Integrative RNA-omics discovers *GNAS* alternative splicing as a phenotypic driver of splicing factor-mutant neoplasms

Emily C. Wheeler[1,2,3$], Shailee Vora[4,5,6,7$], Daniel Mayer[8], Andriana G. Kotini[4,5,6,7], Malgorzata Olszewska[4,5,6,7], Samuel S. Park[1], Ernesto Guccione[4,5,6], Julie Teruya-Feldstein[9], Lewis Silverman[5,7], Roger K. Sunahara[8], Gene W. Yeo[1,2,3*], Eirini P. Papapetrou[4,5,6,7*]

[1]Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA, USA; [2]Stem Cell Program, University of California San Diego, La Jolla, CA, USA; [3]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA; [4]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [5]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [6]Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [7]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [8]Department of Pharmacology, University of California San Diego School of Medicine, La Jolla, CA, USA. [9]Department of Pathology, Molecular and Cell-based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[$]These authors contributed equally

*Correspondence:

Eirini P Papapetrou, MD, PhD
One Gustave L. Levy Place
Box 1044A
New York, NY 10029
eirini.papapetrou@mssm.edu
Phone: 212-824-9337

Gene W Yeo, PhD
2880 Torrey Pines Scenic Drive
La Jolla, CA 92037
geneyeo@ucsd.edu
Phone: 858-246-1497

Running title: Mis-splicing of *GNAS* by mutant SRSF2 and U2AF1 drives MDS

Conflict of interest statement

GWY is co-founder, member of the Board of Directors, on the SAB, equity holder, and paid consultant for Locanabio and Eclipse BioInnovations. GWY is a visiting professor at the National University of Singapore. GWY's interest(s) have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. EPP has received honoraria from Celgene and Merck and research support from Incyte. JTF is on the advisory board of Astellas and Blueprint Medicines and a consultant for Curio Science, Histowitz and Scopio Labs. The authors declare no other competing financial interests.

1

Abstract

Mutations in splicing factors (SFs) are the predominant class of mutations in myelodysplastic syndrome (MDS), but convergent downstream disease drivers remain elusive. To identify common direct targets of mis-splicing by mutant U2AF1 and SRSF2, we performed RNA-Seq and eCLIP in human hematopoietic stem/progenitor cells (HSPCs) derived from isogenic induced pluripotent stem cell (iPSC) models. Integrative analyses of alternative splicing and differential binding converged on a long isoform of *GNAS* (*GNAS-L*), promoted by both mutant factors. MDS population genetics, functional and biochemical analyses support that *GNAS-L* is a driver of MDS and encodes a hyperactive long form of the stimulatory G protein alpha subunit, Gαs-L, that activates ERK/MAPK signaling. SF-mutant MDS cells have activated ERK signaling and consequently are sensitive to MEK inhibitors. Our findings highlight an unexpected and unifying mechanism by which *SRSF2* and *U2AF1* mutations drive oncogenesis with potential therapeutic implications for MDS and other SF-mutant neoplasms.

Statement of Significance

Splicing factor mutations are disease-defining in MDS, but their critical effectors remain unknown. We discover the first direct target of convergent mis-splicing by mutant U2AF1 and SRSF2, a long *GNAS* isoform, which activates G protein and ERK/MAPK signaling, thereby driving MDS and rendering mutant cells sensitive to MEK inhibition.

Introduction

Splicing factor (SF) gene mutations are the most common class of mutations in MDS, present in nearly 60% of MDS patients(1-3). SF mutations are early, sometimes initiating, events in the course of the disease and can thus provide attractive therapeutic targets(4). Canonical SF mutations in MDS affect specific hotspots in 3 main genes that regulate splicing, *SRSF2*, *U2AF1* and *SF3B1*. In the past decade since their discovery, hundreds of splicing alterations have been catalogued in MDS patient cells and in murine and cellular models of SF mutations, in search of those critical to the disease pathogenesis. Transcriptome-wide splicing analyses support that mutant SFs exhibit altered mRNA sequence binding preferences, compared to the wild-type (WT) factors(5-9). However, the specific target transcripts and mechanisms by which these mutations drive MDS remain poorly understood.

SF mutations are always heterozygous and mutually exclusive to one another. While synthetic lethality between SF mutations has also been demonstrated(10), convergence in common downstream targets at the isoform, gene or cellular pathway level, has been hypothesized to underlie their mutual exclusivity. In particular, convergence at the isoform level appears more likely in the case of *SRSF2* and *U2AF1* mutations, as both primarily affect exon usage(5, 11-13). Serine/arginine-rich splicing factor 2 (SRSF2) regulates splicing by promoting inclusion of exons through binding to exonic splicing enhancer (ESE) sequences. Mutant SRSF2 preferentially recognizes 'CCNG'- versus 'GGNG'- containing ESEs, whereas WT SRSF2 binds to both with similar affinity *in vitro*(5, 6). U2 small nuclear RNA auxiliary factor 1 (U2AF1) recognizes the AG splice acceptor dinucleotide at the 3' end of introns. Mutant U2AF1 shows preference for 3' splice sites containing 'CAG*'* or 'AAG*'* - versus 'UAG*'*- sequences, promoting preferential inclusion of the respective exons(7-9).

Efforts to identify common effectors of SF mutations in RNA-sequencing datasets from MDS patient cells and research models have revealed largely non-overlapping changes(11, 12). These datasets have important limitations: primary patient cells are characterized by substantial heterogeneity due to co-occurring mutations and variable clonal composition; alternative splicing (AS) events are largely non-conserved between mouse and human(14); and cellular models engineered in aneuploid leukemia cell lines express mutant SFs at non-physiological levels and stoichiometry that may alter their binding and associated transcript changes.

Here, to identify splicing alterations common to *SRSF2* and *U2AF1* mutations that are directly caused by altered binding of the mutant factors, we developed isogenic CRISPR gene-edited iPSC models of the two mutations that display MDS-related hematopoietic phenotypes. We characterized the transcriptome-wide splicing alterations and RNA binding (eCLIP) of the mutant and WT factors in iPSC-HSPCs. Integrated analyses revealed increased usage of a long isoform of *GNAS* (*GNAS-L*), encoding the α subunit of the stimulatory G protein (Gαs), as an altered splicing event common to both mutations in iPSC-HSPCs and MDS patient cells. We provide functional evidence for a role of *GNAS-L* in driving the MDS phenotype. We further show that *GNAS-L* encodes a more active Gαs protein (Gαs-L). While oncogenic mutant Gαs (R201) activates the cAMP/PKA signaling pathway, we find that Gαs-L results in ERK pathway activation.

These findings reveal activation of ERK/MAPK signaling by a more active Gαs form produced by alternative splicing as a common effector of *SRSF2* and *U2AF1* mutations in MDS and suggest MEK inhibition as a potential therapeutic option for SF-mutant neoplasms.

3

Results

Human HSPCs from isogenic iPSC models of SF mutations recapitulate cellular and molecular phenotypes of SF-mutant MDS

To interrogate the effects of *U2AF1* and *SRSF2* mutations and identify common downstream effectors driving MDS, we used CRISPR/Cas9 to introduce a heterozygous *U2AF1* S34F mutation in a normal iPSC line (N-2.12)(15) (Fig. 1A and B, Supplementary Fig. S1A-S1C). Multiple mutant iPSC lines were generated with two different gRNAs to exclude potential confounding effects of off-target events on cellular and molecular phenotypes (Fig. S1B, Supplementary Table S1). A heterozygous *SRSF2* P95L mutation was previously introduced in the same normal parental iPSC line(16). Homozygous mutant clones were not recovered, in agreement with the previously demonstrated dependency of SF-mutant cells on one WT allele (Supplementary Fig. S1D)(17). To enable allele-specific immunoprecipitation, an epitope tag (3xFLAG) was introduced, in a subsequent gene editing step, at the C-terminus of the endogenous locus of either the mutant allele (in SF-mutant lines) or one WT allele (in the isogenic SF-WT iPSC lines) (Fig. 1A, Supplementary Fig. S2A-J). We confirmed that addition of the epitope tag did not affect expression of the tagged allele and that all lines maintained a normal karyotype (Supplementary Fig. S2).

To assess the effects of the two mutations in hematopoiesis, we employed a human pluripotent stem cell differentiation protocol that produces definitive-type hematopoietic progenitor cells. Directed differentiation of independent lines of each genotype revealed no defects in hematopoietic specification, as indicated by the emergence of CD34+ and CD45+ HSPCs (Supplementary Fig. S3A and S3B). In contrast, the number of hematopoietic colonies generated from SF-mutant iPSC-HSPCs in methylcellulose assays was reduced to approximately half of that generated from WT lines (Fig.1C). SF-mutant iPSC-HSPCs showed severely impaired ability for myeloid maturation (Fig. 1D), as well as decreased proliferation (Fig. 1E). Decreased growth has previously been observed in various cell line and mouse models of SF mutations(16, 18). Additionally, SF-mutant iPSC-HSPCs had decreased viability and showed dysplastic erythroid and myeloid lineage morphologic alterations (Supplementary Fig. S3C and S3D). Collectively, these results show that SF mutations cause differentiation, proliferation and viability defects, as well as dysplastic morphologic changes, recapitulating hallmark phenotypes previously reported by us and others in MDS patient-derived iPSCs and *ex vivo* primary MDS cells(15, 16, 19-21).

To evaluate the impact of SF mutations on RNA splicing, we performed RNA-Seq analyses in sorted CD34+/CD45+ iPSC-HSPCs from at least 3 independent lines per genotype and identified AS events (Fig. 1F, Supplementary Fig. S4, Supplementary Table S1). Cassette exon events were the predominant AS event type in both genotypes, while alternative 3' and 5' splicing events (A3SS, A5SS) were detected at a lower frequency (Fig. 1G, Supplementary Fig. S5A-C, Supplementary Table S2). *'CAG'* and *'AAG'* 3' splice site sequences were enriched in flanking exons preferentially included in *U2AF1*-mutant cells ("S34F exons"), while exons preferentially skipped in U2AF1-mutant cells ("WT exons") were enriched for *'UAG'* 3' splice site sequences (Fig. 1H and 1I). The same motif preferences were also identified at the 3' splice site of A3SS events preferentially included ('CAG/AAG') and skipped ('UAG') in U2AF1-mutant compared to WT cells (Supplementary Fig. S6A). Exons preferentially skipped in *SRSF2*-mutant cells ("WT exons") showed enrichment of 5-mers containing 'GGAG' and 'GGUG' sequences, while 5-mers containing 'GCAG' and *'CCAG'* were enriched in exons preferentially included in *SRSF2*-mutant cells ("P95L exons") (Fig. 1J and 1K, Supplementary Fig. S6B-E). These sequence-specific splicing alterations are consistent with previous findings in other models and

patient cells and establish that our iPSC models recapitulate the characteristic splicing alterations of SF-mutant hematopoietic cells(5-8).

Allele-specific eCLIP identifies bona fide targets of mis-splicing by mutant SRSF2 and U2AF1

To assess the transcriptome-wide protein-RNA interaction landscape of mutant vs WT factors, eCLIP(22) was performed in sorted CD34+/CD45+ iPSC-HSPCs from two independent iPSC lines per genotype (Fig. 1F, Supplementary Table S1) with allele-specific immunoprecipitation of the epitope-tagged mutant (in *SRSF2*-mutant and *U2AF1*-mutant lines) or WT (in isogenic *SRSF2*-WT and *U2AF1*-WT lines) allele (Supplementary Fig. S7A-S7D). We identified ~35,000 U2AF1 binding sites and ~12,000 SRSF2 binding sites across the protein-coding transcriptome, the majority of which occurred within the expected RNA regions: 3' splice sites for U2AF1 (both WT and mutant) and protein-coding exonic regions for SRSF2 (both WT and mutant) (Fig. 2A, Supplementary Fig. S7E-S7G). To identify binding sequence preferences in the binding sites of each mutant or WT factor, we performed motif enrichment analyses and calculated the relative abundance of all 6-mers within peak regions (Fig. 2B-2D). WT U2AF1 binding sites were enriched for 6-mers containing '*UAG*', while U2AF1-S34F binding sites were enriched for 6-mers containing '*CAG*' (Fig. 2B and 2C, Supplementary Fig. S7H). WT SRSF2 binding sites were enriched for '*GC*' and '*GA*'-rich 6-mers, while peaks bound by SRSF2-P95L contained no significantly enriched motifs and displayed a modest enrichment for 6-mers containing '*CCUG*' sequences (Fig. 2B and 2D, Supplementary Fig. S7I). These sequence preferences in binding between the WT and mutant forms of U2AF1 and SRSF2 mimic the sequence alterations that we found in skipped exon events by RNA-Seq and support a causative link between differential RNA sequence recognition by mutant SFs and altered splicing.

To specifically interrogate differential binding in regulatory regions of skipped exons (upstream intron for U2AF1 and cassette exon for SRSF2), we generated 'RNA splicing maps' to visualize position-specific SF binding(23) (Fig. 2E and 2F, Supplementary Fig. S8A and S8B). The binding density of WT U2AF1 was higher than that of U2AF1-S34F specifically at the 3' splice site of "WT exons" (included in WT cells) (Fig. 2E, Supplementary Fig. S8A, green line). Conversely, U2AF1-S34F had higher binding density than WT U2AF1 at the 3' splice site of "S34F exons" (included in mutant cells) (Fig. 2E, Supplementary Fig. S8A, blue line). We also observed increased binding density of WT SRSF2 within "WT exons", but not within upstream or downstream control exons (Fig. 2F, Supplementary Fig. S8B). These results show differences in binding density between the WT and mutant SFs specifically at the regulatory regions of skipped exons, which allow us to couple altered splicing to differential binding by mutant vs WT U2AF1 or SRSF2 at the level of individual AS events.

