blood®
Regular Article



MYELOID NEOPLASIA

# Aberrant *EVI1* splicing contributes to *EVI1*-rearranged leukemia

Atsushi Tanaka,[1,2] Taizo A. Nakano,[3] Masaki Nomura,[1,4] Hiromi Yamazaki,[1] Jan P. Bewersdorf,[5] Roger Mulet-Lazaro,[6,7] Simon Hogg,[5] Bo Liu,[5] Alex Penson,[5] Akihiko Yokoyama,[8] Weijia Zang,[1,9] Marije Havermans,[6,7] Miho Koizumi,[10] Yasutaka Hayashi,[1] Hana Cho,[5] Akinori Kanai,[11,12] Stanley C. Lee,[13,14] Muran Xiao,[1,15] Yui Koike,[1] Yifan Zhang,[1,9] Miki Fukumoto,[1] Yumi Aoyama,[1,9] Tsuyoshi Konuma,[16] Hiroyoshi Kunimoto,[17] Toshiya Inaba,[11] Hideaki Nakajima,[17] Hiroaki Honda,[10] Hiroshi Kawamoto,[2] Ruud Delwel,[6,7] Omar Abdel-Wahab,[5,*] and Daichi Inoue[1,*]

[1]Department of Hematology-Oncology, Institute of Biomedical Research and Innovation, Foundation for Biomedical Research and Innovation at Kobe, Kobe, Hyogo, Japan; [2]Laboratory of Immunology, Institute for Life and Medical Sciences, Kyoto University, Kyoto, Japan; [3]Department of Pediatrics, Section of Hematology, Oncology and Bone Marrow Transplantation, University of Colorado, Aurora, CO; [4]Facility for iPS Cell Therapy, CiRA Foundation, Kyoto, Japan; [5]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, United States; [6]Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands; [7]Oncode Institute, Utrecht, The Netherlands; [8]Tsuruoka Metabolomics Laboratory, National Cancer Center, Yamagata, Japan; [9]Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, Kyoto, Japan; [10]Field of Human Disease Models, Major in Advanced Life Sciences and Medicine, Tokyo Women's Medical University, Tokyo, Japan; [11]Department of Molecular Oncology and Leukemia Program Project, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan; [12]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; [13]Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, WA; [14]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA; [15]Division of Cellular Therapy, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [16]Structural Epigenetics Laboratory, and [17]Department of Stem Cell and Immune Regulation, Graduate School of Medicine, Yokohama City University, Yokohama, Kanagawa, Japan

KEY POINTS

- **A novel *EVI1* splice isoform is frequently expressed in inv(3)/t(3;3) leukemia and drives myeloid transformation.**
- **Frequent *SF3B1* mutations in inv(3)/t(3;3) leukemia generate a novel *EVI1* isoform with an altered second zinc finger domain.**

**Detailed genomic and epigenomic analyses of *MECOM* (the MDS1 and EVI1 complex locus) have revealed that inversion or translocation of chromosome 3 drives inv(3)/t(3;3) myeloid leukemias via structural rearrangement of an enhancer that upregulates transcription of *EVI1*. Here, we identify a novel, previously unannotated oncogenic RNA-splicing derived isoform of *EVI1* that is frequently present in inv(3)/t(3;3) acute myeloid leukemia (AML) and directly contributes to leukemic transformation. This EVI1 isoform is generated by oncogenic mutations in the core RNA splicing factor SF3B1, which is mutated in >30% of inv(3)/t(3;3) myeloid neoplasm patients and thereby represents the single most commonly cooccurring genomic alteration in inv(3)/t(3;3) patients. *SF3B1* mutations are statistically uniquely enriched in inv(3)/t(3;3) myeloid neoplasm patients and patient-derived cell lines compared with other forms of AML and promote mis-splicing of EVI1 generating an in-frame insertion of 6 amino acids at the 3′ end of the second zinc finger domain of EVI1. Expression of this EVI1 splice variant enhanced the self-renewal of hematopoietic stem cells, and introduction of mutant SF3B1 in mice bearing the humanized inv(3)(q21q26) allele resulted in generation of this novel EVI1 isoform in mice and hastened leukemogenesis in vivo. The mutant SF3B1 spliceosome depends upon an exonic splicing enhancer within *EVI1* exon 13 to promote usage of a cryptic branch point and aberrant 3′ splice site within intron 12 resulting in the generation of this isoform. These data provide a mechanistic basis for the frequent cooccurrence of *SF3B1* mutations as well as new insights into the pathogenesis of myeloid leukemias harboring inv(3)/t(3;3).**

## Introduction

Acute myeloid leukemia (AML) with inv(3)(q21q26) or t(3;3)(q21q26) is recognized by the World Health Organization as a unique subtype of AML (henceforth referred to as "inv(3)/t(3;3) AML"),[1] which has a dismal median overall survival (OS) of <1 year after diagnosis.[2-5] This same genomic alteration is also occasionally encountered in patients with myelodysplastic syndromes (MDS), where it is similarly associated with short survival.[4,6] Given the poor outcome of inv(3)/t(3;3) MDS/AML, there have been extensive efforts to dissect the genomic and epigenomic events that give rise to this aggressive disease. Recent efforts have identified that inv(3)/t(3;3) chromosomal rearrangements reposition the *GATA2* enhancer from its normal location at 3q21 to drive ectopic expression of the *EVI1* proto-oncogene from within *MDS1 and EVI1 complex locus* (*MECOM*) at 3q26 (supplemental Figure 1A, available on the *Blood* Web site). Intergenic splicing of *MDS1* and *EVI1* normally results in expression of an *MDS1-EVI1* transcript from *MECOM*, but in patients

with inv(3)/t(3;3) chromosomal alterations, the full-length *MDS1-EVI1* transcript is no longer expressed, and there is monoallelic expression of *GATA2*.[7,8]

*EVI1* encodes a transcription factor that is indispensable for hematopoiesis and contains 2 zinc finger (ZF) DNA binding domains, one at the N-terminus and the other at the protein's C-terminus, which includes 7 and 3 ZFs, respectively. In contrast to our understanding of the pathogenic consequences of inv(3)/t(3;3) chromosomal rearrangements, the role of genomic alterations coexisting with inv(3)/t(3;3) is less well understood. For example, monosomy 7 and mutations in *RUNX1*, *IKZF1*, and RAS pathway genes (*NRAS*, *KRAS*, *PTPN11*, and *NF1*)[5,9-11] are known to occur in inv(3)/t(3;3) AML. However, the specific contribution of additional genomic events to inv(3)/t(3;3) leukemia is not clear.

Here, we identify that mutations in the core RNA splicing factor *SF3B1* are the most common coexisting genetic alterations among patients with inv(3)/t(3;3) MDS and AML. Introduction of mutant SF3B1 in mice bearing the humanized inv(3)(q21q26) allele hastened leukemogenesis. Surprisingly, mutant SF3B1 resulted in the generation of a novel isoform of *EVI1*, which alters its second ZF domain and promotes oncogenicity in vivo. This particular oncogenic EVI1 isoform is an unannotated novel transcript and is expressed in nearly one-third of patients with inv(3)/t(3;3) MDS and AML. These data thereby elucidate a mechanistic basis for the frequent cooccurrence of *SF3B1* mutations in inv(3)/t(3;3) leukemias, identify a novel previously unknown oncogenic form of *EVI1*, and delineate a role for pathologic splicing in inv(3)/t(3;3) AML.

# Methods

## Patient samples

Studies were also approved by the institutional review boards of Memorial Sloan Kettering Cancer Center (MSKCC) and The Medical Ethical Committee of the Erasmus MC and conducted in accordance with the Declaration of Helsinki protocol. Informed consents were obtained from all human subjects. Next-generation sequencing was performed on DNA extracted from bone marrow (BM) mononuclear cells and matched normal from fingernails. Patient samples were sequenced with MSK-IMPACT targeted sequencing panel, with somatic mutations (substitutions and small insertions and deletions), gene-level focal copy number alterations, and structural rearrangements detected with a clinically validated pipeline as previously described.[12,13] All patients with myeloid neoplasms (AML, MDS, chronic myelomonocytic leukemia, and blast-phase chronic myeloid leukemia) and inv(3)/t(3;3) were identified from the electronic medical record and included in this study if next-generation sequencing data were available.

## Animals

All animals were housed at MSKCC and at Foundation for Biomedical Research and Innovation (FBRI, Japan) using a 12-hour light/12-hour dark cycle and with ambient temperature maintained at 72°F ± 2°F (~21.5°C ± 1°C) with 30% to 70% humidity. All animal procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSKCC and FBRI. All mouse experiments were performed in accordance with a protocol approved by the MSKCC (11-12-029) and FBRI (18-06) Institutional Animal Care and Use Committee. *Mx1*-Cre and *Sf3b1*$^{K700E}$ mice were obtained from The Jackson Laboratory and were previously generated, respectively.[14] inv(3)(3q21q26) mouse strain (RBRC09508) was provided by RIKEN BRC through the National BioResource Project of the MEXT/AMED, Japan.[7] All of the primers and polymerase chain reaction (PCR) conditions are listed in supplemental Table 1.

## Cell lines and tissue culture

HEK293T and PlatE cells were obtained from American Type Culture Collection (Manassas, VA) and Toshio Kitamura (University of Tokyo) and cultured in Dulbecco's modified Eagle medium with 10% fetal bovine serum (FBS). K562, MEL270, MOLM-1, Kasumi-3, HNT-34, and 5637 cells were cultured in RPMI 1640 with 10% FBS. Kasumi-4 cells were cultured in RPMI 1640 with 20% FBS and 10 ng/mL granulocyte-macrophage colony-stimulating factor (AF-300-03, PeproTech). YCU-AML1 cells were cultured with OP-9 in Iscove's modified Dulbecco's medium with 10% FBS, 55 μM β-mercaptoethanol (Sigma-Aldrich) and 20 ng/mL granulocyte-macrophage colony-stimulating factor (PeproTech). MUTZ-3 cells were cultured in α minimum essential medium (with ribo- and deoxyribonucleosides)/20% FBS and 20% conditioned medium of cell line 5637. All cell culture media include 100 U/mL penicillin and 100 μg/mL streptomycin (Gibco).

## *EVI1* minigene assay

K562 parental and K562-*SF3B1*$^{K666N/WT}$ cells were seeded into a 12-well plate with culture medium 48 hours before transfection of minigene constructs (more details are provided in supplemental methods) in the presence of X-tremeGENE HP DNA transfection reagent (Roche) according to the manufacturer's directions. Forty-eight hours after transfection, cells were collected, and RNA was extracted using RNeasy mini kit (Qiagen). Extracted RNA was treated with DNase I (Ambion) to ensure complete removal of DNA. Minigene-derived and endogenous *EVI1* transcripts were analyzed by reverse transcription (RT)-PCR using specific primers (supplemental Table 1).

