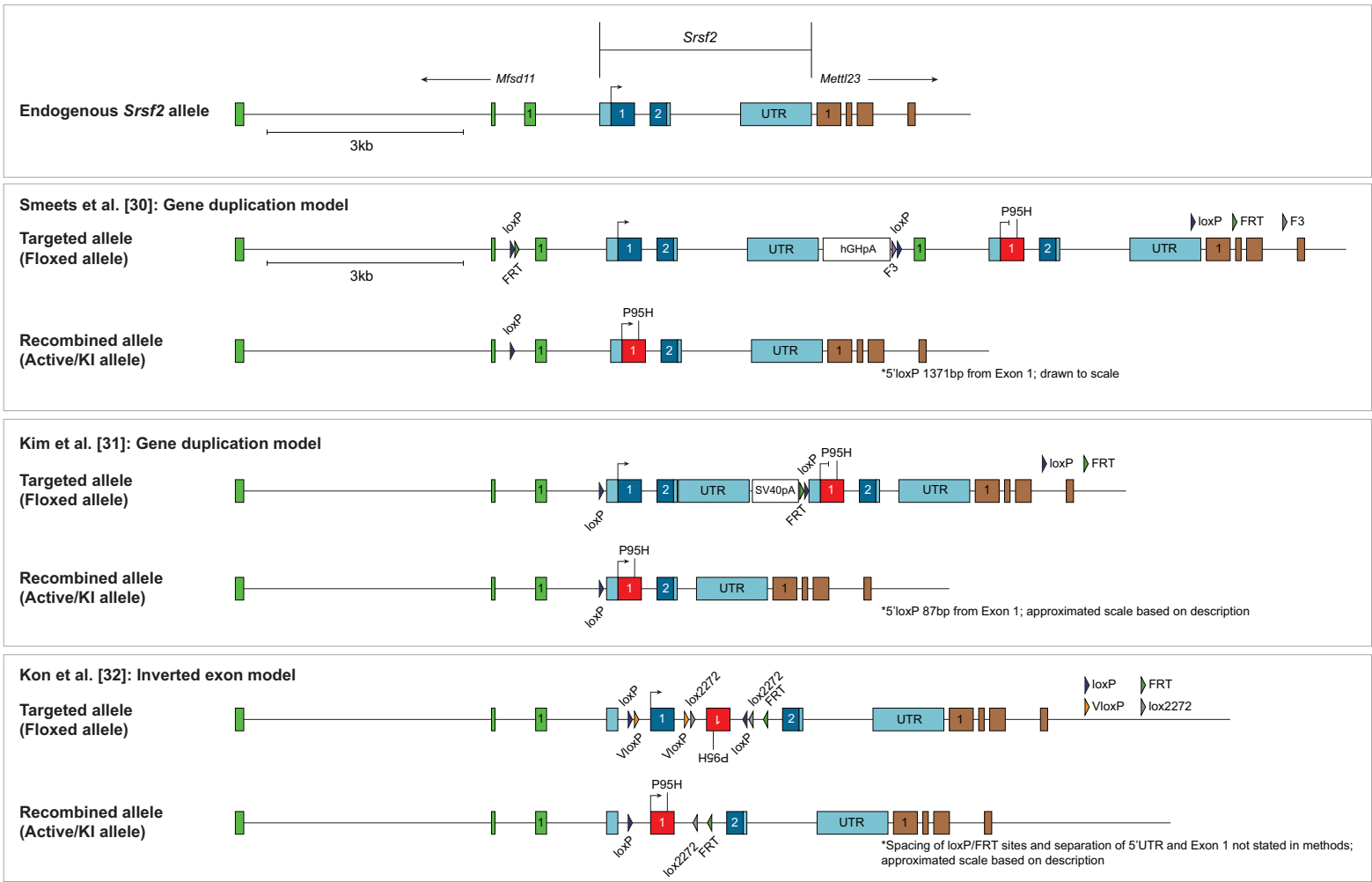


Figure 3



Xu et al., Figure 3_rev1