To this end, we grouped the binding events in regulatory regions of skipped exons (deltaPSI > 5%, FDR < 0.05) into 3 categories: (1) sites bound by the WT SF only ("WT peaks"); (2) sites bound by the mutant SF only ("Mutant peaks"); or (3) sites bound by both WT and mutant SF ("Both peaks") (Fig. 3A). We found that "WT exons" (positive delta PSI WT–S34F) more often contained "WT peaks", and "S34F exons" (negative delta PSI WT–S34F) more often contained "Mutant peaks" (chi-square test, p=0.003). This trend was similar, albeit less strong, in the case of SRSF2, in which the overall number of peaks detected was much lower (Fig. 3A) and was not observed in downstream, control exons (Supplementary Fig. S8C). We expanded the analysis to all observed splicing changes (not just those above significance thresholds), and observed that exons with a corresponding WT or mutant-specific binding event trend towards inclusion in

5

cells of the genotype with a corresponding WT or mutant factor-specific binding event (Supplementary Fig. S8D and S8E).

With these datasets, we were able to assign direct, *in vivo* binding changes between WT and mutant SFs at both novel AS events and events previously reported in association with *U2AF1* or *SRSF2* mutations. For example, a skipped exon event in *MED24*, previously shown to be preferentially included in *U2AF1-* and *SRSF2-* mutant cells(5, 13), contains a binding site specific to both mutant SRSF2 and U2AF1 at the regulated exon (Fig. 3A, Supplementary Fig. S9A and S9B). In contrast, we did not find any direct evidence for differential binding of any factor at a previously reported *EZH2* poison exon(5) (Supplementary Fig. S9C and S9D).

Integrative analysis of mis-splicing and mRNA binding reveals a long isoform of *GNAS* (*GNAS-L*), promoted by both mutant factors, as a phenotypic driver of MDS

Convergence in common downstream genes has been postulated to underlie the mutual exclusivity of SF mutations in MDS patients. We thus harnessed our isogenic model to find altered splicing events common to both SF mutations. Of the 504 and 517 differential AS events that we identified in *U2AF1* S34F and *SRSF2* P95L, compared to isogenic WT cells, respectively, 41 events, affecting 40 genes, were common to both genotypes (Fig. 3B, Supplementary Table S3). We then quantified exon inclusion (delta PSI) for each of these events in a publicly available MDS patient dataset(12). 20 of the 41 events had sufficient read coverage to allow evaluation of splicing. 15 of those 20 were differentially spliced in the same direction in both *U2AF1-* and *SRSF2-* mutant cells (Fig. 3B and 3C). Of these, 3 AS events, in genes *GNAS*, *PSMA4* and *ITGB3BP*, also contained a differential eCLIP peak between S34F and WT U2AF1 at the 3' splice site of the alternatively spliced exon (Fig. 3B and 3C).

Of those 3 genes, *GNAS* (Guanine Nucleotide binding protein, Alpha Stimulating), the gene encoding the α subunit of the stimulatory G protein (Gαs), is a recurrently mutated gene in MDS(24). Furthermore, G protein signaling plays important roles in multiple cellular functions and has been previously linked to oncogenesis in non-hematologic malignancies(25). Both SF mutations promoted the inclusion of exon 3, resulting in the preferential usage of a long *GNAS* isoform (*GNAS-L*), over a short isoform (*GNAS-S*), in both iPSC-HSPC models and in MDS patient cells, giving rise to a longer form of the Gαs protein (Gαs-L) (Fig. 3D-3I, Supplementary Fig. S9E, Supplementary Fig. S10A and S10B). Selective knockdown of *GNAS-L* with two different shRNAs specifically targeting exon 3 rescued the differentiation defect of SF-mutant iPSC-HSPCs and restored it to a level comparable to that of WT iPSC-HSPCs (Fig. 4A and 4B). This effect was specific to the SF-mutant cells, as *GNAS-L* knockdown had no effect in the differentiation of WT iPSC-HSPCs (Supplementary Fig. S10C). Conversely, overexpression of *GNAS-L* in WT iPSC-HSPCs decreased their differentiation potential and viability by approximately half, reproducing the SF-mutant phenotype (Fig. 4C, Supplementary Fig. S10D). Overexpression of *GNAS-S* in SF-mutant iPSC-HSPCs did not improve their colony formation potential (Supplementary Fig. S10E). These results establish that *GNAS-L* is a phenotypic driver of MDS.

6

GNAS-L encodes a hyperactive form of Gαs (Gαs-L), non-redundant to the Gαs R201 mutant form

Gαs is a member of the heterotrimeric family of G proteins that are activated by G protein-coupled receptors (GPCRs). In its inactive state, Gα is bound to GDP. Its activation requires release of GDP and exchange for GTP. Activating *GNAS* mutations, most commonly involving the R201 hotspot, are found in ~1% of MDS patients and in other tumors(24, 25). We thus hypothesized that the *GNAS-L* isoform, promoted by SF mutations, also encodes a hyperactive form of Gαs (Gαs-L). Gα contains two domains, the Alpha Helical Domain (AHD) and the guanine-nucleotide-binding, Ras Homology Domain (RHD). RHD-AHD domain separation, resulting in an "open" conformation, is necessary for G protein activation, i.e. GDP dissociation and nucleotide exchange(26). Exon 3 encodes a 15 amino acid segment in Gαs-L within a hinge-like region located between the AHD and the RHD. Although well-ordered on most other G protein α subunits, the hinge-like region is disordered (i.e. unstructured) in the Gαs crystal structures, whether bound to nucleotide or not(27, 28) (Fig. 4D). To ascertain the role of exon 3, we analyzed the intrinsic nucleotide exchange capacity of purified Gαs-L by measuring [^{35}S]GTPγS or Bodipy-FL GTPγS binding. Gαs-L bound GTPγS faster than Gαs-S (Fig. 4E and 4F, Supplementary Table S4). Furthermore, ectopic expression via transfection of Gαs-S or Gαs-L in HEK293 cells engineered to not express endogenous *GNAS*(29), showed that Gαs-L displays modestly higher maximal activity than Gαs-S upon stimulation with the β$_2$-adrenergic receptor agonist isoproterenol (Fig. 4G and 4H, Supplementary Table S5). We speculated that differences in activity between Gαs-L and Gαs-S may be amplified in the R201 mutant background. We thus measured [^{35}S]GTPγS and Bodipy-FL GTPγS binding and cAMP accumulation of purified R201C and R201H Gαs-L or Gαs-S forms (Fig. 4E-4H, Supplementary Table S4 and S5). These data show a significant difference in the basal levels of G protein activation and adenylyl cyclase activity of the mutant Gαs-L, compared to the Gαs-S, form. Collectively, these results demonstrate that the long Gαs form is more active than the short form.

We then hypothesized that *GNAS* activation by the long isoform and *GNAS* activation by the R201 mutation might be functionally redundant and that this redundancy may manifest as mutual exclusivity between SF mutations (promoting *GNAS-L* expression) and *GNAS* point mutations in MDS patients. To test this, we interrogated a large cohort of 3,222 MDS and acute myeloid leukemia (AML) patients – combined from 3 published cohorts(4, 30, 31) – for signals of genetic interaction between SF mutations and *GNAS* mutations. Unexpectedly, these analyses revealed significant co-occurrence between *SRSF2* and *GNAS* mutations (p=0.01; OR=4.05) (Fig. 5A), strongly suggesting that the effects of these mutations are not redundant, but potentially synergistic. To test the hypothesis that this co-mutation pattern is driven by cooperation of the GNAS-L isoform promoted by SF mutations with the GNAS R201 mutation, we transduced WT iPSC-HSPCs with lentiviral vectors expressing GNAS-L, mutant GNAS R201H or both (GNAS L-R201H) and quantified their effects in colony forming ability. Both GNAS-L and the R201H mutation suppressed colony formation and their combination (GNAS L-R201H) trended towards a more pronounced effect (Supplementary Fig. S10F). Collectively, these results show that inclusion of exon 3, giving rise to the GNAS-L isoform, produces a Gαs protein that is hyperactive. Furthermore, our data suggest that the increased activity of Gαs-L is not equivalent to that of the R201 mutant form, but is rather non-redundant and possibly synergistic to R201 mutant Gαs.

Gαs-L activates ERK/MAPK signaling

Activated Gαs directly activates adenylyl cyclase, the enzyme responsible for producing cAMP. cAMP is the second messenger that activates, among other things, protein kinase A (PKA). We thus investigated the signaling pathways activated by Gαs-L in K562 cells, iPSC-HSPCs, and cord blood (CB) CD34+ cells transduced with lentiviral vectors expressing WT or R201 GNAS-L or GNAS-S. These experiments collectively showed that the R201 mutation (present on either the long or short GNAS form) activates canonical cAMP/PKA signaling. In contrast, we found no additional PKA activation by the GNAS-L form, with or without the R201 mutation (Fig. 5B and 5C, Supplementary Fig. S11A and S11B) or any evidence of increased PKA activation in *SRSF2 P95L* or *U2AF1 S34F* iPSC-HSPCs or primary cells (Fig. 5D and 5E).

We thus hypothesized that the Gαs-L form may result in distinct signaling outcomes than the R201 mutant form. Since many GPCRs potently regulate ERK activity(32, 33), we hypothesized that an increase in Gαs-L may affect ERK pathway activation. Indeed, we found that ectopic Gαs-L expression, with or without the R201 mutation, in WT iPSC-HSPCs and CB CD34+ cells, resulted in ERK1/2 and AKT activation (Fig. 6A and 6B, Supplementary Fig. S11C and S11D). In contrast, the R201 mutation alone in the short isoform did not activate ERK or AKT signaling. Additionally, we found increased ERK activation in *SRSF2 P95L* and *U2AF1 S34F*, compared to WT, iPSC-HSPCs (Fig. 6C, Supplementary Fig. S11E), as well as increased levels of dual-specificity phosphatase 6 (DUSP6), a classical ERK target gene (Fig. 6C). Forced expression of *GNAS-L* activated RAF and MEK and *SRSF2 P95L* and *U2AF1 S34F* iPSC-HSPCs also had increased RAF and MEK activation, compared to WT iPSC-HSPCs (Fig. 6C and 6D, Supplementary Fig. S11F and S11G). Specific shRNA knockdown of GNAS-L resulted in striking decrease in pERK in both *SRSF2 P95L* and *U2AF1 S34F* iPSC-HSPCs (Fig. 6E). Furthermore, primary MDS and AML cells with *U2AF1* or *SRSF2* mutations had increased Gαs-L form and ERK, but not PKA, activation, compared to SF-WT cells (Fig. 6F, Supplementary Fig. S11H and S11I, Supplementary Table S6), while primary MDS and AML patient cells with both *SRSF2 P95L* and *GNAS R201H* mutations showed increased cAMP and PKA substrate phosphorylation, compared to patient cells with *SRSF2 P95L* alone (Fig. 5E, Supplementary Fig. S11J). Collectively, these results show that, while the R201 *GNAS* mutation activates the canonical cAMP/PKA pathway, the long isoform activates the ERK/MAPK pathway.

Given the ERK/MAPK activation status of SF-mutant cells we uncovered here, we next wanted to test the dependency of SF-mutant cells on ERK/MAPK signaling. Mitogen-activated protein kinase kinase (MEK) is immediately upstream of ERK1/2 in the MAPK pathway. Several MEK inhibitors are currently US FDA-approved oncology drugs for melanoma and other solid tumors. To test the sensitivity of SF-mutant cells to MEK inhibitors we treated *U2AF1* S34F, *SRSF2* P95L and isogenic WT iPSC-HSPCs with 4 MEK inhibitors, US FDA-approved or currently in clinical testing: trametinib, cobimetinib, selumenitib and CH5126766. *U2AF1*- and *SRSF2*-mutant cells showed marked sensitivity to all 4 MEK inhibitors (but not to a mutant-BRAF inhibitor, vemurafenib, as control) (Fig. 7A-7C). Importantly, ex vivo cultured primary cells from MDS and sAML patients with SRSF2 P95 mutations were more sensitive to MEK inhibition than cells from patients without SF mutations (Fig. 7D). These results suggest that targeting the MEK/ERK pathway holds therapeutic promise for MDS patients with SF mutations.

Discussion

Here we developed a genetically faithful, human, isogenic, karyotypically normal iPSC-based model of SF-mutant MDS, with which we quantified alternative splicing and *in vivo* mRNA binding. We identify preferential usage of the *GNAS-L* isoform as a convergent consequence of *U2AF1 S34F* and *SRSF2 P95L* mutations. We further provide evidence that *GNAS-L* is an MDS phenotypic driver and that it mediates ERK/MAPK pathway activation.