# Results

## *SF3B1* mutations are frequent within inv(3)/t(3;3) AML

Prior genomic analyses of cohorts of patients with inv(3)/t(3;3) myeloid neoplasm identified a number of genomic rearrangements coexisting with inv(3)/t(3;3) rearrangement.[10,11] To further evaluate the landscape of genomic alterations within inv(3)/t(3;3) myeloid neoplasms, we compiled prior cohorts of genomic analysis of patients with inv(3)/t(3;3) myeloid neoplasm (n = 63 patients from the Dutch-Belgian Cooperative Trial Group for Hematology-Oncology/the German-Austrian AML Study Group, Leucegene, and BeatAML)[10,11,15] with a previously unpublished cohort of 46 patients with inv(3)/t(3;3) myeloid neoplasm (supplemental Table 2). Interestingly, this revealed that the single most commonly mutated gene in patients with inv(3)/t(3;3) myeloid neoplasm was the core RNA splicing factor *SF3B1* (Figure 1A), although RAS pathway mutations (*NRAS*, *PTPN11*, *KRAS*, and *NF1*) as a group in aggregate (present in 63 out of 109 patients [57.8%]) are more frequent than mutations in

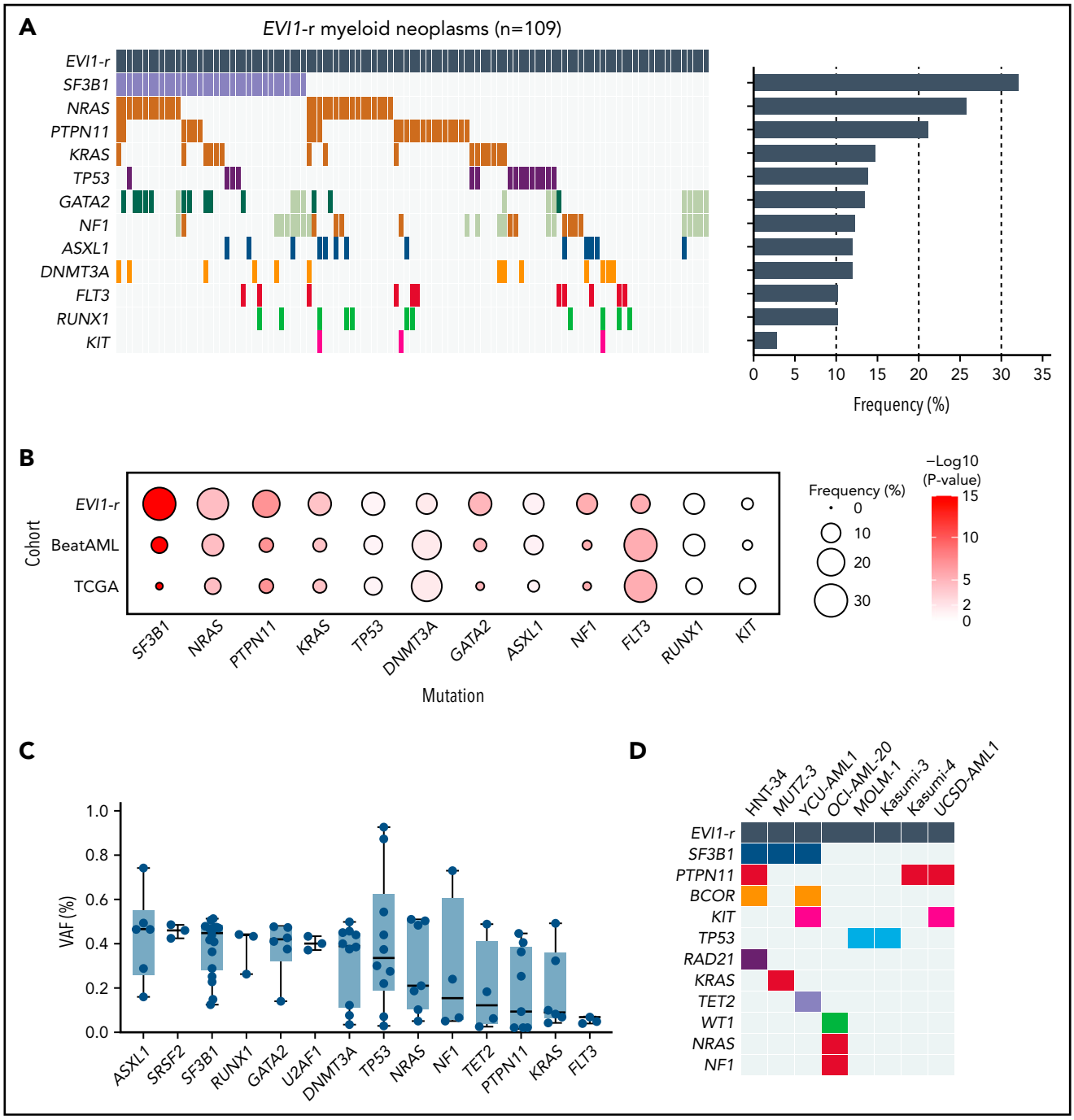



**Figure 1. Frequent cooccurrence of *SF3B1* mutations in myeloid malignancies with inv(3)(q21q26) or t(3;3)(q21q26).** (A) Oncoprint of recurrently mutated genes in 109 patients with *EVI1*-rearranged (*EVI1*-r) myeloid neoplasms. Horizontal bars show the mutational frequency of each gene. Gray color indicates data not available. (B) Frequency (indicated by bubble size) and statistical enrichment (indicated by color gradient) of mutations (x-axis) across AML (y-axis; inv(3)/t(3;3) patients from panel A, n = 109; BeatAML study, n = 622; TCGA AML study, n = 200). *P* values of Fisher's exact test are color-coded. (C) VAF of mutations in *SF3B1* and *RAS*-associated genes relative to mutations in transcriptional factors, chromatin modifiers, RNA splicing factors in patients with *EVI1*-r myeloid neoplasm. (D) Oncoprint of recurrently mutated genes in *EVI1*-r leukemia cell lines.

*SF3B1*. *SF3B1* mutations were present in 32.1% of patients and located at hotspot residues in HEAT repeat domains, such as K700 and K666 (supplemental Figure 1B).[16,17] In contrast to the high frequency of mutations in *SF3B1* within *EVI1*-rearranged myeloid neoplasms, mutations in *SRSF2* or *U2AF1* were not uniquely enriched among *EVI1*-rearranged AML combined with overall patients with AML from TCGA/BeatAML cohorts (6.52% vs 8.03% in *SRSF2* mutation [*P* = .6126], 6.52% vs 4.33% [*P* = .4485] in *U2AF1* mutation).[15,18] RNA splicing factor mutations were mutually exclusive in our MSKCC cohort except for 1 patient harboring *SF3B1*-K666R (variant allele frequencies [VAFs] = 47.1%) and *SRSF2*-P95H (VAF = 42.4%).

This high frequency of *SF3B1* mutations within inv(3)/t(3;3) AML was particularly conspicuous because *SF3B1* mutations are most commonly enriched in patients with myeloid leukemia with MDS

with ringed sideroblasts (MDS-RS) (65%-81%),[16,17,19-21] whereas they are relatively rare in AML (~4%).[22] Indeed, compared with mutational analyses of AML cohorts without chromosome 3 alterations,[15,18] inv(3)/t(3;3) AML was characterized by a remarkably higher rate of mutations in *SF3B1* ($P = 1.16 \times 10^{-17}$), RAS-associated pathway genes, and *GATA2*. On the other hand, *EVI1*-rearranged (*EVI1*-r) AML exhibited lower rates of mutations in *FLT3* and *DNMT3A*, when compared with AML without *EVI1*-r (Figure 1B). In contrast to RAS-associated pathway mutations, *SF3B1* mutations tended to contribute to the founder clones in most patients with *SF3B1*-mutated *EVI1*-r myeloid based on the relative VAFs (median VAF of *SF3B1* mutations in inv(3)/t(3;3) samples was 0.448) (Figure 1C). Interestingly, across all inv(3)/t(3;3) human AML cell lines (HNT-34, Kasumi-3, Kasumi-4, MOLM-1, MUTZ-3, OCI-AML20, and UCSD-AML1, and YCU-AML1), 3 of 8 of these lines contain a well-described heterozygous *SF3B1* hotspot mutation (Figure 1D; supplemental Tables 3 and 4). In fact, these 3 cell lines (HNT-34, MUTZ-3, and YCU-AML1) represent the only known AML cell lines with naturally occurring *SF3B1* mutations. Moreover, in the analysis of correlations across all pairwise combinations among 12 frequently mutated drivers, inv(3)/t(3;3) AML exhibited a distinct pattern of positive and negative correlations relative to that in patients with whole MDS/AML (supplemental Figure 1C).[23] For example, *SF3B1* is significantly comutated with *GATA2* within inv(3)/t(3;3) myeloid neoplasms, whereas this correlation was not detected in wider MDS/AML cohorts.[23] These results highlight a unique enrichment of *SF3B1* mutations in inv(3)/t(3;3) AML compared with other AML subtypes. In our MSKCC cohort, we found no significant difference between *SF3B1*-mutated and wild-type (WT) patients in OS from the time of inv(3)/t(3;3) detection (188 vs 179 days; hazard ratio 0.83 [0.41-1.69]). Moreover, OS from the time of initial disease diagnosis was comparable (520 vs 366 days; hazard ratio 0.63 [0.31-1.27]) (supplemental Figure 1D).

## *SF3B1* mutations promote leukemogenicity in humanized inv(3)(q21q26) mice

Given the recurrent nature of *SF3B1* mutations in inv(3)/t(3;3) AML, we hypothesized that spliceosomal alterations are important in the development of inv(3)/t(3;3) myeloid malignancies. We therefore set out to test this hypothesis by generating a mouse model permitting time- and tissue-specific induction of mutant *Sf3b1* in the presence of the human inv(3)(q21q26) allele. This was accomplished by generating *Mx1*-Cre *Sf3b1*$^{K700E/WT}$ inv(3)(q21q26) mice (along with single-mutant and WT control mice; Figure 2A). We used transgenic mice harboring a human bacterial artificial chromosome encompassing the human inv(3)(q21q26) allele[7] (hereafter referred to as "inv(3) mice") (supplemental Figure 1A) and crossed these animals to mice with conditional knock-in of mutant *Sf3b1* K700E.[14] We thereby generated animals with 4 different genotypes: *Mx1*-Cre control, *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$.

We first evaluated the impact of mutant SF3B1 on the clonogenic capacity of inv(3)(q21q26) hematopoietic cells. We collected whole BM cells 4 weeks after polyinosinic-polycytidylic acid (pIpC) injection and evaluated colony formation in methylcellulose media optimized for each hematopoietic lineage (Figure 2B). Hematopoietic cells from either *Mx1*-Cre inv(3) or *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ mice produced virtually no BFU-E erythroid progenitor or pre-B lymphoid progenitor colonies (supplemental Figure 2A). In contrast, these BM cells enhanced the clonogenic capacity of myeloid progenitors. Although both were replatable into the fifth round, *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ produced more colonies in the second and fifth platings than inv(3) mice alone.