Two key features of our study allowed us to pinpoint *GNAS* as a high-priority direct target common to both mutant factors: the isogenic conditions that empowered the identification of convergent targets of both *U2AF1* and *SRSF2* mutations; and the high quality eCLIP-Seq data in relatively homogeneous iPSC-derived HSPC populations and faithful genomic context (diploid human genome with one SF-WT and one mutant allele), which allowed us to couple altered exon inclusion/exclusion with altered binding of the mutant factors. We were, thus, able to identify specific exons that are both differentially bound by the mutant SFs and preferentially included or excluded in the mutant cells. One such event, resulting in increase of the *GNAS-L* isoform, was functionally validated to mediate the differentiation defect observed in our MDS model. Although our differentiation protocol generates definitive-type hematopoiesis, iPSC-derived blood cells may resemble fetal cells more than adult cells. Nonetheless, we and others have shown that iPSC models of myeloid malignancies capture phenotypic and molecular characteristics of disease and can be used to discover new disease mechanisms and therapeutic vulnerabilities, as highlighted by the present study (15, 16, 19, 34-40). Previous studies that used previous versions of CLIP technology, without size-matched input controls, to interrogate binding of mutant SRSF2 or U2AF1 transfected into immortalized cell lines did not observe a correlation between mis-splicing events and binding at their regulatory regions to pinpoint specific splicing events directly coupled with altered binding(41-43). Furthermore, while the *GNAS* AS event we report here was present in previous RNA-Seq datasets of *U2AF1*-mutant or *Srsf2*-mutant human and murine cells(13, 44-46), it was not prioritized for follow up in previous studies, in the absence of evidence of differential binding.

Since its discovery in the 1980s, two major splice forms of Gαs were identified, but no evidence of different functional consequences contributing to disease in humans existed prior to this study. A few earlier reports suggested differences between these splice forms in terms of nucleotide affinities, receptor coupling and cyclase activation, while others suggested that the two splice variants are functionally equivalent(47-50). Our biochemical and functional studies, together with MDS patient population genetics data, show that the Gαs form encoded by the long isoform is more active than the short and that this enhanced activity is non-redundant and possibly synergistic to Gαs activation by hotspot mutations that primarily activate the canonical cAMP/PKA effector pathway.

Activation of Gαs upon ligand binding to cell surface GPCRs can result in complex signaling outcomes through interactions of Gαs with a diverse array of signaling partners. Gαs activation promotes functional dissociation of the Gαs and Gβγ subunits, which can also recruit proteins to the plasma membrane and produce signaling responses. Previous links have been made between G protein signaling and the modulation of ERK/MAPK pathway, both stimulating and inhibiting, at different cellular contexts(51). The mechanism by which G proteins modulate ERK activation is not well understood. Our data show activation of RAF and MEK upstream of ERK (Fig. 6A and 6C and 6D, Supplementary Fig. S11F and S11G). This is also consistent with our finding that SF-mutant cells are sensitive to the MEKi CH5126766, which inhibits RAF-bound MEK(52) (Fig. 7). It is currently unclear how this activation occurs and whether it involves direct

9

engagement of RAF-MEK by Gαs-L, indirect activation through dissociation of the Gβγ subunits or other mechanisms. Dual activation of cAMP/PKA and ERK/MAPK signaling could mediate the cooperative effects of the R201 mutation with the long form of Gαs and provide a mechanistic basis for the co-occurrence of *SRSF2* and *GNAS* mutations that we report. Alternatively, the contribution of other mis-spliced targets of mutant SRSF2 and/or U2AF1 may underlie this cooperation. While specific knockdown of *GNAS-L* completely rescued the colony formation deficit of SF-mutant cells, additional splicing alterations may also contribute to the disease pathogenesis and phenotypic manifestations that cannot be detected with our system and assays. The continued study of the effects of SF mutations will shed more light into the relative contributions of different splicing alterations as drivers of MDS and other SF-mutant neoplasms.

We did not find conclusive evidence of increase in the *GNAS-L* isoform in MDS with *SF3B1* or *ZRSR2* mutations (Supplementary Fig. S10A and S10B). Since SRSF2 and U2AF1 both primarily affect exon inclusion/exclusion, convergence in a common exon event was a priori more likely for *SRSF2* and *U2AF1* mutations. The mutual exclusivity of *SF3B1* and *ZRSR2* mutations could still be explained by convergence at the level of the pathway or cellular processes affected. Future studies investigating the signaling consequences and target gene expression changes induced by increased *GNAS-L* in detail may point to candidate genes and isoforms that may be altered by these other mutations. Alternatively, other, as yet unidentified, targets of mis-splicing common to all mutant factors may exist or *SF3B1* and *ZRSR2* mutations may not converge in common targets with *SRSF2* and *U2AF1* mutations and their mutual exclusivity may be due to common effects in cellular function that are ultimately redundant in the pathogenesis of MDS or due to synthetic lethality alone(10).

MDS is a disease with poor prognosis and few therapeutic options and thus presents a high unmet clinical need(53). Activation of MAPK signaling by other mechanisms has been documented in AML, primarily in AML subsets with RAS and FLT3 mutations(54, 55). MDS has also been linked to RAS/MAPK pathway activation and inhibitors of RAS and MAPK signaling are in preclinical and clinical testing(56, 57). Furthermore, germline RAS pathway mutations cause inherited syndromes, such as Noonan syndrome and neurofibromatosis type 1, that predispose to development of MDS and related neoplasms (such as chronic myelomonocytic leukemia, CMML). The MEKi trametinib showed limited efficacy as a single agent in a phase 1/2 clinical study in relapsed/refractory AML and high-risk MDS, however most patients enrolled in this study were heavily pretreated and preselected for RAS (NRAS and KRAS) mutations, which typically occur late in disease progression(58). Our results suggest that MEK inhibitors may be a viable therapeutic option for MDS patients with *SRSF2* and *U2AF1* mutations at an earlier disease stage, a finding that could be readily translated, as these inhibitors are FDA-approved for other cancers and could be easily repurposed(59). Since toxicity is frequently a limiting factor with MEK inhibitors, direct targeting of Gαs or even Gαs-L might provide an alternative therapeutic option with potentially more favorable toxicity profile, although this remains to be tested. While there are currently no small molecule inhibitors of Gαs, approaches leveraging specific splicing modulation of *GNAS* exon 3 or proteolysis targeting chimeras (PROTACs) can be envisioned in the future.

The importance of the findings we report here may extend beyond MDS and AML, as *U2AF1* and *SRSF2* mutations are also present in CLL and some solid tumors(9). It remains to be seen whether other cancers with SF mutations are driven by MAPK activation through alternative splicing of *GNAS*. Interestingly, activating *GNAS* mutations in solid cancers are primarily found in low-grade tumors and benign metaplasias, which indicates a role for G protein signaling in

early steps of oncogenesis(60, 61). Since MDS is also a preleukemic condition with a low proliferation index akin to that of low-grade solid tumors, it is possible that Gαs activation is a common feature of low-grade malignant and premalignant cells and, as such, its targeting may offer new opportunities for cancer prevention.

Acknowledgements

12

Methods

CRISPR/Cas9 gene editing of human iPSCs

We used the previously described normal iPSC line N-2.12-D1-1 as the parental line(19). Generation of SRSF2 P95L iPSC lines was previously described(15, 16). The *U2AF1* S34F mutation and all allele-specific 3xFLAG tags were introduced using CRISPR/Cas9-mediated homology directed repair (HDR) in two steps.

To introduce the *U2AF1* S34F mutation, two different gRNAs targeting the *U2AF1* locus within exon 3 (cutting site between 1 bp and 10 bp from the 101C>T mutation site, sequences shown in Supplementary Fig. S1A) were designed, assembled by a two-step overlapping PCR reaction downstream of the U6 promoter sequence and cloned in the gRNA/Cas9 plasmid, also expressing Cas9 driven by the CMV promoter linked to mCitrine with a P2A(15, 16). Two sets (one for each gRNA) of two donor DNA plasmids, one containing the S34F mutation and one the corresponding wild-type (WT) sequence, containing 5' (1032 bp) and 3' (1064 bp) homology arms consisting of nucleotides 43104347 – 43105378 and 43103283 – 43104346 (hg38 human genome assembly), respectively, were constructed. The donor plasmids also contained silent mutations (shown in Supplementary Fig. S1A) to introduce a new restriction site sequence (SphI) and to prevent further cleavage by Cas9. The entire 5'+3' homology sequence was amplified from N-2.12-D1-1 genomic DNA and the c.101C>T and/or silent mutations to introduce new restriction enzyme recognition sites and to prevent cleavage by Cas9 were introduced by two-step overlapping PCR before subsequent cloning into the donor plasmid.

To generate *U2AF1* S34F iPSCs, the N-2.12-D1-1 iPSC line was cultured in hESC media containing 10 mM Y-27632 for at least one hour before nucleofection. The cells were dissociated into single cells with accutase and 1 million cells were used for nucleofection with 5 µg of gRNA/Cas9 plasmid and 5 µg of each donor plasmid (WT and S34F) using Nucleofector II (Lonza) and program B-16. Immediately after nucleofection the cells were replated on MEFs. mCitrine+ cells were FACS-sorted 48 hr after transfection and plated as single cells at clonal density (1000 FACS-sorted cells per 60-mm dish). After 10-12 days, single colonies were picked in separate wells of a 6-well plate, allowed to grow for approximately 3-6 days and screened by PCR. 1-3 medium-sized colonies from each individual clone were picked directly into a 0.2 ml tube, pelleted and lysed. Restriction Fragment Length Polymorphism (RFLP) analysis was performed after PCR with primers F: AGGAAAGTGGAGGGGATTTG and R: CCATGGCCACTGGTTTAGTT and digestion of the product with SphI. Bi-allelically targeted clones were Sanger sequenced to select clones heterozygous for the U2AF1[S34F] mutation. Heterozygous status was confirmed by cloning the PCR product into the PCR-4 TOPO TA vector (Invitrogen) and sequencing.

Four independent heterozygous *U2AF1* S34F clones – one generated with gRNA1 and three with gRNA2 – were confirmed to be karyotypically normal and after preliminary phenotypic characterization to exclude potential outliers, one clone (S34F-1) was selected for the subsequent editing step, along with P95L-1 clone and the parental N-2.12 line (Supplementary Table S1). A 3xFLAG tag was inserted into the C-terminus of the mutant or WT allele with a strategy similar to the one described above. A gRNA targeting the *U2AF1* locus (cutting site 12 bp 5' to the stop codon, with sequence AGATCTTTCACGATCTCTCG) or the *SRSF2* locus (cutting site 14 bp 3' to the stop codon, with sequence TAGGGGAATGGTAATGTCTG) was designed and cloned in the gRNA/Cas9 plasmid described above. Donor templates containing 5' (1098 bp) and 3' (1023 bp) homology arms (*U2AF1* locus), consisting of nucleotides 43093105 – 43094202 and 43092082 – 43093104 (hg38 human genome assembly),

13

respectively, and 5' (946 bp) and 3' (851 bp) homology arms (*SRSF2* locus) consisting of nucleotides 76736164 – 76737109 and 76735313 – 76736157, respectively, and the 3xFLAG sequence (GACTACAAGGACGACGATGACAAGGATTACAAAGATGACGACGATAAGGACTATAAGGAC GATGATGATAAA) were constructed. Gene targeting was performed as above. Colonies were screened by PCR and confirmed by TOPO TA cloning and sequencing. Monoallelically targeted clones with the desired allele tagged were selected and confirmed to be karyotypically normal. Expression of the FLAG-tagged protein (U2AF1 or SRSF2) of the expected size was confirmed by Western blot with an anti-FLAG antibody. To ensure clonality, an additional step of single-cell cloning was performed after each step of gene editing.

Human iPSC culture, hematopoietic differentiation and phenotypic characterization
Culture of human iPSCs on mitotically inactivated MEFs or feeder-free conditions, was performed as previously described(15, 16). All iPSC lines used in this study were tested periodically (every 2-4 weeks) and confirmed to be free of Mycoplasma contamination. Line authentication (SF genotype) was periodically performed by Sanger sequencing. All lines were confirmed to be karyotypically normal and were always maintained at a passage number not exceeding 100.

Hematopoietic differentiation was performed using a spin-EB protocol previously described(16). Briefly, cells were dissociated into single cells with accutase and plated at 3,500 cells per well in round-bottom low-attachment 96-well plates in APEL2 medium containing 5% protein-free hybridoma medium (PFHM-II), 30 ng/ml bone morphogenetic protein 4 (BMP4) and 10 µM Y-27632. The plates were centrifuged at 800 rpm for 5 min to induce EB aggregation. After 24 hours, the medium was replaced by APEL2 medium containing 5% PFHM-II, 30 ng/mL BMP4 and 50 ng/mL FGF2. After 2 days, the cytokine cocktail was changed to 5% PFHM-II, 20 ng/ml vascular endothelial growth factor (VEGF), 10 ng/ml FGF2, 100 ng/ml stem cell factor (SCF), 20 ng/ml Flt3 ligand (FL), 20 ng/ml thrombopoietin (TPO), 40 ng/ml IL-3. On day 8, EBs were collected and resuspended in Stem Pro34 SFM medium with 1% nonessential amino acids (NEAA), 1 mM L-glutamine and 0.1 mM β-mercaptoethanol (β-ME), supplemented with 100 ng/ml SCF, 20 ng/ml Flt3L, 20 ng/ml TPO, 40 ng/ml IL-3. The medium was thereafter replaced every two days. At the end of the differentiation culture, the cells were collected and dissociated with accutase into single cells and used for flow cytometry or clonogenic assays, as described(15, 16). Competitive growth assays were performed using an isogenic GFP-marked iPSC line (N-2.12-GFP), as previously described(15, 16).

Flow cytometry
The following antibodies were used: CD34-PE (clone 563, BD Pharmingen, RRID:AB_393871), CD45-APC (clone HI30, BD Pharmingen, RRID:AB_398600), CD14-APC (clone M5E2, BD Pharmingen, RRID:AB_398596), CD15-BV785 (clone W6D3, BioLegend, RRID:AB_2632921), CD16-BV510 (clone 3G8, BD Horizon, RRID:AB_2744296). Cell viability was assessed with DAPI (Life Technologies). Cells were assayed on a BD Fortessa and data were analyzed with FlowJo software (Tree Star, RRID:SCR_008520).

RNA sequencing
At least three clones of each genotype were subjected to hematopoietic differentiation. Magnetic-activated cell sorting (MACS) with anti-CD45 MACS cell separation microbeads and reagents (Miltenyi Biotec) was performed on an empirically determined day of differentiation culture when nearly 100% of CD45+ cells are also still CD34+ (ranging from day 10 to 13, depending on the individual line and differentiation experiment). RNA was extracted with the Direct-zol RNA purification kit (Zymo R2061). Sequencing libraries were prepared using the

14

TruSeq Stranded mRNA library prep kit (Illumina 20020594) from 500ng input RNA. Libraries used to call alternative splicing were sequenced to a depth of ~40 million reads in PE100 mode on an Illumina HiSeq4000. Libraries used to quantify gene expression were sequenced to a depth of ~15 million reads in SE75 mode on the Illumina HiSeq4000.

eCLIP library preparation
eCLIP was performed as previously described(22, 62). Briefly, sorted CD34+/CD45+ cells were UV-crosslinked (400 mJ/cm^2, 254 nm) and snap-frozen. Crosslinked cell pellets from independent differentiations were combined to obtain 10 million cells per replicate. Lysed pellets were sonicated and treated with RNaseI for RNA fragmentation. 2% of lysate was retained for preparation of a size-matched input library, and the remaining 98% was subject to immunoprecipitation (IP) using an anti-FLAG antibody (Sigma F1804, RRID:AB_262044), coupled to magnetic dynabeads (Invitrogen 11203D). Bound RNA fragments were dephosphorylated and 3'-end ligated with an RNA adapter. Protein-RNA complexes from both input and IP samples were run on SDS polyacrylamide gel and transferred to nitrocellulose membrane for extraction of bound RNA fragments. Membrane regions from the size of the tagged protein to 75 kDa above the protein size were cut and RNA was released with proteinase K. Input samples were then dephosphorylated and 3'-end ligated with an RNA adapter. Reverse transcription was performed with AffinityScript (Agilent) and cDNAs were 5'-end ligated with a DNA adapter. cDNA products were amplified with Q5 PCR mix (NEB) to obtain a sequencing library. Libraries were sequenced on the Illumina HiSeq4000 in SE75 mode to a depth of ~20 million reads per library.

Biotin-based visualization of RBP-coupled RNA
Biotin labeling and visualization was performed as described previously(63). Briefly, the cells were UV-crosslinked (400 mJ/cm^2) and snap-frozen in pellets of 1 million cells. Pellets were lysed and processed as described in eCLIP library preparation above through the immunoprecipitation (IP) step. Following IP, biotinylated cytidine (Thermo 20160) was ligated on-bead to the 3' end of RNA fragments. Samples were washed and loaded on SDS-PAGE gel. Ligation of biotinylated cytidine was performed on input samples that were taken before IP and loaded directly on the gel (no post-ligation cleanup was performed). Samples were transferred to a nitrocellulose membrane and developed with the Chemiluminescent Nucleic Acid Kit (Thermo 89880) following the manufacturer's instructions for visualization of RNAs.

RNA-seq data processing
RNA-Seq reads were trimmed of adapter sequences using cutadapt (v1.4.0, RRID:SCR_011841) and mapped to repetitive elements (RepBase v18.04, RRID:SCR_021169) with STAR (v2.4.0i, RRID:SCR_004463). Reads that did not map to repetitive elements were carried through and mapped to the human genome (hg19). GENCODE v19 gene annotations (RRID:SCR_01496) and featureCounts (v1.5.0, RRID:SCR_012919) were used to assign reads to genes and genic regions.

Quantification of alternative splicing
rMATS v4.0.2 (RRID:SCR_013049) was used to perform alternative splicing (AS) analysis among replicate RNA-seq datasets. Significant events were calculated as those with >5% change in isoform ratio between genotypes with a FDR < 0.05. Each event was required to contain an average of at least 10 reads supporting the inclusion and exclusion isoform across replicates in one genotype and at least an average of 10 total reads of both isoforms in the other genotype. We previously reported that rMATS often calls multiple splicing events with different flanking regions but overlapping alternatively spliced regions(23). To remove these artifacts and

15

avoid double-counting of AS events, we used custom scripts to remove overlapping AS events for analysis (subset_rmats_junctioncountonly.py found in https://github.com/YeoLab/rbp-maps). Unprocessed fastq files of RNA-Seq data of MDS patients from Pellagatti et al.(12) and Madan et al.(64) were downloaded from GEO (accession GSE114922 and GSE63816 respectively) and processed as described above.

Selection of unchanged cassette exons

To generate a background list of cassette exons that are unchanged between genotypes, we randomly selected exons that were matched for inclusion levels. The splicing events of interest were grouped in ranges of inclusion levels from: 0-0.25, 0.25-0.5, 0.5-0.75, and 0.75-1. We then calculated the total number of events that fell within each range and randomly selected that same number of exons from a list that were not differentially regulated between genotypes and were within the same range of inclusion level.

RNA-Seq motif analyses

Weblogo (https://weblogo.berkeley.edu/logo.cgi) was used to generate nucleotide enrichment figures at 3' splice sites flanking regulatory exons. 5-mer sequences within cassette exons were quantified with Kvector (https://github.com/olgabot/kvector). The resulting counts were summed by genotype and enrichment was calculated using a chi-squared test (*Scipy v1.2.0, chi2_contingency*, RRID:SCR_008058). Motif enrichment within cassette exons and downstream control exons was determined with the HOMER function (findmotifs.pl, RRID:SCR_010881) in rna mode with a sequence length of 6. Sequences within 'WT exons' were used as the foreground, and sequences within 'P95L exons' or 'S34F exons' were used as the background to calculate enrichment in "WT exons". The reverse orientation was used to calculate sequence enrichment in "P95L exons" or "S34F exons". P-values were calculated with a cumulative binomial distribution.

eCLIP-seq data processing

Code used for eCLIP data processing is available on GitHub (https://github.com/YeoLab/eclip). Briefly, reads were adapter-trimmed and mapped to a database of repetitive elements (RepBase v18.04) with STAR (v2.4.0i). Reads that did not map to repetitive elements were mapped to the human genome (hg19) with STAR. Removal of PCR-duplicated reads was performed using the unique molecular identifier sequences in the 5' adaptor, and non-duplicated reads were retained as 'usable reads'. Peaks were called on 'usable reads' with CLIPper and assigned to gene regions annotated in GENCODE v19 (RRID:SCR_01496) in the following order of priority: 3'splice site, 5' splice site, coding sequence, 3' UTR, 5' UTR, proximal intron, distal intron, noncoding regions. Peaks with multiple annotations were assigned to the region with the highest priority. Peaks were deemed significant at ≥ 4-fold enrichment relative to input and $p<0.001$ (chi-squared or Fisher's Exact test). Peaks that passed significance thresholds in one of two replicate experiments were retained for analysis. Splicing maps were generated with RBP-maps(23) (https://github.com/YeoLab/rbp-maps). Raw sequencing data and processed files are available at GEO with accession code GSE164666.

Kmer enrichment analysis and calculation of enriched sequences of eCLIP peaks

Kmer sequences were counted in regions of interest using kvector (https://github.com/olgabot/kvector). The resulting counts were summed by genotype and enrichment was calculated using a chi-squared test (Scipy v1.2.0, chi2_contingency, RRID:SCR_008058) comparing the kmer frequencies observed in each genotype. FDR was calculated to correct for multiple hypothesis testing and 6-mer sequences with FDR < 0.05 were reported as significantly enriched.

16

HOMER analysis of motif enrichment in eCLIP peaks

The findmotifs.pl function of HOMER (RRID:SCR_010881) was used in RNA mode with a sequence length of 6 to identify motifs enriched in binding sites of the mutant and WT factors. To identify sequences enriched in WT peaks, the WT binding sites were input as the foreground and the mutant binding sites were included as the background for each genotype. The reverse comparison was performed to identify sequences enriched in the mutant factor binding sites.

RT-PCR for GNAS-L and GNAS-S

RNA was isolated with Trizol (Life Technologies). Reverse transcription was performed with Superscript III (Life Technologies) and PCR was performed with primers F: AAGCACCATTGTGAAGCAGA, R: TTCAATCGCCTCTTTCAGGT with varying number of amplification cycles (from 25 to 35). Products were separated on an agarose gel. The lowest cycle count in which both products were visible was used for quantification of band intensity using ImageJ (RRID:SCR_003070).

GNAS-L knockdown and overexpression

Two shRNAs targeting exon 3, specific to GNAS-L isoform, were designed with sequence shRNA 1: ACCCACCATAGGGCATGATTA; shRNA 2: TAAAGCCTTAAGCACAATTAA. The GNAS-L shRNA and a scrambled shRNA sequences were inserted into the 3′ UTR of the G-U6 lentiviral vector(19). Lentiviral vectors were packaged as described(19). 500,000 iPSC-HSPCs on day 11 of hematopoietic differentiation were transduced with a lentiviral vector expressing either the GNAS-L shRNA or scrambled shRNA. Two days after transduction, 5000 cells were plated on methylcellulose media for colony forming assay and the rest were used for quantification of GNAS-L and GNAS-S expression by RT-PCR.

For overexpression, GNAS cDNA sequences encoding the short or long isoform with or without the R201H mutation – short wild-type (S-WT), long wild-type (L-WT), short R201H (S-R201H), long R201H (L-R201H) – or GFP (empty vector) were inserted into the mP2A lentiviral vector co-expressed with mCherry through a P2A peptide(16). 500,000 iPSC-HSPCs on d11 of hematopoietic differentiation, K562 or CB CD34+ cells were transduced in parallel with each lentiviral vector. Two days after transduction (d13 of hematopoietic differentiation) the cells were harvested for Western blots and/or cAMP ELISA assay. For cytokine starvation, cells on d12 of hematopoietic differentiation were washed with 1X PBS and cultured for 24 h prior to collection in Stem Pro34 medium with 1% nonessential amino acids (NEAA), 1 mM L-glutamine and 0.1 mM β-ME without cytokines.

Protein expression and purification

The human Gαs-S and Gαs-L cDNAs, each containing TEV cleavable N-terminal hexa-histidine tag, were cloned into pQE60 vector (Qiagen, RRID:Addgene_12553). Chemically competent WK6 cells were transformed with the pQE60- Gαs plasmid and expression was induced by 30 mM Isopropyl-β-D-thiogalactopyranoside for approximately 20 hours. Cell pellets were lysed and homogenized in 50 mM HEPES pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10% glycerol, protease inhibitor cocktail (35 µg/ml phenylmethanesulfonyl fluoride, 32 µg/ml tosyl phenylalanyl chloromethyl ketone, 32 µg/ml tosyl lysyl chloromethyl ketone, 3.2 µg/ml leupeptin and 3.2 µg/ml soybean trypsin inhibitor), 6 mM β-ME, 10 mM GDP and DNAse (Roche Diagnostics). The lysed cell suspension was sonicated three times (with each cycle 8 min long and consisting of 10 sec pulse and 10 sec pause) and centrifuged for 40 min at 35.000 rpm at 4°C. Imidazole was added to the clarified supernatant at 20 mM final concentration, which was then applied to a 1.5 mL Ni-NTA resin (Thermo Scientific) gravity-flow column (Bio-rad) pre-equilibrated with 20 mM HEPES pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, 6 mM bME and 10 mM GDP. The column was then washed with 15 column volumes (CV) of high salt wash buffer followed by a low salt wash of 10 CV (20 mM HEPES pH 8.0, 20 mM NaCl, 1 mM $MgCl_2$, 6 mM bME and 10 mM

17

GDP). The protein was eluted with 7 mL of elution buffer (low salt wash buffer with 200 mM imidazole added). The pooled elution fractions were diluted 4x with ion-exchange buffer A (20 mM HEPES pH 8.0, 5% glycerol and 6 mM bME) and immediately loaded on a 4% buffer B (ion-exchange buffer A with 1 M NaCl added)-equilibrated 8 mL Q-Sepharose column (GE Healthcare). The column was washed with 5 CV of 4% buffer B and eluted with a linear gradient of 4% to 50% buffer B over 200 mL. 4 mL fractions were collected directly into tubes already containing GDP to reach a final concentration of 10 mM. The Gαs-containing fractions were concentrated using a centrifugal spin concentrator with a molecular cut-off of 30 kDa (Amicon) and flash frozen until further use. The protein concentration was determined by Bradford (Bio-rad) and the functional active protein by [^{35}S]GTPγS binding assay.