We next evaluated the cell-autonomous effects of hematopoietic stem cells from each mouse model by performing BM transplantation assays into lethally irradiated CD45.1 recipient mice (Figure 2A). After confirming successful engraftment, we treated recipient mice with pIpC 4 weeks after transplantation. Interestingly, inv(3) *Sf3b1*$^{K700E/WT}$ double-mutant mice exhibited significant leukopenia and macrocytic anemia compared with either mutation alone (Figure 2C; supplemental Figure 2B). However, 6 months after pIpC injection, hematopoietic stem and progenitor cell (HSPC) fractions, including LSK ($Lin^-c\text{-}Kit^+Sca1^+$), multipotent progenitors 2/3 (MPP2/3; $CD135^-CD150^+CD48^+$LSK and $CD135^-CD150^-CD48^+$LSK, respectively), and common myeloid progenitors ($Lin^-c\text{-}Kit^+Sca1^-CD34^+Fc\gamma R^-$) (Figure 2D-E; supplemental Figure 2C), were significantly expanded in the double-mutant model, suggesting ineffective hematopoiesis. In the peripheral blood, myeloid-lineage skewing was observed at the expense of B-cell commitment, which was most significant in inv(3)/Sf3b1 double-mutant mice (Figure 2F). In addition to splenomegaly, histological and morphological analysis of BM and spleen cells in the double-mutant mice revealed hypercellularity, destruction of normal architecture, morphological abnormalities, and frequent immature blasts with slight differentiation toward myeloid lineage (Figure 2G; supplemental Figure 3). As a result, *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ mice had hastened death owing to MDS and AML (supplemental Figure 2D-E) even when compared with *Mx1*-Cre inv(3) ($P = .0389$) (Figure 2H). Moreover, serial transplantation of double-mutant mouse cells led to more rapidly lethal leukemia vs inv(3) leukemia cells alone (Figure 2I). These results indicate that the *SF3B1* mutation enhanced and accelerated inv(3)-associated myeloid malignancies in a genetically accurate murine model.

## Inv(3) rescues the fitness disadvantage of *SF3B1*-mutated HSPCs

A series of studies reported significantly lower peripheral blood chimerism derived from *Sf3b1*$^{K700E/WT}$ mouse hematopoietic precursors in competitive transplantation,[14,24] suggesting that *SF3B1* mutations impair cell-autonomous repopulating when present alone. We therefore evaluated the in vivo self-renewal of *Mx1*-Cre control, *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ models by performing BM competitive transplantation assays. Equal numbers of $CD45.2^+$ BM cells from each of these models were mixed with $CD45.1^+$ WT BM cells and injected into lethally irradiated CD45.1 recipient mice (supplemental Figure 4A). Five months later, whole BM cells of each group were serially transplanted into new $CD45.1^+$ recipient mice to determine the further reconstitution capacity. Interestingly, in stark contrast to the near-complete loss of hematopoiesis in *Mx1*-Cre *Sf3b1*$^{K700E/WT}$ mice, the inv(3) transgene rescued the impaired in vivo clonogenic capacity of *Sf3b1*-mutated HSPCs, especially in secondary transplant (supplemental Figure 4B). This was associated with rescued chimerism of myeloid-lineage cells in the BM as well as stem and progenitor cells in BM 5 months following primary

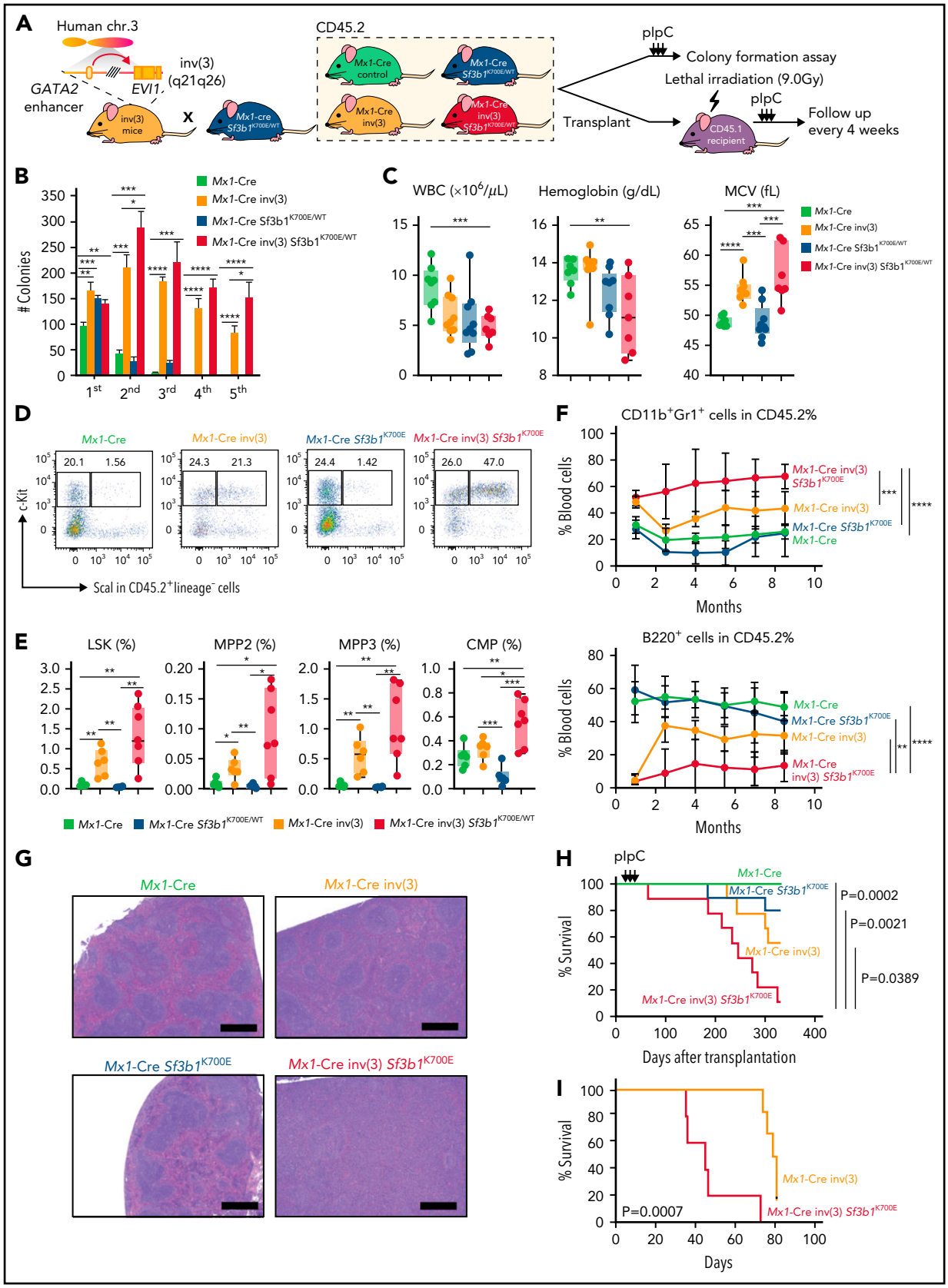



**Figure 2.**

transplantation. In line with the results of in vitro colony-forming assay, no rescue was observed in B220$^+$ cells (supplemental Figure 4C-D).

## Transcriptome analysis of SF3B1 mutant/inv(3) HSPCs

We next performed RNA-seq analyses of mouse and human leukemias to identify the impact of inv(3) and *SF3B1* mutations alone and together on splicing and gene expression. RNA-seq analysis of fluorescence-activated cell sorter–purified lineage-negative c-Kit$^+$ cells of the mouse models described above was performed 2 months after pIpC injection, in biological triplicate for each model. Unsupervised hierarchical clustering analysis, using Euclidean distance and Complete linkage method, revealed that transcriptomes of lineage-negative c-Kit$^+$ cells from *Mx1*-Cre inv(3) and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ groups differed substantially from *Mx1*-Cre control and *Mx1*-Cre *Sf3b1*$^{K700E/WT}$ groups, suggesting potent gene expression effects of the inv(3) transgene (Figure 3A). Similarly, principal component (PC) analysis revealed differences in the transcriptome of *Mx1*-Cre inv(3) and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ groups vs *Mx1*-Cre control *Mx1*-Cre *Sf3b1*$^{K700E/WT}$ groups. PC1 accounted for 65% of the variance and defined the presence of the inv(3) transgene (Figure 3B). At the same time, PC2 clearly distinguished *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ from *Mx1*-Cre inv(3), indicating that Sf3b1 mutant exerted additive gene expression effects on inv(3) HSPCs, which is consistent with the collaborative biological outcomes of double-mutant mice described above (Figure 2).

Differential gene expression analysis identified 6659, 2609, and 6796 genes dysregulated compared with control (Student *t* test, $P < .01$; fold change >2 or <0.5) in inv(3), *Sf3b1*$^{K700E/WT}$, inv(3) *Sf3b1*$^{K700E/WT}$ mice, respectively. Interestingly, a large portion of differentially expressed genes in the inv(3) *Sf3b1*$^{K700E/WT}$ group were shared with those of the inv(3) group rather than the *Sf3b1*$^{K700E/WT}$ group (65.9%, 4481/6796 genes vs 22.3%, 1514/6796 genes) (Figure 3C). We identified significant enrichment for Gene Ontology (GO) and pathway terms reflecting the development of MDS/AML in inv(3)/Sf3b1 double-mutant mice, including MAP kinase, interferon, tumor necrosis factor-α, and NF-κB signaling as well as interleukin-6 activation (Figure 3D; supplemental Figure 5). Consistent with the results in Figure 3C, the dysregulated pathways and GO terms were mainly attributed to the effect of inv(3), especially in upregulated pathways (Figure 3D; supplemental Figure 5A).

In order to evaluate these findings in patient samples, we analyzed RNA-seq data of unfractionated BM mononuclear cells isolated from 7 *SF3B1*-mutant/*EVI1*-r, 5 *SF3B1*-WT/*EVI1*-r, 2 *SF3B1*-mutant, and 7 *SF3B1*-WT normal karyotype AML patient samples. In line with the mouse RNA-seq results (Figure 3C-D), in the presence of the *EVI1*-r, the *SF3B1* mutation had limited impact on gene expression in terms of dysregulated pathways evaluated by using GO enrichment terms in the algorithm of Enrichr (https://maayanlab.cloud/Enrichr/) (Figure 3E). As such, a large proportion of differentially expressed genes in the *SF3B1*-mutant/*EVI1*-r group overlapped with those of the *SF3B1*-WT/*EVI1*-r patient samples (Figure 3E).

We next sought to identify aberrant splicing events associated with *SF3B1* mutation and *EVI1*-r. Of note, the number of aberrantly spliced genes was largest in *SF3B1*-mutant/inv(3) patient samples where the majority of splicing changes could be attributed to the presence of the *SF3B1* mutation (Figure 3F). Consistent with previous studies,[25,26] aberrant 3′ splice site (3′ss) usage was the most prevalent mis-splicing event in the *SF3B1*-mutant/inv(3) group (Figure 3G), although such effects were relatively modest in murine models (supplemental Figure 5B). These transcriptome analyses suggested cooperative effects of the *SF3B1* mutation and inv(3) allele on aberrant splicing and gene expression, respectively.

## *EVI1* mis-splicing in *SF3B1*-mutant cells

As noted in prior studies,[24,27,28] despite striking similarities in global aberrant splicing patterns induced by SF3B1 mutants across diverse species, there was modest overlap in aberrantly spliced events in mouse vs human cells. Interestingly, however, we identified that human *EVI1* itself was recurrently targeted in human *SF3B1*-mutant/inv(3) patient samples (Figure 4A).

Intriguingly, the *SF3B1* mutation was associated with aberrant 3′ss selection at the intron 12-exon 13 junction of *EVI1* (NM_001105078.4), which encodes the C-terminal end of the second ZF domain of EVI1. EVI1 contains 10 ZFs that are arranged in 2 separate domains, each of which binds different consensus DNA sequences (Figure 4A).[29] This *EVI1* mis-splicing event was exclusively observed in *SF3B1*-mutant myeloid malignancy patient samples and cell lines and present across *SF3B1* hotspot mutations (Figure 4B; supplemental Table 5). For example, the inv(3)/t(3;3) AML cell lines HNT-34 and YCU-AML1 harboring *SF3B1*K700E heavily express this variant *EVI1* isoform (Figure 4C).[30,31] Expression of several *SF3B1* mutations (K700E, K666N, and G740E), but not WT *SF3B1*, into K562 cells similarly resulted in the aberrant *EVI1* isoform generation (supplemental Figure 6A). Sanger sequencing of this isoform verified that this splice variant results in the introduction of 18 nucleotides inserted between exons 12 and 13 of the *EVI1* transcript, at the very 3′ end of intron 12 (Figure 4D). This nucleotide insertion gives rise to an in-frame insertion of 6 amino acids (FLLHTG) (Figure 4D). This exact same +18 *EVI* splice variant was

**Figure 2. *SF3B1* mutations enhance the leukemogenicity of hematopoietic cells expressing the inv(3)(q21q26) allele.** (A) Schema of generation of CD45.2 *Mx1*-cre inv(3) *Sf3b1*$^{K700E/WT}$ mice (left) and schema of in vitro and in vivo analyses of hematopoiesis from these mice and single-mutant controls. (B) Number of myeloid colonies on first to fifth plating of *Mx1*-cre inv(3) *Sf3b1*$^{K700E/WT}$ mice and controls. (C) Box-and-whisker plots of white blood cell count (WBC), hemoglobin, and mean corpuscular volume (MCV) from CD45.1 recipient mice following 8.5 months of transplantation of CD45.2 mice from panel A. For box-and-whiskers plots throughout, bar indicates median, box edges first and third quartile values, and whisker edges minimum and maximum values. (D) Representative fluorescence-activated cell sorter plots of CD45.2$^+$ LSK (lineage-negative Sca1$^+$ and c-Kit$^+$) and LK (lineage-negative Sca1$^-$ and c-Kit$^+$) cells from BM of CD45.1 recipient mice at 4 months posttransplant. % of cells within gate is shown. (E) Box-and-whisker plots of percentage of BM CD45.2$^+$ LSK, multipotent progenitor cells 2 and 3 (MPP2 and MPP3, respectively), and common myeloid progenitor (CMP) cells. (F) % of CD11b$^+$Gr1$^+$ and B220$^+$ cells among CD45.2$^+$ cells in peripheral blood over time following transplantation. Mean ± standard deviation are shown. (G) Representative hematoxylin-and-eosin stain (original magnification ×100) of spleen of CD45.1 primary recipient mice. Scale bars, 400 μm. (H) Kaplan-Meier survival curve of primary CD45.1 recipient mice. *P* values were calculated by log-rank test. (I) Kaplan-Meier survival curve of secondarily transplanted CD45.1 recipient mice following sublethal irradiation (4.5 Gy). *P* values were calculated by 2-sided Student *t* test or log-rank test. *$P < .05$, **$P < .01$, ***$P < .001$, and ****$P < .0001$. chr., chromosome.

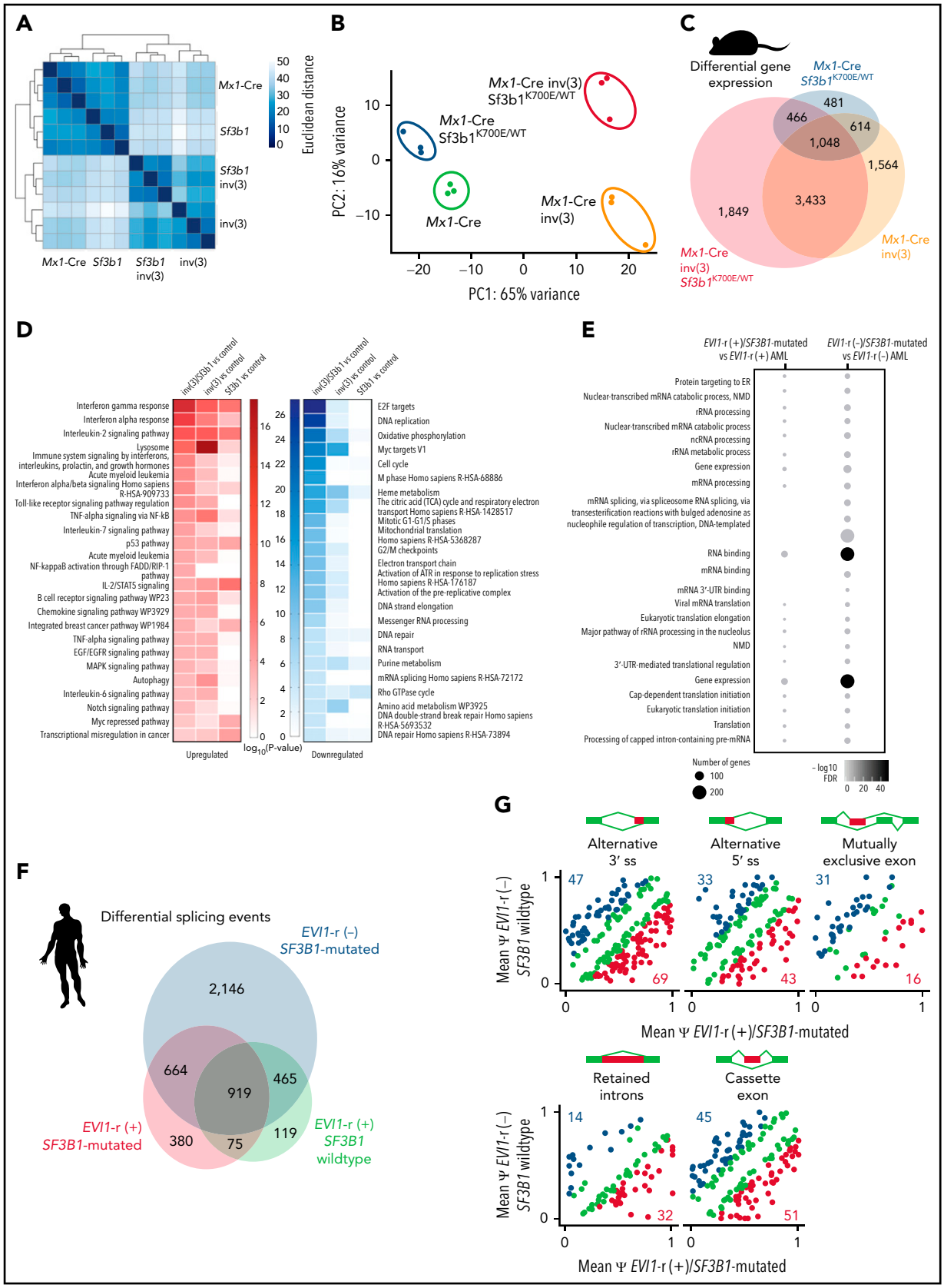



**Figure 3.**

expressed in *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ double-mutant mice, indicating that murine Sf3b1 mutant similarly induces aberrant 3′ss selection of human *EVI1* contained within the bacterial artificial chromosome transgene recapitulating the inv(3)(q21q26) allele (Figure 4E; supplemental Figure 6B). Interestingly, a model structure by AlphaFold2,[32] which predicts 3-dimensional protein structure from amino acid sequence using a deep learning system, revealed that the EVI1 + 18 splicing event introduces 6 amino acids immediately after the second ZF domain of EVI1 (supplemental Figure 6C). This change in protein sequence may alter DNA recognition by the second ZF domain. Hereafter, the newly discovered *EVI1* (*MECOM*) splice variant will be referred to as "*EVI1*+18."

## Functional impact of the novel *EVI1*+18 splice variant

We next sought to understand the biological effects of the aberrant +18 isoform and generated a pMYs-IRES-GFP retrovirus vector to express WT *EVI1* (NM_001105078.4, EVI1-145 kDa) or *EVI1*+18 cDNA. Several studies have demonstrated that ectopic *EVI1* expression in hematopoietic stem cells leads to increased colony output and immortalization in methylcellulose media,[33] which was confirmed by our experiments. Interestingly, however, *EVI1*+18 further enhanced proliferation capacity compared with WT *EVI1* in colony formation assay (Figure 4F). We performed RNA-seq analysis of the fourth colonies and validated the dysregulated genes by quantitative RT-PCR. This revealed prominent upregulation of *Hes1*, *Meis1*, *Bcl11a*, and *Cd34* in cells expressing the *EVI1*+18 isoform (supplemental Figure 6D). In addition, we observed a competitive advantage of *EVI1*+18 expressing HSPCs over WT *EVI1*-transduced HSPCs (supplemental Figure 6E-F). Consistent with these data, gene set enrichment analysis demonstrated that genes upregulated in leukemic stem cells are enriched in *EVI1*+18 transduced K562 cells compared with the *EVI1* WT-expressing cells (supplemental Figure 6G). These results indicate that SF3B1-mutants generate a previously unknown *EVI1* variant with enhanced self-renewal capacity.

To test if *EVI1*-r/*SF3B1*-mutated AML cells depend on the aberrant splicing machinery, we evaluated the 50% inhibitory concentration of *EVI1*-r cells with or without *SF3B1* mutation to indisulam, a selective degrader of the RNA splicing factor RBM39.[34] Of note, 3-day dose-response experiments revealed that HNT-34 and MUTZ-3, both of which harbor *SF3B1*$^{K700E}$/*EVI1*r, were sensitive to indisulam compared with *SF3B1* WT/*EVI1*-r cell lines, Kasumi-3 and K562 cells with t(3;8),[35] as well as *SF3B1* WT/*EVI1* WT cell lines (K562) (supplemental Figure 7). These data suggest that spliceosomal disruption may be therapeutically effective against *EVI1*-rearranged AML with *SF3B1* mutations.

Given the altered second ZF domain of EVI1 + 18, we next sought to evaluate the genomic distribution and transcriptional effects of this EVI1 isoform compared with the most common annotated EVI1 isoform produced in inv(3) AML. To identify genome-wide binding preferences of EVI1 vs EVI1 + 18, we performed anti-EVI1 chromatin immunoprecipitation (ChIP)-seq using HNT-34 (*SF3B1*$^{K700E}$/*EVI1*-r), MUTZ-3 (*SF3B1*$^{K666N}$/*EVI1*-r), and MOLM-1 (*SF3B1* WT/*EVI1*-r) cells, as well as 1 primary *SF3B1*$^{K700E}$/*EVI1*-r AML patient sample (Figure 4G). Leukemias with concomitant *SF3B1* mutation and *EVI1*-r, but not the *SF3B1* WT/*EVI1*-r cell line, expressed the *EVI1*+18 isoform (Figure 4C). Interestingly, we identified 5698 exclusive *EVI1*+18 peaks in *SF3B1* mutant/*EVI1*-r cell lines but not in MOLM-1 (*SF3B1* WT/*EVI1*-r) cells (supplemental Figure 8A). The majority of such peaks (5206 of 5698 peaks, 91.4%) were also detected in primary *SF3B1*$^{K700E}$/*EVI1*-r AML cells expressing *EVI1*+18, and nearly all are located at promoters. Of note, transcription factor enrichment analysis identified PU.1 (SPI1) motifs significantly enriched in EVI1 + 18-specific peaks (supplemental Figure 8B). Finally, we visualized these ChIP-seq results at several genes important for leukemic transformation, such as *MEIS1* (Figure 4H; supplemental Figure 8C). Given that expression of *Meis1*, a crucial regulator of leukemogenesis,[36] was remarkably increased in the immortalized colonies with *EVI1*+18 (compared with those with *EVI1* WT; supplemental Figure 6D), we speculate that the link between *EVI1*+18 and oncogenic transcriptional program contributes to leukemia development. We also performed ChIP-seq using exogenously EVI1- vs EVI1 + 18-expressing 293T cells (supplemental Figure 9A). In such a model, the genomic distributions of EVI1 and EVI1 + 18 were not strikingly different (supplemental Figure 9B-E) except for a limited number of regions, indicating that the transcriptional regulation by the endogenous promoter and cellular contexts may be important.