[^{35}S]GTPγS binding assay

GTPγS binding was initiated by the addition of 100 mL of a buffer containing 2.5 mM cold GTPγS and 3.5 nM [^{35}S]GTPγS to equal volumes (100 mL) of 400 nM purified G protein in 20 mM HEPES pH 7.7, 100 mM NaCl, 10 mM MgCl$_2$, 14.3 mM β-ME, for a final assay concentration of 200 nM G protein. At specific time points (0, 3, 6, 10, 16, 22, 30 and 40 min) a 20 mL aliquot was withdrawn from the reaction and quenched in 100 μL ice-cold quenching buffer (20 mM HEPES pH 7.7, 100 mM NaCl, 10 mM MgCl$_2$, 2 mM β-ME and 100 mM GTP). Quenched samples were immediately vacuum-filtered through PROTRAN BA-83 0.22 mm nitrocellulose filters (Whatman) presoaked in assay buffer containing 20 mM HEPES pH 7.7, 100 mM NaCl, 10 mM MgCl$_2$, 14.3 mM β-ME and 1 mM GDP. Filters were washed twice with 4 mL of a buffer containing 20 mM HEPES pH 7.7, 100 mM NaCl and 2 mM MgCl$_2$ and dried at room temperature overnight. Dried filters were transferred to scintillation vials containing 5 mL of liquid scintillation cocktail (CytoScint ES, MP Biomedicals). After 1 hour of equilibration the radioactivity was determined using a LS 6000IC Beckman counter. Association rate constants were determined by fitting the data to a single exponential using Prism 6 (GraphPad LLC, CA, RRID:SCR_002798).

Bodipy FL-GTPγS binding assay

We measured real-time GTPγS binding to purified Gas protein preparations (prepared as described above) using a fluorescent analogue, Bodipy FL- GTPγS (Invitrogen)(65), as described(66). Briefly, purified Gas protein (400 nM final) was incubated with Bodipy FL- GTPγS (100 nM final) in a buffer containing 20 mM HEPES pH 8.0, 200 mM NaCl, 10 mM MgCl$_2$ and 7 mM β-ME in a total volume of 100 uL. Fluorescence was measured in a quartz cuvette using Horiba Fluomax 4 with excitation @ 480 nm (10 nm slit width) and emission @ 510 nm (5 nm slit width) at 4 second recording intervals. Association rate constants were determined by fitting the data to a single exponential using Prism 6 (GraphPad LLC, CA, RRID:SCR_002798).

cAMP assay

The assay was performed in a clonal selected ΔGs-HEK293 cell line(29), stably expressing the cAMP biosensor pink flamido (encoded on a pcDNA4.0/TO/Zeocin plasmid). The cells were cultured in DMEM with 10% FBS (Peak Serum) and 100 mg/ml zeocin. Doxycycline (4 mg/mL) was added to the media to induce expression of the cAMP biosensor and two days later the cells were transiently transfected with plasmid DNA encoding each G protein isoform.

Cells were harvested after 6 hours, washed in Hanks' Balanced Salt Solution (HBSS, Sigma) and replated onto clear bottom poly-D-lysine coated, black 96-well polystyrene assay plates at 2 x 10^6 cells per well. An additional 2 x 10^6 cells were used for Western Blotting. Isoproterenol in HBSS buffer with 20 mM HEPES, 600 mM 3-Isobutyl-1-methylxanthine (IBMX) and 3 mM ascorbic acid was added to the cells (as a 3x stock) at concentrations 10^{-5}, 10^{-6}, 10^{-7}, 10$^{-7.5}$, 10$^-$

18

8.0, $10^{-8.5}$ and $10^{-9.5}$ M and fluorescence was read using a fluorescence plate reader (BioTek, Winooski, VT). The cAMP accumulation was monitored for 17.5 min. For each isoproterenol concentration the initial rates for the linear portion of the cAMP accumulation curve (30s to 330s) were fitted by linear regression. Rates were expressed as a function of isoproterenol concentration and fitted to a logistic curve using Prism (GraphPad LLC, CA, RRID:SCR_002798).

Western blot

500,000 to 1 million iPSC-HSPCs, K562 and CB CD34+ cells or 300,000 to 500,000 primary MDS or sAML patient bone marrow or peripheral blood mononuclear cells were collected and lysed with high salt buffer (0.3 M KCl) supplemented with protease inhibitor and phosphatase inhibitor. Protein concentration was determined by bicinchoninic acid assay (Pierce Biotechnology Inc.) and 20 µg of protein from each extract were diluted in Laemmli SDS sample buffer and resolved by electrophoresis on Bolt 4% to 12% Bis-Tris precast gels (Invitrogen) and blotted on nitrocellulose membranes. The membranes were blocked with 5% BSA (Fischer bioreagents) in Tris-buffered saline and incubated with primary antibody p-(Ser/Thr) PKA substrate (9621S, Cell Signaling Technologies, RRID:AB_330304), anti-FLAG (14793S, Cell Signaling Technologies), anti-P-p44/42 MAPK (Erk1/2) (4370S, Cell Signaling Technologies), p44/42 MAPK(Erk1/2) (4696S, Cell Signaling Technologies), anti-P-Raf-1 (Ser338) (05-538, Millipore), DUSP6/MKP3 (39441, Cell Signaling Technologies, RRID:AB_2246226), P-MEK1/2 (Ser217/221) (9154S, Cell Signaling Technologies), P-AKT (S473) (4060S, Cell Signaling Technologies), anti-Gαs-Subunit C-terminal (371732, Millipore), anti-mCherry (ab213511, Abcam, RRID:AB_281489) or anti-β-Actin (5125S, Cell Signaling Technologies). After washing, blots were incubated with HRP-conjugated secondary antibody and developed using ECL Western Blotting Detection Reagents (Pierce ECL Western Blotting Substrate, Thermo Scientific). Band intensity was quantified using ImageJ (RRID:SCR_003070). All samples shown in each individual figure panel were processed in parallel on the same blot. Empty space was added to indicate samples that were not run in lanes adjacent to one another. For each antibody independent blots were used from the same cell lysate with identical loading conditions.

Primary MDS and AML patient cells

AML and MDS patient bone marrow or peripheral blood mononuclear cells were obtained with written informed consent under protocols approved by a local Institutional Review Board at the Icahn School of Medicine at Mount Sinai in accordance with recognized ethical guidelines. Information on cytogenetic abnormalities and mutations for each sample was obtained from the Mount Sinai tissue bank. A subset of samples was sequenced for a more extended gene panel. Specifically, a custom capture bait set was used to sequence the coding regions of 163 MDS and AML associated genes. Samples were sequenced with pair-end Illumina Hi-Seq at a median coverage of 600x per sample (range 127-2480x). Variants with VAF < 2%, less than 20 total reads or less than 5 mutant supporting reads were excluded. After pre-filtering of artefactual variants, likely germline SNPs were filtered out by considering the VAF density of variants, their presence in the Genome Aggregation Database (gnomAD), their annotation in the human variation database ClinVar (RRID:SCR_006169) and their recurrence in a panel of normal samples. From the list of likely somatic variants, putative oncogenic variants were distinguished from variants of unknown significance (VUS) based on the mutational consequence and their recurrence in various databases of somatic mutations in cancer. All MDS- and AML- associated mutations and chromosomal abnormalities found in the samples are listed in Supplementary Table S6.

Cryopreserved cells were thawed and cultured in X-VIVO15 containing 20% BIT, 1% nonessential amino acids (NEAA), 1 mM L-glutamine and 0.1 mM β-ME supplemented with 100 ng/ml SCF, 50 ng/ml Flt3L, 50 ng/ml TPO and 20 ng/ml IL-3, for 1-3 days and collected for Western blots and cAMP ELISA assay.

cAMP ELISA assay
2.5 million primary AML cells were collected after incubation with 100uM 3-isobutyl-1-methylxanthine (IBMX) at 37°C for 30 min. All samples were run in duplicate according to the manufacturer's protocol (cAMP Parameter Assay, R&D Systems).

Treatment with MEK inhibitors
iPSC-HSPCs were plated on day 11 of hematopoietic differentiation on 96-well tissue culture-treated clear flat bottom plates (Corning, 3903 or 3603 respectively) at a density of 20,000 per well. The compounds trametinib-S2673 (GSK1120212), cobimetinib-S8041 (GDC-0973), selumetinib-S1008 (AZD6244) and vemurafenib-S1267 (PLX4032) were purchased from Selleckchem. All compounds were dissolved in DMSO for stock solutions at a concentration of 5mM or 10mM and subsequently diluted in StemPro media and added to a total volume of 50uL of media per well at a final concentration of 10μM, 1μM, 100nM, 50nM, 10nM, 5nM, 1nM, 0.1nM or 0.01nM in triplicate wells. After 3 days, cell viability was measured using the CellTiter-Glo Luminescent Cell Viability Assay (Promega, G7570) per the manufacturer's suggested conditions. Percent viability at each compound concentration was calculated as: (Signal-Blank)/(DMSO Control-Blank) x 100. IC50 value calculations and generation of IC50 curves were performed using the Prism 8 software (Graphpad, RRID:SCR_002798). For MEKi treatment of primary MDS and sAML patient bone marrow or peripheral blood mononuclear cells, 10,000 to 20,000 cells per well were plated and treated with the compounds and cell viability was measured after 1 to 2 days.

Statistical analysis
Statistical analysis was performed with GraphPad Prism software (Graphpad, RRID:SCR_002798). Pairwise comparisons between different groups were performed using a two-sided unpaired unequal variance t-test. For all analyses, $p < 0.05$ was considered statistically significant, unless otherwise stated.

Data availability

RNA-sequencing and eCLIP-sequencing data of this study have been deposited in GEO with the accession code GSE164666.

20

References

1.   Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet. 2011 Dec 11;44(1):53-7.

2.   Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med. 2011 Oct 13;365(15):1384-95.

3.   Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011 Oct 6;478(7367):64-9.

4.   Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013 Nov 21;122(22):3616-27; quiz 99.

5.   Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SC, Ramakrishnan A, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell. 2015 May 11;27(5):617-30.

6.   Zhang J, Lieu YK, Ali AM, Penson A, Reggio KS, Rabadan R, et al. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc Natl Acad Sci U S A. 2015 Aug 25;112(34):E4726-34.

7.   Okeyo-Owuor T, White BS, Chatrikhi R, Mohan DR, Kim S, Griffith M, et al. U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. Leukemia. 2015 Apr;29(4):909-17.

8.   Ilagan JO, Ramakrishnan A, Hayes B, Murphy ME, Zebari AS, Bradley P, et al. U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res. 2015 Jan;25(1):14-26.

9.   Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer. 2016 Jul;16(7):413-30.

10.   Lee SC, North K, Kim E, Jang E, Obeng E, Lu SX, et al. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. Cancer Cell. 2018 Aug 13;34(2):225-41 e8.

11.   Shiozawa Y, Malcovati L, Galli A, Sato-Otsubo A, Kataoka K, Sato Y, et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. Nature communications. 2018 Sep 7;9(1):3649.

12.   Pellagatti A, Armstrong RN, Steeples V, Sharma E, Repapi E, Singh S, et al. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. Blood. 2018 Sep 20;132(12):1225-40.

13.   Shirai CL, Ley JN, White BS, Kim S, Tibbitts J, Shao J, et al. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell. 2015 May 11;27(5):631-43.

14.   Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci U S A. 2005 Feb 22;102(8):2850-5.

15.   Kotini AG, Chang CJ, Chow A, Yuan H, Ho TC, Wang T, et al. Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. Cell Stem Cell. 2017 Mar 2;20(3):315-28 e7.

16.   Chang CJ, Kotini AG, Olszewska M, Georgomanoli M, Teruya-Feldstein J, Sperber H, et al. Dissecting the Contributions of Cooperating Gene Mutations to Cancer Phenotypes and Drug Responses with Patient-Derived iPSCs. Stem Cell Reports. 2018 May 8;10(5):1610-24.

17.   Lee SC, Dvinge H, Kim E, Cho H, Micol JB, Chung YR, et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat Med. 2016 Jun;22(6):672-8.

18.   Saez B, Walter MJ, Graubert TA. Splicing factor gene mutations in hematologic malignancies. Blood. 2017 Mar 9;129(10):1260-9.

19.   Kotini AG, Chang CJ, Boussaad I, Delrow JJ, Dolezal EK, Nagulapally AB, et al. Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. Nat Biotechnol. 2015 Jun;33(6):646-55.

20.   Hsu J, Reilly A, Hayes BJ, Clough CA, Konnick EQ, Torok-Storb B, et al. Reprogramming identifies functionally distinct stages of clonal evolution in myelodysplastic syndromes. Blood. 2019 Jul 11;134(2):186-98.

21.   Flores-Figueroa E, Gutierrez-Espindola G, Guerrero-Rivera S, Pizzuto-Chavez J, Mayani H. Hematopoietic progenitor cells from patients with myelodysplastic syndromes: in vitro colony growth and long-term proliferation. Leuk Res. 1999 Apr;23(4):385-94.

22.   Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods. 2016 Jun;13(6):508-14.