## Molecular regulation of aberrant *EVI1* splicing by mutant SF3B1

SF3B1, as part of the U2 small nuclear ribonucleoprotein complex, is responsible for recognition of the intronic branchpoint sequence (BPS), which facilitates 3′ss selection.[16,25] To identify *cis*-acting elements, including BPS, polypyrimidine tract, and exonic splicing enhancers (ESEs), required for *EVI1* aberrant splicing by SF3B1 mutants, we generated a minigene construct containing the mis-spliced intron (intron 12) and flanking exons (exons 12 and 13).[37,38] We then mutagenized a variety of sequences within the minigene and transduced WT and mutant minigenes into *SF3B1* WT and K666N knock-in K562 cells to identify those *cis* elements within *EVI1* essential for aberrant splicing by mutant SF3B1 (Figure 5A). As expected, mutant SF3B1, but not WT SF3B1, induced the *EVI1*+18 variant in both the minigene-derived RNA and the endogenous RNA (Figure 5B). Disruption of the cryptic 3′ss eliminated the *EVI1*+18

**Figure 3. Combined impact of mutations in *SF3B1* and inv(3)/t(3;3) on gene expression and RNA splicing.** (A) Similarity matrix and hierarchical clustering of 4 groups (*Mx1*-Cre control, *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$) by differential gene expression. Three samples were independently collected in each group. (B) Principal component (PC) analysis of gene expression from 12 samples (4 groups, biologically triplicated). (C) Overlap of differentially expressed genes compared with *Mx1*-Cre control. (D) Significantly dysregulated pathways. *P* values are color-coded. (E) Significantly dysregulated pathways. Number of genes and statistical significance ($-\log_{10}$FDR) were shown. The impact of *SF3B1* mutation on gene expression was analyzed under the condition with or without *EVI1* rearrangement. (F) Overlap of differentially spliced genes compared with AML without *SF3B1* mutation or *EVI1* rearrangement. (G) Aberrant splicing detected in AML with *EVI1* rearrangement and *SF3B1* mutations. x-axis and y-axis indicate the percent spliced in (ψ) of each splicing event in the presence/absence of genetic alterations. Alternate splice sites, mutually exclusive exons, retained introns, or cassette exons are shown when $P < .01$. Red and blue dots represent individual splicing events or coding genes that are promoted or repressed in each condition; green dots are shown when the difference in percent spliced is <10%. The number of aberrantly spliced genes is indicated in blue or red.

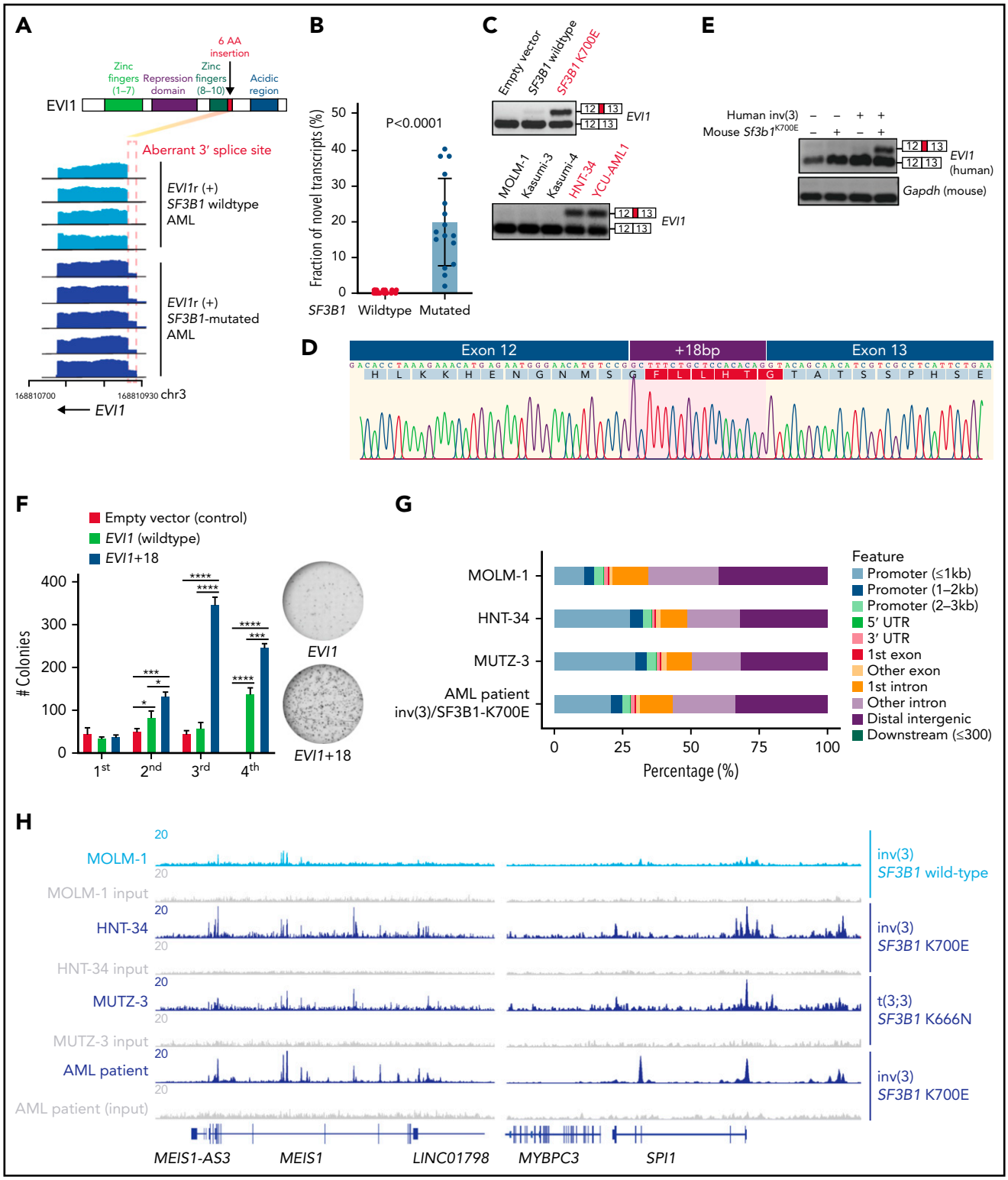



**Figure 4. *SF3B1* mutations promote expression of a novel *EVI1* isoform that enhances *EVI1*'s self-renewal capacity.** (A) Schematic of EVI1 protein with 6 amino acid insertion (top) and representative RNA-seq coverage plot of *SF3B1* WT and mutated inv(3) AML (bottom). (B) Fraction of the novel transcript (*EVI1*+18) compared with normal transcript in *SF3B1* WT and *SF3B1* mutated *EVI1*-rearranged AML. (C) RT-PCR illustrating the inclusion of intronic sequences in *SF3B1* K700E-transduced MEL270 cells (top, red) and endogenously *SF3B1* K700E harboring leukemia cells (bottom, red). (D) Sanger sequencing of complementary DNA (cDNA) arising from the top band in panel C. The nucleotide sequences and corresponding amino acids are indicated. (E) RT-PCR of human *EVI1* and mouse *Gapdh* using cDNA derived from peripheral blood of 4 murine models. (F) Number of myeloid colonies on first to fourth plating of c-Kit$^+$ BM cells transduced with empty vector (control), *EVI1* (WT), or *EVI1*+18 cDNA (left). Representative images (right) of the sixth colony. (G) Genomic distribution of anti-EVI1 ChIP-seq peaks. (H) Coverage tracks showing EVI1 ChIP-seq occupancy at the indicated genomic loci. *P* values were calculated by 2-sided Student *t* test. **P* < .05, ***P* < .01, ****P* < .001, and *****P* < .0001.

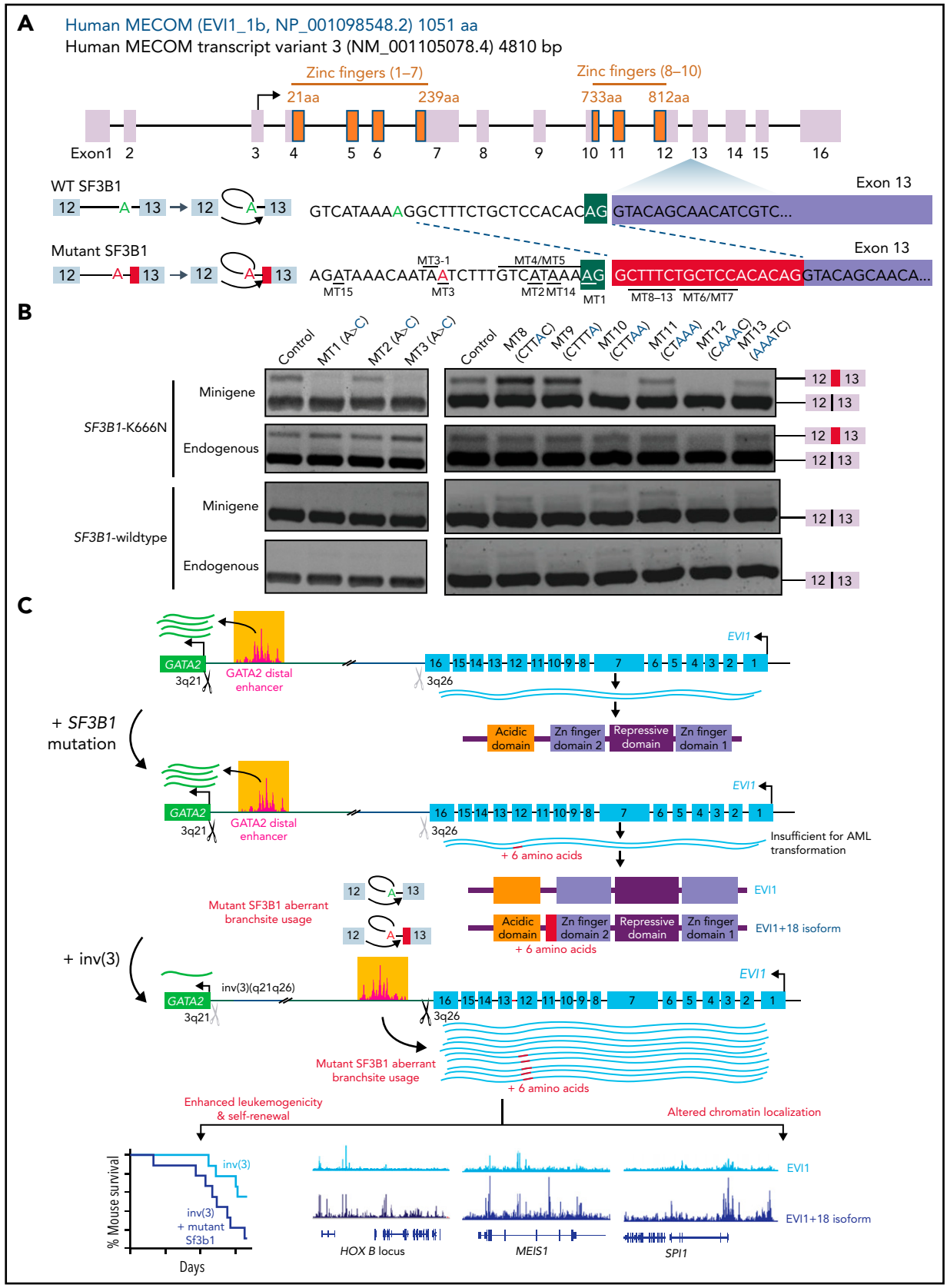



**Figure 5.**

isoform (this mutant minigene is noted as "MT1"). We then searched for potential cryptic branchpoints used by mutant SF3B1 by mutagenizing 2 adenines upstream of the cryptic 3′ss (MT2 and MT3 minigenes) and concluded that the adenine located at the −16 position upstream of the cryptic 3′ss was indispensable for *EVI1*+18 generation. We mapped branch sites by lariat-sequencing RNA derived from K562 cells with endogenous knock-in of the *SF3B1* K700E mutation (supplemental Figure 10A). We sequenced 6 single colonies to detect branchpoints within this intron and found that the adenosine nucleotides corresponding to the regions at MT1 (3/6), MT2 (1/6), and MT3 (2/6) are used (supplemental Figure 10B). However, it is hard to exclude the possibility that an adjacent adenosine is never used as the branch site because the branchpoint is occasionally deleted or mutated during PCR. Indeed, we found that mutagenesis of the adenosine nucleotide immediately adjacent to MT3 (MT3-1 in Figure 5A, A>C) erased the +18 isoform (supplemental Figure 10C). We next investigated the role of the polypyrimidine tract in generation of this mutant form of *EVI1*. Although a polypyrimidine tract is not consistently observed upstream of an aberrant cryptic 3′ss, the forced introduction of a polypyrimidine tract upstream of the cryptic 3′ss or disruption of the canonical 3′ss's polypyrimidine tract enhanced expression of the *EVI1*+18 by mutant SF3B1 (supplemental Figure 10D, MT4-7). Considering that altering the branchpoint affected generation of the *EVI1*+18 isoform more profoundly than altering the polypyrimidine tract, it is likely that SF3B1 mutants are less dependent on polypyrimidine tract sequences in inducing the aberrant 3′ss selection at this region of *EVI1*. This finding is similar to that seen with mutant SF3B1's of *BRD9* (supplemental Figure 10E).[37]

Next, we sought to identify ESEs necessary for the production of the aberrant *EVI1* transcript by mutant SF3B1. In our prior analyses of *BRD9* mis-splicing by mutant SF3B1, we identified a "TTTCT" sequence as a cryptic ESE within *BRD9* exclusively used by mutant SF3B1.[37] Of note, the *EVI1*+18 variant also contained a "TTTCT" sequence close downstream of the 3′ss. Mutagenizing either 2 or 3 bases around this same sequence within *EVI1* remarkably inhibited the production of *EVI1*+18 (Figure 5B, MT8-13). These results highlight the dependency of SF3B1 mutant-induced mis-splicing of *EVI1* on cryptic BPS selection and a specific ESE. Moreover, we confirmed that AG to AA mutation at the 3′ss completely ablated the normal 3′ss in *SF3B1*-WT condition and promoted usage of the "AG" located at −18 bp upstream as an aberrant 3′ss as observed in *SF3B1* mutated patients (supplemental Figure 10F-G). Finally, because there are annotated single nucleotide polymorphisms (SNPs) (c.2651-23A>G and c.2651-44A>G) near the 3′ss of intron 12 that may affect *EVI1* splicing, we evaluated the impact of these SNPs using our *EVI1* minigene (MT14 and MT15 in Figure 5A, respectively). This revealed that these SNPs modestly affected the splicing in the minigene reporter assay, suggesting that the SNPs in intron 12 may influence the generation of the *EVI1*+18 isoform (supplemental Figure 10G).

## Discussion

Although mutations in *SF3B1* are heavily enriched in MDS-RS[21] and there are accumulating data on the impact of *SF3B1* mutations and their consequent mis-splicing in the aberrant erythropoiesis of MDS-RS,[14,24,39,40] specific roles for *SF3B1* mutations in AML are not well explored. Here, we identify a strikingly high frequency of *SF3B1* mutations in adverse myeloid malignancies with inv(3)/t(3;3). Although the pathogenic consequences of how chromosome 3 rearrangements alter the enhancer landscape of *GATA2* and *MECOM* expression have been exquisitely dissected, here we make a novel observation that mutations in *SF3B1* induce aberrant splicing of *EVI1* itself and give rise to a novel oncogenic isoform of *EVI1*.

The splicing event within *EVI1* reported here is distinct from any prior *EVI1* or *MECOM* isoforms described previously. Although there have been extensive prior studies of global splicing alterations as well as individual mis-splicing events induced by mutations in *SF3B1* in a variety of cell and cancer types,[25,26,37,41-44] the current report is the first to report this novel unannotated splicing alteration within *MECOM/EVI1*. Most prior studies have focused on the impact of *SF3B1* mutations on aberrant intron proximal 3′ss usage resulting in out-of-frame transcripts predicted to result in nonsense-mediated messenger RNA (mRNA) decay.[25,26,37] However, this study as well as several important studies on the impact of SF3B1 mutant mis-splicing on aberrant exon inclusion[37] and intron removal[26,44] highlights a diversity of functionally important splicing changes induced by mutant SF3B1.

Although this mis-splicing event in *MECOM/EVI1* is present in *SF3B1* mutant human cells lacking inv(3)/t(3;3) rearrangements, we believe that detection of *EVI1* mis-splicing was facilitated in our studies by the uniquely elevated *EVI1* mRNA expression created by the structural alteration of chromosome 3 studied here (Figure 5C). Although it is clear that expression of this *EVI1*+18 isoform from the endogenous inv(3) allele promoted leukemogenesis in vivo compared with expression of known *EVI1* isoforms from the inv(3) allele, the contribution of the *EVI1*+18

**Figure 5. *Cis* elements within *EVI1* required for generation of the *EVI1*+18 bp splice variant by mutant SF3B1.** (A) *EVI1* gene structure and protein domains (top). Inset illustrates the transcripts when +18 nucleotides (red rectangle) are excluded (top) or included (bottom). Green A and red A indicate the branchpoint for canonical and aberrant transcripts, respectively. Single underlining indicates sequence motifs that were subsequently mutated in the minigene assay (each individual minigene construct is named "MT1" to "MT13"). aa, amino acid. (B) RT-PCR analysis of the +18 nucleotides inclusion in a minigene (top) or endogenous (bottom) context following transfection of minigenes with the illustrated mutations into *SF3B1*-K666N knocked-in K562 cells and *SF3B1*-WT K562 cells. (C) Schematic of the model proposed by which *EVI1* rearrangements and *SF3B1* mutations promote leukemia development. As previously demonstrated, structural rearrangements at chromosome 3q reposition the *GATA2* distal enhancer to upregulate *EVI1* expression while simultaneously downregulating *GATA2*. As shown in this study, approximately one-third of patients with *EVI1* rearrangements harbor concomitant change-of-function mutations in SF3B1, which promote use of an aberrant intron-proximal branch site within intron 12 of *EVI1*. This splicing alteration generates a stable unannotated transcript of EVI1 ("*EVI1*+18"), which is translated to express an EVI1 protein with insertion of 6 amino acids at the C-terminal end of the second ZF domain of EVI1. The EVI1+18 isoform is expressed whenever any recurrent cancer hotspot mutations in SF3B1 is present in cells with human EVI1 expression. Although EVI1+18 is not sufficient for leukemia transformation on its own, EVI1+18 enhances leukemogenicity in the setting of the EVI1 rearrangement and alters the chromatin localization of EVI1 to loci well known to be involved in leukemia development (such as *MEIS1* and the *HOXB* locus).

isoform in cells without *EVI1* rearrangement remains to be clarified. Although expression of the *EVI1*+18 isoform was seen in every inv(3)/t(3;3) rearranged patient with an *SF3B1* mutation, we witnessed varying levels in the magnitude of expression of this unique *EVI1* isoform.

Currently, it is unclear if the variation in the degree of *EVI1*+18 expression occurred due to differences in allelic frequency of *SF3B1* mutations, the exact SF3B1 mutant residue, potential germline SNPs within *EVI1* intron 12, and/or technical variability owing to leukemia cell purity. Use of emerging long-read RNA-seq technologies may be helpful in illuminating the exact full-length *EVI1* isoforms created in the setting of inv(3)/t(3;3) AML and may even be applied to *EVI1* isoform expression at the single-cell level.[45] It is also possible that structural rearrangements involving the *GATA2* distal enhancer with distinct breakpoints could result in varying levels of *EVI1* upregulation. This point will be important to study in future efforts focusing on cohorts of patients with *EVI1* rearrangements paired with bulk and single-cell genomic data. Moreover, it will be important for future clinical studies involving larger numbers of patients with *EVI1* rearrangements to dissect the clinical impact of coexisting *SF3B1* mutations and presence of distinct *EVI1* isoforms. Finally, given that *SF3B1* mutations are often seen in the setting of clonal hematopoiesis,[46,47] it will be interesting to determine the order of acquisition of *SF3B1* mutations and chromosome 3 structural rearrangements in future studies.

Combined expression of the *SF3B1* mutation with the human inv(3) allele in mice enhanced myeloid lineage skewing, HSPC expansion, and leukemia development, supporting the notion that mutant SF3B1 gives rise to an additional cancer program within inv(3)-inducing leukemia. Evaluation of shared mis-splicing events across human inv(3)/t(3;3) *SF3B1*-mutant AML and our murine models consistently identified this shared *EVI1* mis-splicing event. Importantly, mis-splicing of *EVI1* was only seen in mice when the human *EVI1* sequence was present in mutant SF3B1 mouse cells (as we believe differences in nucleotide sequence between mouse and human intron 12 precluded mis-splicing of mouse *Evi1*). Although our data do not eliminate the possibility that additional mutant SF3B1 mis-splicing events can contribute to inv(3)/t(3;3) leukemogenesis, ectopic expression of this *EVI1*+18 isoform enhanced self-renewal of HSPCs.

Mechanistically, the novel *EVI1*+18 appears to alter chromatin and/or DNA binding relative to WT *EVI1* (Figure 5C). Structural predictions suggest the possibility that the additional 6 amino acids generated by the *EVI1*+18 splicing event may alter the function of the second ZF domain of EVI1.[32] Interestingly, most monoallelic mutations within MECOM in patients with the *MECOM*-associated syndrome patients occur in the second ZF domain and many occur at this same splice site.[48] These observations suggest that mutations or mis-spliced forms of this ZF domain occur in multiple EVI1-associated diseases.

It is hoped future biochemical and structural studies of this *EVI1* isoform will clarify the biophysical impact of the *EVI1*+18 on *EVI1* function in more detail. For example, it is possible that this mis-splicing event or mutations in this domain of EVI1 may disrupt the secondary structure of the second Zn finger domain and/or interfere with zinc coordination. Overall, given the dismal outcome in patients with inv(3)/t(3;3) MDS and AML, the models developed here will be an important resource for future therapeutic and mechanistic studies, and this subtype of myeloid malignancy is in need of better outcomes.

Although our data suggested that spliceosomal inhibitors exert therapeutic effects on *EVI1*-r AML with *SF3B1* mutations, there is debate on whether leukemia-associated mutations in RNA splicing factors drive disease development and/or maintenance owing to mis-splicing of key mRNAs and/or via impacts on cellular processes distinct from RNA splicing.[38,49] It will therefore be very exciting to use the SF3B1 mutant models here to study the impact of correcting the individual mis-splicing event in *EVI1* as well as more global alterations in RNA splicing created by mutant SF3B1 using these models.