23.   Yee BA, Pratt GA, Graveley BR, Van Nostrand EL, Yeo GW. RBP-Maps enables robust generation of splicing regulatory maps. RNA. 2019 Feb;25(2):193-204.

24.   Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011 Jun 30;364(26):2496-506.

25.   O'Hayre M, Vazquez-Prado J, Kufareva I, Stawiski EW, Handel TM, Seshagiri S, et al. The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. Nat Rev Cancer. 2013 Jun;13(6):412-24.

26.   Dror RO, Mildorf TJ, Hilger D, Manglik A, Borhani DW, Arlow DH, et al. SIGNAL TRANSDUCTION. Structural basis for nucleotide exchange in heterotrimeric G proteins. Science. 2015 Jun 19;348(6241):1361-5.

27.   Sunahara RK, Tesmer JJ, Gilman AG, Sprang SR. Crystal structure of the adenylyl cyclase activator Gsalpha. Science. 1997 Dec 12;278(5345):1943-7.

28.   Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature. 2011 Jul 19;477(7366):549-55.

29.   Stallaert W, van der Westhuizen ET, Schonegge AM, Plouffe B, Hogue M, Lukashova V, et al. Purinergic Receptor Transactivation by the beta2-Adrenergic Receptor Increases Intracellular Ca(2+) in Nonexcitable Cells. Mol Pharmacol. 2017 May;91(5):533-44.

30.   Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016 Jun 9;374(23):2209-21.

31.   Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014 Feb;28(2):241-7.

32.   Gutkind JS. Regulation of mitogen-activated protein kinase signaling networks by G protein-coupled receptors. Sci STKE. 2000 Jul 11;2000(40):re1.

33.   Jain R, Watson U, Vasudevan L, Saini DK. ERK Activation Pathways Downstream of GPCRs. Int Rev Cell Mol Biol. 2018;338:79-109.

34.   Papapetrou EP. Modeling myeloid malignancies with patient-derived iPSCs. Exp Hematol. 2019 Mar;71:77-84.

35.   Wesely J, Kotini AG, Izzo F, Luo H, Yuan H, Sun J, et al. Acute Myeloid Leukemia iPSCs Reveal a Role for RUNX1 in the Maintenance of Human Leukemia Stem Cells. Cell Rep. 2020 Jun 2;31(9):107688.

36.   Ruiz-Gutierrez M, Bolukbasi OV, Alexe G, Kotini AG, Ballotti K, Joyce CE, et al. Therapeutic discovery for marrow failure with MDS predisposition using pluripotent stem cells. JCI Insight. 2019 Apr 30;5.

37.   Fong JY, Pignata L, Goy PA, Kawabata KC, Lee SC, Koh CM, et al. Therapeutic Targeting of RNA Splicing Catalysis through Inhibition of Protein Arginine Methylation. Cancer Cell. 2019 Aug 12;36(2):194-209 e9.

38.   Chao MP, Gentles AJ, Chatterjee S, Lan F, Reinisch A, Corces MR, et al. Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease. Cell Stem Cell. 2017 Mar 2;20(3):329-44 e7.

39.   Singh S, Ahmed D, Dolatshad H, Tatwavedi D, Schulze U, Sanchi A, et al. SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications. Leukemia. 2020 Sep;34(9):2525-30.

40.   Fidler TP, Xue C, Yalcinkaya M, Hardaway B, Abramowicz S, Xiao T, et al. The AIM2 inflammasome exacerbates atherosclerosis in clonal haematopoiesis. Nature. 2021 Apr;592(7853):296-301.

41.   Liang Y, Tebaldi T, Rejeski K, Joshi P, Stefani G, Taylor A, et al. SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. Leukemia. 2018 Dec;32(12):2659-71.

42.   Palangat M, Anastasakis DG, Fei DL, Lindblad KE, Bradley R, Hourigan CS, et al. The splicing factor U2AF1 contributes to cancer progression through a noncanonical role in translation regulation. Genes Dev. 2019 May 1;33(9-10):482-97.

43.   Esfahani MS, Lee LJ, Jeon YJ, Flynn RA, Stehr H, Hui AB, et al. Functional significance of U2AF1 S34F mutations in lung adenocarcinomas. Nat Commun. 2019 Dec 13;10(1):5712.

44.   Fei DL, Zhen T, Durham B, Ferrarone J, Zhang T, Garrett L, 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 Oct 30;115(44):E10437-E46.

45.   Przychodzen B, Jerez A, Guinta K, Sekeres MA, Padgett R, Maciejewski JP, et al. Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. Blood. 2013 Aug 8;122(6):999-1006.

46.   Kon A, Yamazaki S, Nannya Y, Kataoka K, Ota Y, Nakagawa MM, et al. Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. Blood. 2018 Feb 8;131(6):621-35.

47.   Graziano MP, Freissmuth M, Gilman AG. Expression of Gs alpha in Escherichia coli. Purification and properties of two forms of the protein. J Biol Chem. 1989 Jan 5;264(1):409-18.

48.   Kvapil P, Novotny J, Svoboda P, Ransnas LA. The short and long forms of the alpha subunit of the stimulatory guanine-nucleotide-binding protein are unequally redistributed during (-)-isoproterenol-mediated desensitization of intact S49 lymphoma cells. Eur J Biochem. 1994 Nov 15;226(1):193-9.

49.   Seifert R, Wenzel-Seifert K, Lee TW, Gether U, Sanders-Bush E, Kobilka BK. Different effects of Gsalpha splice variants on beta2-adrenoreceptor-mediated signaling. The beta2-adrenoreceptor coupled to the long splice variant of Gsalpha has properties of a constitutively active receptor. J Biol Chem. 1998 Feb 27;273(9):5109-16.

50.   Unson CG, Wu CR, Sakmar TP, Merrifield RB. Selective stabilization of the high affinity binding conformation of glucagon receptor by the long splice variant of Galpha(s). J Biol Chem. 2000 Jul 14;275(28):21631-8.

51.   Goldsmith ZG, Dhanasekaran DN. G protein regulation of MAPK networks. Oncogene. 2007 May 14;26(22):3122-42.

52.   Khan ZM, Real AM, Marsiglia WM, Chow A, Duffy ME, Yerabolu JR, et al. Structural basis for the action of the drug trametinib at KSR-bound MEK. Nature. 2020 Dec;588(7838):509-14.

53.   Cazzola M. Myelodysplastic Syndromes. N Engl J Med. 2020 Oct 1;383(14):1358-74.

54.   Assi SA, Imperato MR, Coleman DJL, Pickin A, Potluri S, Ptasinska A, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. Nat Genet. 2019 Jan;51(1):151-62.

55.   Gilliland DG, Griffin JD. Role of FLT3 in leukemia. Curr Opin Hematol. 2002 Jul;9(4):274-81.

56.   Navas TA, Mohindru M, Estes M, Ma JY, Sokol L, Pahanish P, et al. Inhibition of overactivated p38 MAPK can restore hematopoiesis in myelodysplastic syndrome progenitors. Blood. 2006 Dec 15;108(13):4170-7.

57.   Navada SC, Garcia-Manero G, OdchimarReissig R, Pemmaraju N, Alvarado Y, Ohanian MN, et al. Rigosertib in combination with azacitidine in patients with myelodysplastic syndromes or acute myeloid leukemia: Results of a phase 1 study. Leuk Res. 2020 Jul;94:106369.

58.   Borthakur G, Popplewell L, Boyiadzis M, Foran J, Platzbecker U, Vey N, et al. Activity of the oral mitogen-activated protein kinase kinase inhibitor trametinib in RAS-mutant relapsed or refractory myeloid malignancies. Cancer. 2016 Jun 15;122(12):1871-9.

59.   Samatar AA, Poulikakos PI. Targeting RAS-ERK signalling in cancer: promises and challenges. Nat Rev Drug Discov. 2014 Dec;13(12):928-42.

60.   Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Science translational medicine. 2011 Jul 20;3(92):92ra66.

61.   Landis CA, Masters SB, Spada A, Pace AM, Bourne HR, Vallar L. GTPase inhibiting mutations activate the alpha chain of Gs and stimulate adenylyl cyclase in human pituitary tumours. Nature. 1989 Aug 31;340(6236):692-6.

62.   Van Nostrand EL, Nguyen TB, Gelboin-Burkhart C, Wang R, Blue SM, Pratt GA, et al. Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). Methods Mol Biol. 2017;1648:177-200.

63.   Van Nostrand EL, Pratt GA, Yee BA, Wheeler EC, Blue SM, Mueller J, et al. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. Genome Biol. 2020 Apr 6;21(1):90.

64.   Madan V, Kanojia D, Li J, Okamoto R, Sato-Otsubo A, Kohlmann A, et al. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nature communications. 2015 Jan 14;6:6042.

65.   McEwen DP, Gee KR, Kang HC, Neubig RR. Fluorescent BODIPY-GTP analogs: real-time measurement of nucleotide binding to G proteins. Anal Biochem. 2001 Apr 1;291(1):109-17.

66.   Westfield GH, Rasmussen SG, Su M, Dutta S, DeVree BT, Chung KY, et al. Structural flexibility of the G alpha s alpha-helical domain in the beta2-adrenoceptor Gs complex. Proc Natl Acad Sci U S A. 2011 Sep 20;108(38):16086-91.

25

Figure Legends

Figure 1: SF-mutant iPSC-HSPC models recapitulate cellular phenotypes and motif preferences of SF-mutant MDS. A. Schematic overview of the generation of isogenic clonal iPSC lines with the canonical *U2AF1 S34F* and *SRSF2 P95L* mutations and allele-specific epitope tags through CRISPR/Cas9-mediated gene editing. **B.** Percentage of mutant allele of total *U2AF1* or *SRSF2* transcripts (from RNA-Seq data) confirming heterozygous state and showing equal expression of the mutant and WT alleles. Mean and SEM of 3-4 replicates is shown (Supplementary Table S1). **C.** Number of colonies from 5,000 cells seeded in methylcellulose assays on day 14 of hematopoietic differentiation of the indicated iPSC lines. Mean of 1-6 independent differentiation experiments for each line is shown. **D.** Fraction of CD15+, CD14+ and CD16+ myeloid cells on days 12 and 14 of hematopoietic differentiation of WT and SF-mutant iPSC lines (WT-1, S34F-6, P95L-2). Mean and SEM of 2-4 independent differentiation experiments are shown. **E.** Competitive growth assay. The cells were mixed 1:1 at the onset of hematopoietic differentiation with an isogenic WT iPSC line stably expressing GFP (derived from the parental line WT-1). The relative population size was estimated as the percentage of GFP- cells (measured by flow cytometry) at each time point (days 4-12 of differentiation), relative to the population size on day 2. Results from 1 or 2 independent experiments per line are shown. **F.** iPSC lines were differentiated along the hematopoietic lineage and CD34+/CD45+ HSPCs were sorted for RNA-Seq and allele-specific eCLIP analyses. **G.** Number of alternative splicing (AS) events detected for each genotype in comparison to WT cells (FDR < 0.05, delta PSI > 5%). SE: skipped exon; A3SS: alternative 3' splice site; A5SS: alternative 5' splice site; RI: retained intron. **H.** Sequence logos of 3' splice sites flanking skipped exon events and downstream exons. "WT exons" and "S34F exons" denote exons skipped or included, respectively, in U2AF1-mutant cells. **I.** Count of skipped exon events containing the indicated 3' splice site sequence. "WT exons" and "S34F exons" denote exons skipped or included, respectively, in U2AF1-mutant cells. "Unchanged" denotes a randomly selected control set of exons that are not differentially spliced in U2AF1-mutant cells. **J.** HOMER motif enrichment of sequence preferences within skipped exons and downstream control exons. "WT exons" and "P95L exons" denote exons skipped or included, respectively, in SRSF2-mutant cells. P-values were calculated with a cumulative binomial distribution. **K.** Individual 5-mer sequences with most significant differential enrichment in "WT exons" and "P95L exons". p < 0.005 by chi-squared test with Bonferroni-correction for all 5-mers plotted.