## Acknowledgments

Inv(3) mouse strain (RBRC09508, 3q21q26 mouse line B) was kindly provided by Masayuki Yamamoto (Tohoku University) and RIKEN BRC through the National BioResource Project of the MEXT/AMED, Japan.

This work was supported by American Society of Hematology (S.C.L. and D.I.), Japanese Society of Hematology (D.I.), Leukemia & Lymphoma Society (D.I. and O.A.-W.), National Institutes of Health (NIH), National Cancer Institute grants R01 CA242020 (O.A.-W.), R01 CA251138 (O.A.-W.), and R00 CA218896 (S.C.L.), National Heart, Lung, and Blood Institute grant R01HL128239 (O.A.-W.), P50 254838 (O.A.-W.), the Edward P. Evans Foundation (O.A.-W. and S.C.L.), and the Vera and Joseph Dresner Foundation (S.C.L.). D.I. is supported by The Naito Foundation, The Uehara Memorial Foundation, Senri Life Science Foundation, The Sumitomo Foundation, The Kanae Foundation for the Promotion of Medical Science, The Mitsubishi Foundation, KAKETSUKEN, Bristol Myers Squibb Foundation, JSPS KAKENHI (JP20H00537 and JP20H03717) and AMED (21ck0106697h0001). A.T. is supported by JSPS KAKENHI (JP21J15620, 16H06279 (PAGS)).

## Authorship

Contribution: A.T., D.I., and O.A.-W. designed the study; H. Kawamoto supervised retroviral experiments; J.P.B. and T.A.N. provided clinical data; D.I., W.Z., A.T., M.N., R.M.-L., A.P., and R.D. performed computational analyses of mutational/RNA-seq data; M.N., S.H., and A.P. performed computational analyses of RNA-seq data; A.Y. and M.H. performed ChIP-seq; D.I., W.Z., A.K., T.I., and M.N. performed computational analyses of ChIP-seq data; H. Kunimoto and H.N. provided YCU-AML1 cells; A.T. and B.L. performed minigene splicing assay; T.K. performed protein structural analysis; A.T., H.Y., S.H., M.K., Y.H., H.C., S.C.L., M.X., Y.K., Y.Z., W.Z., M.F., Y.A., H.H., S.C.L., and D.I. performed animal experiments; A.T., D.I., M.N., A.P., and O.A.-W. wrote the manuscript with approval from all coauthors.

Conflict-of-interest disclosure: O.A.-W. has served as a consultant for H3B Biomedicine, Foundation Medicine Inc, Merck, Janssen, and Loxo Oncology/Lilly, is on the Scientific Advisory Board of Envisagenics Inc and Harmonic Discovery Inc, and has received prior research funding from H3B Biomedicine, Loxo Oncology/Lilly, and Nurix Therapeutics unrelated to the current manuscript. D.I. has received prior research funding from Abbvie and Sumitomo Dainippon Pharma unrelated to the current manuscript. The remaining authors declare no competing financial interests.

ORCID profiles: A.T., 0000-0002-9449-4894; R.M.-L., 0000-0003-0298-7948; A.P., 0000-0003-2429-2370; A.Y., 0000-0002-5639-8068; T.K., 0000-0002-4412-2849; T.I., 0000-0002-3455-6010; D.I., 0000-0002-7947-6540.

Correspondence: Omar Abdel-Wahab, Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10045; e-mail: abdelwao@mskcc.org; Daichi Inoue, Department of Hematology-Oncology, Institute of Biomedical Research and Innovation, Foundation for Biomedical Research and Innovation at Kobe, Kobe, Japan 6500047; e-mail: d-inoue@fbri.org.

## Footnotes

Submitted 28 December 2021; accepted 6 June 2022; prepublished online on *Blood* First Edition 16 June 2022. DOI 10.1182/blood.2021015325.

*O.A.-W. and D.I. contributed equally to this study.

The data reported in this article have been deposited in the Gene Expression Omnibus database (accession number GSE49642; mouse model: accession GSE190655, human cell line: GSE190653, mouse colony with retroviral EVI1 expression: GSE202208).

RNA-seq reads for the human samples reported in Leucegene were downloaded from the Gene Expression Omnibus (accession number GSE49642). RNA-seq data generated by the Beat AML data (dbGaP accession phs001657.v1.p1) were downloaded from the National Cancer Institute Genomic Data Commons. RNA-seq data generated in this study have been deposited in the Gene Expression Omnibus (mouse model: accession GSE190655, human cell line: GSE190653, mouse colony with retroviral EVI1 expression: GSE202208), and the human ChIP-seq data are deposited in GSE190652, GSE202207, and available from the investigators upon request.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.



## REFERENCES

1. Sun J, Konoplev SN, Wang X, et al. De novo acute myeloid leukemia with inv(3)(q21q26.2) or t(3;3)(q21;q26.2): a clinicopathologic and cytogenetic study of an entity recently added to the WHO classification. *Mod Pathol.* 2011;24(3):384-389.

2. Lugthart S, Gröschel S, Beverloo HB, et al. Clinical, molecular, and prognostic significance of WHO type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and various other 3q abnormalities in acute myeloid leukemia. *J Clin Oncol.* 2010;28(24):3890-3898.

3. Cui W, Sun J, Cotta CV, Medeiros LJ, Lin P. Myelodysplastic syndrome with inv(3)(q21q26.2) or t(3;3)(q21;q26.2) has a high risk for progression to acute myeloid leukemia. *Am J Clin Pathol.* 2011;136(2):282-288.

4. Rogers HJ, Vardiman JW, Anastasi J, et al. Complex or monosomal karyotype and not blast percentage is associated with poor survival in acute myeloid leukemia and myelodysplastic syndrome patients with inv(3)(q21q26.2)/t(3;3)(q21;q26.2): a Bone Marrow Pathology Group study. *Haematologica.* 2014;99(5):821-829.

5. Gröschel S, Lugthart S, Schlenk RF, et al. High EVI1 expression predicts outcome in younger adult patients with acute myeloid leukemia and is associated with distinct cytogenetic abnormalities. *J Clin Oncol.* 2010;28(12):2101-2107.

6. Summerer I, Haferlach C, Meggendorfer M, Kern W, Haferlach T, Stengel A. Prognosis of *MECOM* (*EVI1*)-rearranged MDS and AML patients rather depends on accompanying molecular mutations than on blast count. *Leuk Lymphoma.* 2020;61(7):1756-1759.

7. Yamazaki H, Suzuki M, Otsuki A, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell.* 2014;25(4):415-427.

8. Gröschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell.* 2014;157(2):369-381.

9. Haferlach C, Bacher U, Haferlach T, et al. The inv(3)(q21q26)/t(3;3)(q21;q26) is frequently accompanied by alterations of the RUNX1, KRAS and NRAS and NF1 genes and mediates adverse prognosis both in MDS and in AML: a study in 39 cases of MDS or AML. *Leukemia.* 2011;25(5):874-877.

10. Lavallée VP, Gendron P, Lemieux S, D'Angelo G, Hébert J, Sauvageau G. EVI1-rearranged acute myeloid leukemias are characterized by distinct molecular alterations. *Blood.* 2015;125(1):140-143.

11. Gröschel S, Sanders MA, Hoogenboezem R, et al. Mutational spectrum of myeloid malignancies with inv(3)/t(3;3) reveals a predominant involvement of RAS/RTK signaling pathways. *Blood.* 2015;125(1):133-139.

12. Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn.* 2015;17(3):251-264.

13. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients [published correction appears in *Nat Med.* 2017;23(8):1004]. *Nat Med.* 2017;23(6):703-713.

14. Obeng EA, Chappell RJ, Seiler M, et al. Physiologic expression of Sf3b1(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. *Cancer Cell.* 2016;30(3):404-417.

15. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature.* 2018;562(7728):526-531.

16. Inoue D, Bradley RK, Abdel-Wahab O. Spliceosomal gene mutations in myelodysplasia: molecular links to clonal abnormalities of hematopoiesis. *Genes Dev.* 2016;30(9):989-1001.

17. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011;478(7367):64-69.

18. Ley TJ, Miller C, Ding L, et al; Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059-2074.

19. Papaemmanuil E, Cazzola M, Boultwood J, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med.* 2011;365(15):1384-1395.

20. Malcovati L, Karimi M, Papaemmanuil E, et al. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood.* 2015;126(2):233-241.

21. Malcovati L, Stevenson K, Papaemmanuil E, et al. SF3B1-mutant MDS as a distinct disease subtype: a proposal from the International Working Group for the Prognosis of MDS [published correction appears in *Blood.* 2021;137(21):3003]. *Blood.* 2020;136(2):157-170.

22. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med.* 2016;374(23):2209-2221.

23. Ochi Y, Kon A, Sakata T, et al. Combined cohesin-RUNX1 deficiency synergistically perturbs chromatin looping and causes myelodysplastic syndromes. *Cancer Discov.* 2020;10(6):836-853.

24. Mupo A, Seiler M, Sathiaseelan V, et al. Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia.* 2017;31(3):720-727.

25. Darman RB, Seiler M, Agrawal AA, et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3′ splice site selection through use of a different branch point. *Cell Rep.* 2015;13(5):1033-1045.

26. Shiozawa Y, Malcovati L, Gallì A, et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat Commun.* 2018;9(1):3649.

27. Obeng EA, Stewart C, Abdel-Wahab O. Altered RNA processing in cancer pathogenesis and therapy. *Cancer Discov*. 2019;9(11):1493-1510.

28. Inoue D, Abdel-Wahab O. Modeling SF3B1 mutations in cancer: advances, challenges, and opportunities. *Cancer Cell*. 2016;30(3): 371-373.

29. Birdwell C, Fiskus W, Kadia TM, DiNardo CD, Mill CP, Bhalla KN. EVI1 dysregulation: impact on biology and therapy of myeloid malignancies. *Blood Cancer J*. 2021; 11(3):64.

30. Kunimoto H, Fukuchi Y, Murakami K, et al. Establishment of a high-risk MDS/AML cell line YCU-AML1 and its xenograft model harboring t(3;3) and monosomy 7. *HemaSphere*. 2020;4(5):e469.

31. Hamaguchi H, Suzukawa K, Nagata K, Yamamoto K, Yagasaki F, Morishita K. Establishment of a novel human myeloid leukaemia cell line (HNT-34) with t(3;3)(q21;q26), t(9;22)(q34;q11) and the expression of EVI1 gene, P210 and P190 BCR/ABL chimaeric transcripts from a patient with AML after MDS with 3q21q26 syndrome. *Br J Haematol*. 1997;98(2): 399-407.

32. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589.

33. Kataoka K, Sato T, Yoshimi A, et al. Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. *J Exp Med*. 2011; 208(12):2403-2416.

34. Uehara T, Minoshima Y, Sagane K, et al. Selective degradation of splicing factor CAPERα by anticancer sulfonamides. *Nat Chem Biol*. 2017;13(6):675-680.

35. Ottema S, Mulet-Lazaro R, Erpelinck-Verschueren C, et al. The leukemic oncogene EVI1 hijacks a MYC super-enhancer by CTCF-facilitated loops. *Nat Commun*. 2021;12(1):5679.

36. Argiropoulos B, Yung E, Humphries RK. Unraveling the crucial roles of Meis1 in leukemogenesis and normal hematopoiesis. *Genes Dev*. 2007;21(22):2845-2849.