Figure 2: Allele-specific eCLIP identifies direct targets of mis-splicing in human HSPCs from isogenic iPSC models of SRSF2- and U2AF1- mutant MDS. A. Count of significantly enriched binding sites (eCLIP peaks) identified in each genic region indicated. Significantly enriched peaks are peaks with fold change > 4 relative to input and p-value < 0.001 (chi-squared test) in at least one of two replicate experiments. **B.** Top enriched motifs from all peaks identified per genotype. P-values are reported to the right of each motif (cumulative binomial distribution). **C,D.** Relative abundances of 6-mer sequences counted in all eCLIP peak regions. Differentially used 6-mers were identified with a chi-squared test and FDR < 0.05 and are colored according to the 6-mer category defined in the legend. 6-mers that are not differentially used are plotted in light grey. **E,F.** Splicing maps of eCLIP binding density in regions flanking skipped exons that are differentially used between WT and U2AF1-mutant (E) or SRSF2 mutant (F) cells. A random background set of exons that are unchanged between WT and mutant cells was used as control. Binding density is averaged across regions (50 bp on each end of exon sequences and 300 bp into surrounding introns) and normalized to the input sample. Data from one of two replicate samples are shown. (The other replicate is shown in Supplementary Fig. S8A and S8B.) AU: arbitrary units

26

Figure 3: Intersectional analysis of mis-splicing and mRNA binding reveals a long isoform of GNAS (GNAS-L) as a direct convergent effect of both mutant SFs. A. Upper panel: schematic of strategy for identification of skipped exon events with differential peaks in regulatory regions (upstream intron for U2AF1 and cassette exon for SRSF2). Skipped exons are categorized in 3 groups: "WT peak" include peaks bound by the WT SF only; "Mutant peak" include peaks bound by the mutant SF only; and "Both peak" include sites bound by both WT and mutant SF. Lower panels: Delta PSI (WT–mutant) of skipped exon events divided among the 3 binding categories. A chi-square test was used to determine dependence of "WT peak" or "Mutant peak" status on the direction of exon inclusion (positive Delta PSI in "WT peak" and negative Delta PSI in "Mutant peak"). **B.** Workflow showing the integration of splicing analyses in both iPSC-HSPC genotypes with datasets from *U2AF1-* and *SRSF2-* mutant MDS patient cells and with eCLIP analyses of differentially bound exons. 41 splicing events, affecting 40 genes, were common to both *U2AF1-* and *SRSF2-* mutant iPSC-HSPCs. Of those, 20 had sufficient read coverage to allow evaluation of splicing in publicly available data from MDS patients (Pellagatti et al. 2018(12)). 15 of the 20 genes were differentially spliced in the same direction in both *U2AF1-* and *SRSF2-* mutant cells, compared to cells from MDS patients without SF mutations (SF-WT). Of those 15, the 3 genes indicated (*GNAS*, *ITGB3BP* and *PSMA4*) also contained a differential eCLIP peak between mutant and WT factor. **C.** Delta PSI of the 20 AS events common to both iPSC-HSPC genotypes that could be evaluated in MDS patient cells (sufficient coverage). Black dots represent AS events trending in the same direction in both genotypes vs SF-WT in the MDS patient cells, as well as in our iPSC-HSPC models of the two mutations. Blue circles show events that also contain a differential eCLIP peak in the regulatory region (as determined in a). **D.** Visualization of G/A and G/C-rich sequence motifs in exons 2-4 of the *GNAS* transcript. Only exon 3, which is more included in *SRSF2* P95L cells, contains a G/C-rich sequence motif. **E.** eCLIP binding density at exons 2 - 5 of the GNAS transcript showing normalized read density for one replicate of WT U2AF1 (green) and U2AF1-S34F (blue) and sashimi plots showing inclusion levels of exon 3 in WT (green), S34F (blue) and P95L (orange) iPSC-HSPCs. Read counts are from one representative sample of each genotype and Ψ (Percent spliced in) is the average of all biological replicates. An eCLIP peak at exon 3 can be seen in the U2AF1-S34F but not the WT U2AF1 track. (There were no peaks for either WT or mutant SRSF2 called in this exon due to insufficient coverage. The SRSF2 eCLIP read density from this region is shown in Supplementary Fig. S9E.) Ψ: Percent spliced in. **F, G.** PSI (percent spliced in) of *GNAS* exon 3 in iPSC-HSPCs (F) or MDS patient cells (from Pellagatti et al. 2018(12)) (G). Data points represent independent lines (F) or patients (G). Boxes represent the IQR (25th, 50th and 75th percentiles) and whiskers represent 1.5 times the IQR from the 25th and 75th percentiles. ***p<0.005, *p<0.05, n.s.: not significant (Wilcoxon rank-sum test). **H.** Quantification of *GNAS* isoforms by RT-PCR in WT and SF-mutant iPSC-HSPCs, as indicated. Upper: schematic of the two isoforms. Middle: Ethidium bromide staining showing the two isoforms. Lower: Quantitation by image quantitation software of the two isoforms from the gel shown in the middle panel. **I.** Immunoblot showing the long and short forms of Gαs in WT and SF-mutant iPSC-HSPCs.

Figure 4: GNAS-L is a phenotypic MDS driver and encodes a hyperactive Gαs protein. A. SF-mutant iPSC-HSPCs were transduced with a lentiviral vector encoding either shRNA specifically targeting GNAS-L exon 3 or a scrambled shRNA. GNAS isoform expression was quantified by RT-PCR 48 hours later. One representative experiment out of 3 is shown. The upper and lower bands correspond to GNAS-L and GNAS-S, respectively. The bars show quantification of the respective bands. **B.** Number of methylcellulose colonies from 5,000 SF-mutant iPSC-HSPCs transduced with GNAS-L shRNA or scrambled shRNA. Mean from 6 (WT), 3 (GNAS-L shRNA with two different shRNAs) and 2 (Scrambled shRNA) experiments is shown. **C.** Methylcellulose colonies from WT iPSC-HSPCs transduced with GNAS-L or GNAS-S. Mean

and SEM of 3 different WT lines (WT-1, WT-2, WT-3) is shown. **D.** Crystal structure of Gαs bound to GTPγS (PDB: 1AZT). Demarcated is the disordered region (dotted line) encoded by exon 3. **E.** Time course of [^{35}S]GTPγS binding to Gαs-S and Gαs-L forms with the R201C mutation, the R201H mutation or no mutation. Mean and SEM of 3 experiments is shown. (See also Supplementary Table S4) **F.** Real-time measurement of BodipyGTPγS-FL binding to R201 mutant or WT, Gαs-S or Gαs-L forms, as indicated. Mean and SEM from 3 experiments is shown. AU: arbitrary units. (See also Supplementary Table S4) **G.** Dose-response of cAMP accumulation under isoproterenol stimulation in HEK293 cells lacking endogenous GNAS(29) and stably expressing a cAMP sensor transfected with the indicated constructs. FU: fluorescence units. Data are mean and SEM of 6-7 independent transfections. **H.** Basal and maximal adenylyl cyclase activity (cAMP accumulation) estimated by non-linear curve fit of the dose-response curves shown in G. Mean and SEM of 6-7 independent transfections is shown. *$p<0.05$, **$p<0.005$, ***$p<0.0001$ (See also Supplementary Table S5)
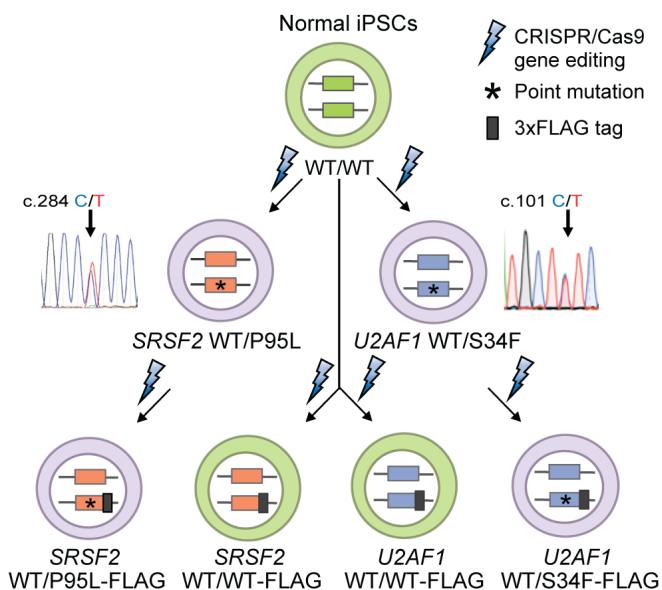
Figure 5: GNAS-L is non-redundant to the R201 mutant form. A. Mutational co-occurrence in combined MDS and AML patient cohorts from Papaemmanuil et al. 2013(4); Papaemmanuil et al. 2016(30) and Haferlach et al. 2014(31). p value was calculated with a Fisher's exact test. OR: odds ratio. **B.** PKA substrate phosphorylation in K562 cells transduced with GNAS-L or GNAS-S with or without the R201H mutation, as indicated (EV: empty vector; S-WT: Short WT; L-WT: Long WT; S-R201H: Short Mutant; L-R201H: Long Mutant). (Note that multiple bands are visible and expected in these blots because a phospho-PKA substrate antibody is used.) **C.** PKA substrate phosphorylation in CB CD34+ cells transduced with the same vectors as in B. **D.** PKA substrate phosphorylation in WT (WT-1) and SF-mutant (P95L-4 and S34F-5) sorted CD34+/CD45+ iPSC-HSPCs. **E.** PKA substrate phosphorylation in primary cells from AML and MDS patients with *SRSF2* mutation with or without *GNAS R201H* mutation.

Figure 6: GNAS-L drives MDS through ERK/MAPK pathway activation. A, B. ERK, MEK and AKT phosphorylation in WT iPSC-HSPCs (WT-1) (A) and CB CD34+ cells (B) transduced with lentiviral vectors expressing GNAS-L or GNAS-S with or without the R201H mutation, as indicated (S-WT: Short WT; L-WT: Long WT; S-R201H: Short Mutant; L-R201H: Long Mutant). mCherry is co-expressed with GNAS in all lentiviral constructs through a P2A peptide. The long and short forms of Gαs (Gαs-L, Gαs-S) can also be seen. **C.** ERK, MEK, CRAF phosphorylation and total DUSP6 in WT and SF-Mutant iPSC-HSPCs (lines WT-1, S34F-6, and P95L-2). **D.** ERK and CRAF phosphorylation and total DUSP6 in WT iPSC-HSPCs transduced with an empty vector (EV) or a lentiviral vector expressing GNAS-L. **E.** ERK phosphorylation in *SRSF2* P95L and *U2AF1* S34F iPSC-HSPCs transduced with lentiviral vectors expressing a GNAS-L shRNA or scrambled shRNA. **F.** ERK phosphorylation and Gαs isoform expression at the protein level in mononuclear cells from three MDS patients with *U2AF1* mutations (S34Y, Q157P and S34F in patients U2AF1-MUT 1, U2AF1-MUT 2 and U2AF1-MUT 3, respectively, see also Supplementary Table S6), two MDS patients with *SRSF2* mutations (P95H and P95R in patients SRSF2-MUT 3 and SRSF2-MUT 6, respectively, see also Supplementary Table S6) and 3 MDS patients without any SF mutations (SF-WT 1-3, see also Supplementary Table S6), as indicated.
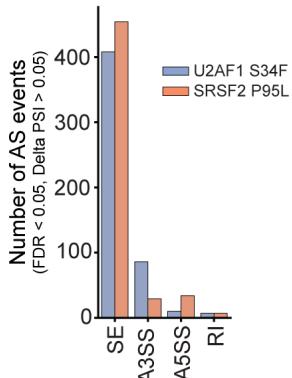
Figure 7: *SRSF2*- and *U2AF1*- mutant cells are sensitive to MEK inhibition. A. Representative IC50 curves of treatment of *SRSF2 P95L* iPSC-HSPCs (P95L-2) with the MEK inhibitors trametinib, cobimetinib, selumetinib and CH5126766 (or the mutant-BRAF inhibitor vemurafenib as negative control) for 3 days. **B.** IC50 values of *SRSF2 P95L* (P95L-2) and

U2AF1 S34F (S34F-6) iPSC-HSPCs treated with the MEK inhibitors trametinib, cobimetinib, selumetinib and CH5126766 (and the mutant-BRAF inhibitor vemurafenib as negative control) for 3 days. Mean and SEM of 3-6 replicates for each MEK inhibitor from independent differentiation experiments is shown. **C.** IC50 values of WT (WT-1), *SRSF2 P95L* (P95L-2) and *U2AF1 S34F* (S34F-6) iPSC-HSPCs treated with the indicated MEK inhibitors (and the mutant-BRAF inhibitor vemurafenib as negative control) for 3 days. Mean and SEM of 3-8 replicates for each MEK inhibitor from independent differentiation experiments is shown. **D.** IC50 curves from primary SF-WT or SRSF2-mutant mononuclear cells from MDS and sAML patients (see Supplementary Table S6 for details) treated with the indicated MEK inhibitors.

Figure 2

Figure 3

Figure 5

Figure 6

Figure 7

CANCER DISCOVERY

AACR American Association for Cancer Research

Integrative RNA-omics discovers GNAS alternative splicing as a phenotypic driver of splicing factor-mutant neoplasms

Emily C Wheeler, Shailee Vora, Daniel Mayer, et al.
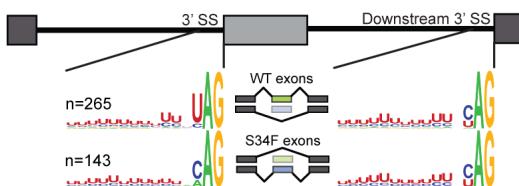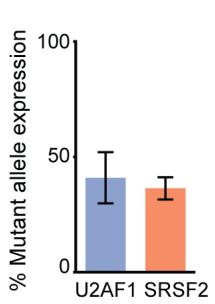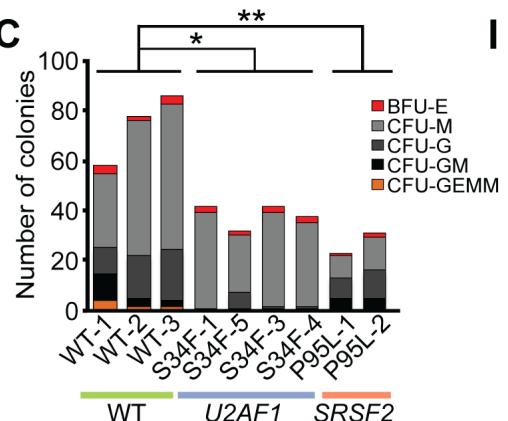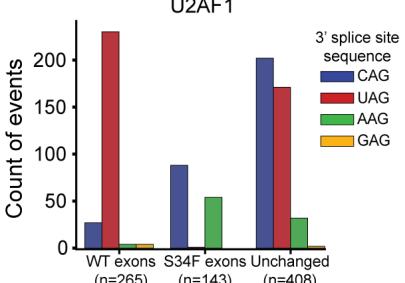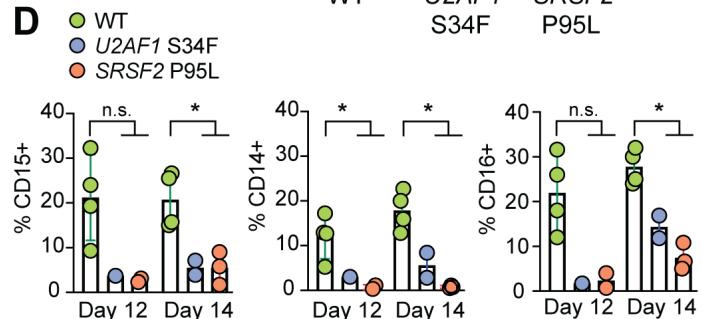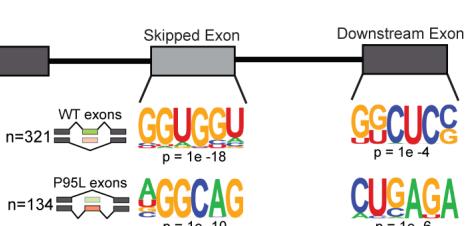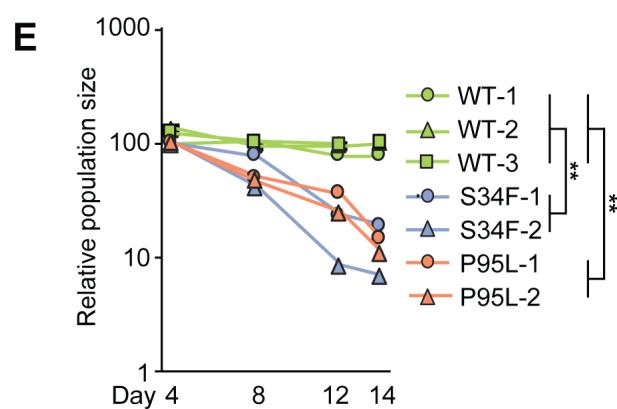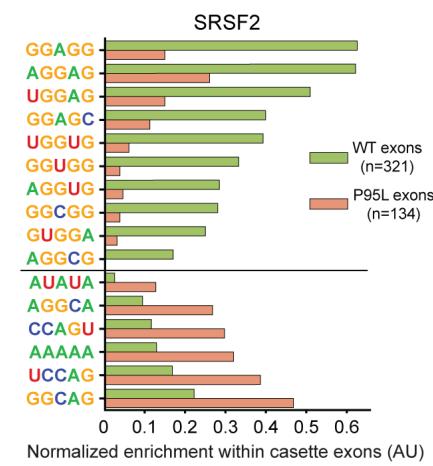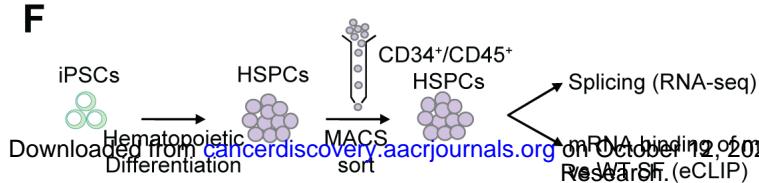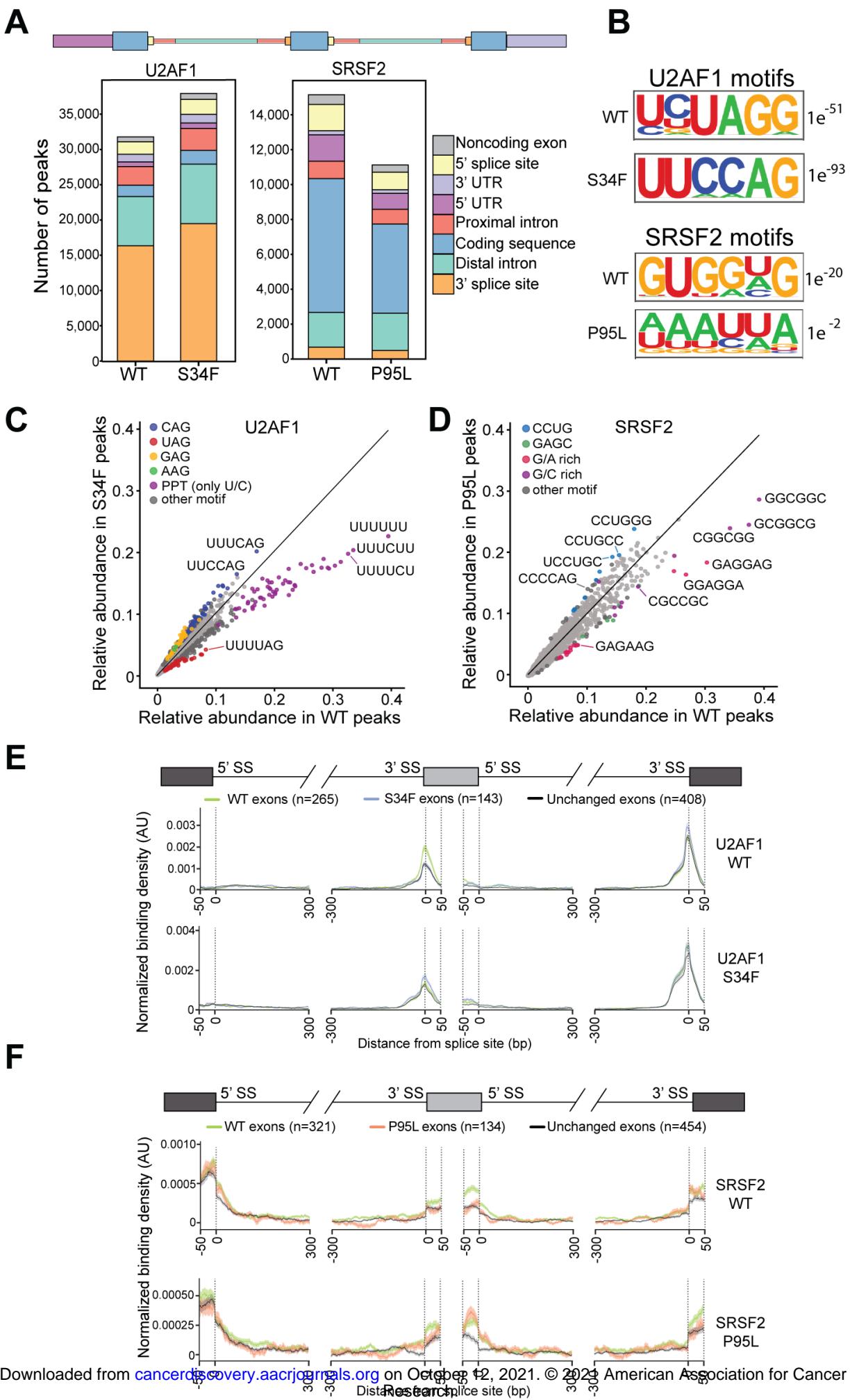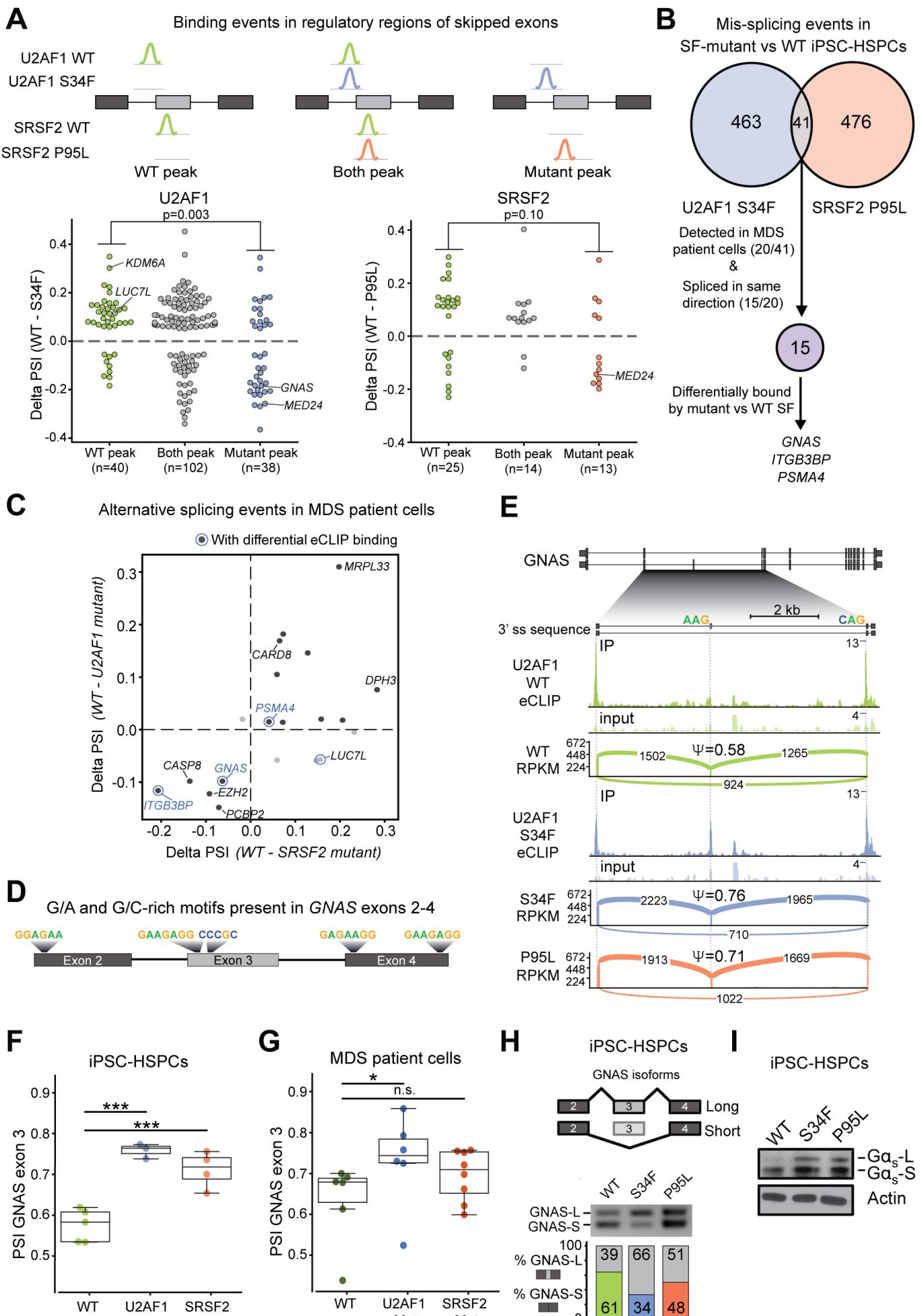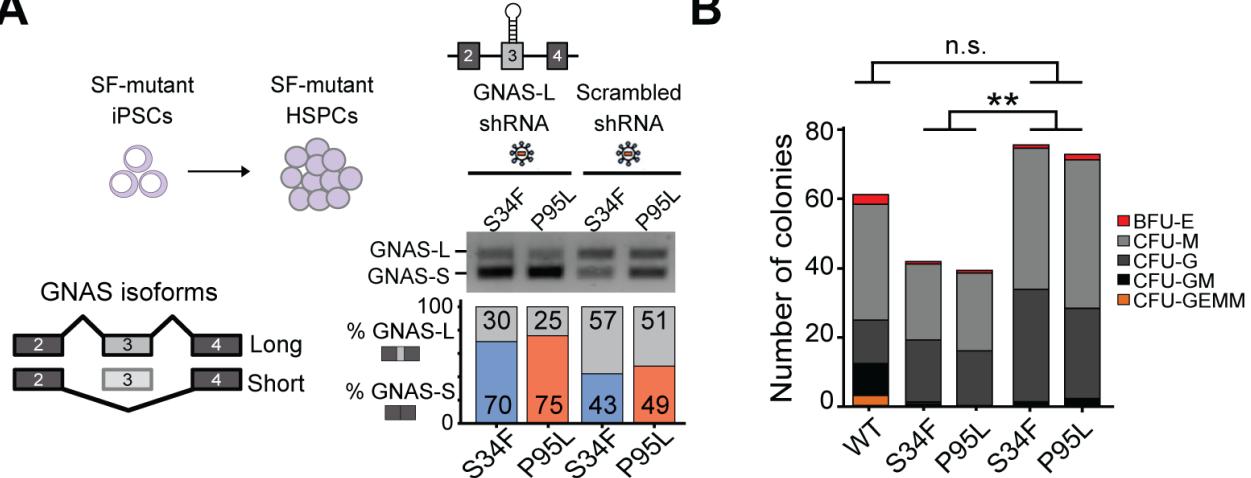
Cancer Discov Published OnlineFirst October 7, 2021.