37. Inoue D, Chew GL, Liu B, et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature*. 2019;574(7778): 432-436.

38. Inoue D, Polaski JT, Taylor J, et al. Minor intron retention drives clonal hematopoietic disorders and diverse cancer predisposition. *Nat Genet*. 2021;53(5):707-718.

39. Dolatshad H, Pellagatti A, Fernandez-Mercado M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia*. 2015; 29(5):1092-1103.

40. Jin S, Su H, Tran NT, et al. Splicing factor SF3B1K700E mutant dysregulates erythroid differentiation via aberrant alternative splicing of transcription factor TAL1. *PLoS One*. 2017;12(5):e0175523.

41. Furney SJ, Pedersen M, Gentien D, et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov*. 2013;3(10):1122-1129.

42. Alsafadi S, Houy A, Battistella A, et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun*. 2016;7(1): 10615.

43. Wang L, Brooks AN, Fan J, et al. Transcriptomic characterization of SF3B1 mutation reveals its pleiotropic effects in chronic lymphocytic leukemia. *Cancer Cell*. 2016;30(5):750-763.

44. Tang AD, Soulette CM, van Baren MJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun*. 2020;11(1):1438.

45. Gupta I, Collier PG, Haase B, et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat Biotechnol*. 2018;36(12):1197-1202.

46. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015; 126(1):9-16.

47. McKerrell T, Park N, Moreno T, et al; Understanding Society Scientific Group. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep*. 2015;10(8): 1239-1245.

48. Germeshausen M, Ancliff P, Estrada J, et al. MECOM-associated syndrome: a heterogeneous inherited bone marrow failure syndrome with amegakaryocytic thrombocytopenia. *Blood Adv*. 2018;2(6): 586-596.

49. Chen L, Chen JY, Huang YJ, et al. The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol Cell*. 2018;69(3):412-425.e6.

## Supplementary Methods

### Bone marrow (BM) transplantation

Freshly dissected femora and tibiae were isolated from *Mx1*-Cre control, *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ CD45.2$^{+}$ mice. The BM was spun at 0.5 g by centrifugation and red blood cells (RBCs) were lysed in ammonium chloride-potassium bicarbonate lysis buffer for 5 min. After centrifugation, cells were resuspended in PBS plus 3% fetal bovine serum (FBS), passed through a cell strainer, and counted. Finally, 1.0 x $10^6$ total BM cells of *Mx1*-Cre control, *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ CD45.2$^{+}$ mice were mixed with 1.0 x $10^6$ WT CD45.1$^{+}$ support BM and transplanted via tail vein injection into lethally irradiated (4.5Gy, twice) CD45.1$^{+}$ recipient mice. Chimerism was measured by flow cytometry from the peripheral blood at 4 weeks after transplant (week 0, pre-pIpC [polyinosinic-polycytidylic acid] (p1530, Sigma-Aldrich)). After confirming the engraftment, the recipient mice were treated with pIpC (dissolved with PBS, intraperitoneal injection, 500 μg/body, every other day, three times). Chimerism was followed via flow cytometry from the peripheral blood every 4 weeks. Additionally, for each bleeding, whole blood cell counts were measured on a blood analyzer, and peripheral blood smears were scored. For the serial transplant, we collected whole BM samples eight months after the initial transplant and transplanted the viably frozen whole BM cells into sublethally irradiated (4.5Gy, once) wild-type recipients (1 x $10^6$ cells per recipient). For noncompetitive transplantation experiments, 2.0 x $10^6$ total BM cells of *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ CD45.2$^{+}$ mice were injected into lethally irradiated (4.5Gy, twice) CD45.1$^{+}$ recipient mice. Five months after pIpC injection, whole BM cells (2.0 x $10^6$ cells/recipient) were serially transplanted into lethally irradiated CD45.1$^{+}$ recipient mice.

### Retroviral Infection

Retroviral vector pMYs-IRES-GFP encoding EVI1 (wild-type and +18) and SF3B1 (wild-type, K700E, K666N, and G740E) and, if necessary, pVSV-G vector (631530, TakaraBio) were transfected into Plat-E or GP2-293 (631530, Takara) cells using Xtremegene 9 (6365809001, Sigma-Aldrich). About 48 hours after transfection, the culture supernatant was filtered through a 0.22 μm filter unit (SLGVR33RS, Merck) and either used for infection or frozen for stock. For the infection into primary cells, the prestimulated cells were infected for 60 hours using 6-well dishes coated with RetroNectin (T100A, Takara Bio). For the infection into cell lines, cells were

incubated in complete media with 20% viral supernatant and 2.5 μg/ml polybrene (12996-81, nacalai tesque). The media was replaced with complete media 8-24 hours after infection.

### *In vitro* colony-forming assays

Whole BM cells from *Mx1*-Cre control, *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ mice and seeded at a density of 20,000, 100,000, 200,000 cells/replicate into cytokine-supplemented methylcellulose medium (Methocult M3434, M3436, and M3630, respectively; STEMCELL Technologies). Colonies propagated in culture were scored at day 7-10 and M3434 colonies were replated into new M3434 semisolid media. For cDNA expression experiment, c-Kit$^+$ cells were selected by anti-mouse CD117 MicroBeads (Miltenyi Biotech) from 14-week-old primary CD45.1$^+$ mice, and were cultured overnight in Iscove's Modified Dulbecco's Medium (IMDM, I3390, Sigma-Aldrich) and10% FBS medium supplemented with 50 ng/ml recombinant murine SCF (250-03; PeproTech), 20 ng/ml recombinant murine IL-3 (213-13; PeproTech), 20 ng/ml recombinant murine IL-6 (216-16; PeproTech), and 20 ng/ml recombinant murine TPO (315-14; Peprotech). The next day, those cells were infected with retroviral supernatent expressing pMYs-IRES-GFP empty vector, N-terminal 3X HA tagged full-length *EVI1* (NM_001105078.4), and *EVI1* +18 variant cDNA. Three days after infection, 2.0 x $10^4$ of GFP$^+$ cells were FACS-sorted and plated in cytokine-supplemented methylcellulose medium (Methocult M3434; STEMCELL Technologies) in triplicate. For the *in vitro* competitive assay, 1.0 x $10^4$ each of GFP$^+$ and tdTomato$^+$ cells were FACS-sorted and plated in M3434 semisolid media, followed by replating every 10 days.

### Antibodies, FACS, and Western blot analysis

All FACS antibodies were purchased from BD Biosciences, eBioscience, or BioLegend. BM mononuclear cells were stained with a lineage cocktail comprised of antibodies targeting CD3, CD4, CD8, B220, CD19, NK1.1, Gr-1, CD11b, Ter119, and IL-7Rα. Cells were also stained with antibodies against c-Kit, Sca1, CD150, and CD48. Cell populations were analyzed using an LSR Fortessa (BD Biosciences) and a FACSLyric (BD Biosciences) and sorted with a FACSAria II instrument (BD Biosciences). We used the following antibodies: B220-APC-Cy7 (clone; RA3-6B2; BioLegend; catalog #: 103224; dilution: 1:200); B220-PerCP-Cy5.5 (RA3-6B2; eBioscience; 45-0452-82; 1:200); CD3-PE-Cy7 (17A2; BioLegend; 100220; 1:200); CD3-APC-Cy7 (17A2; BioLegend; 100222; 1:200); Gr1-APC (RB6-8C5; eBioscience; 25-5931-82; 1:500); CD11b-FITC (M1/70; Biolegend; 101206; 1:200); CD11b-APC-Cy7 (M1/70; BioLegend; 101226; 1:200); NK1.1-APC-Cy7 (PK136; BioLegend; 108724; 1:200); Ter119-APC-Cy7 (Ter119,

BioLegend; 116223: 1:200); c-Kit-APC (2B8; BioLegend; 105812; 1:100); c-Kit-PerCP-Cy5.5 (2B8; BioLegend; 105824; 1:100); c-Kit-BV605 (ACK2; BioLegend; 135120; 1:100); Sca1-PE-Cy7 (D7; BioLegend; 108102; 1:100); CD45.1-FITC (A20; BioLegend; 110706; 1:200); CD45.1-PerCP-Cy5.5 (A20; BioLegend; 110728; 1:200); CD45.1-BV711 (A20; BioLegend; 110739; 1:200); CD45.1-APC (A20; BioLegend; 110714; 1:200); CD45.2-PE (104; eBioscience; 12-0454-82; 1:200); CD45.2-Alexa700 (104; BioLegend; 109822; 1:200); CD45.2-BV605 (104; BioLegend;109841; 1:200); CD48-PerCP-Cy5.5 (HM48-1; BioLegend; 103422; 1:100); CD150-PE (9D1; eBioscience; 12-1501-82; 1:100); CD135-APC (A2F10; BioLegend; 135310; 1:200); CD16/CD32 (FcγRII/III)-Alexa700 (93; eBioscience; 56-0161-82; 1:100); CD34-FITC (RAM34; BD Biosciences; 553731; 1:50); CD34-PerCP (8G12; BD Biosciences; 345803; 1:50); CD117-PE-Cy7 (104D2; eBioscience; 25-1178-42; 1:100); CD45-APC-H7 (2D1; BD Biosciences; 560178; 1:200); Ly-5.1-PE (BP-1; BD Bioscience; 553735; 1:200). For Western blot analysis, lysate of cultured cells with RIPA buffer (150 mM NaCl, 50 mM Tris, pH 8.0, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) with Halt Protease and Phosphatase Inhibitor Cocktail (78446, Thermo Fisher Scientific) was mixed with Pierce Lane Marker Reducing Sample Buffer (39000, Thermo Fisher Scientific) and denatured by boiling for 5 minutes. The mixture was loaded onto 4–12% Bis-Tris NuPAGE Gels (Thermo Fisher Scientific) followed by wet transfer. The following antibodies were used for Western Blot analysis: HA (#3724, Cell Signaling Technology; 1:1,000), Actin (A-5441, Sigma-Aldrich, 1:5,000), rabbit IgG-HRP (#7074, Cell Signaling Technology, 1:10,000), and mouse IgG-HRP (#7076, Cell Signaling Technology; 1:10,000).

**Histological analysis**

Mice were sacrificed and autopsied, and the dissected tissue samples were fixed for 24 hours in 4% paraformaldehyde, dehydrated, and embedded in paraffin. Paraffin blocks were sectioned at 4 mm and stained with H&E. Images were acquired using an Axio Observer A1 microscope (Carl Zeiss).

**Peripheral blood analysis**

Blood was collected by retro-orbital bleeding using heparinized microhematocrit capillary tubes (02-668-25, Thermo Fisher Scientific). Automated peripheral blood counts were obtained using a HemaVet 950 (Drew Scientific) according to standard manufacturer's instruction. Differential blood counts were realized on blood smears stained using Wright-Giemsa staining and visualized using an Axio Observer A1 microscope.

**Indisulam treatment and the IC50 measurements**

For *in vitro* experiments, indisulam (22759-5, Cayman Chemical) was dissolved in DMSO to make a 100 micromolar stock solution, and this was then added to complete culture media to the appropriate final concentration. Cell lines were plated in 96 well plates and incubated the media including indisulam at concentrations ranging from 0.1 to 100 micromolar with a minimum of three technical replicates per concentration per cell line. Cell viability was measured with the CellTiter Glo reagent (G7570, Promega) as per manufacturer's instructions. Absolute viability values were converted to percentage viability versus DMSO control treatment, and then non-linear fit of log (inhibitor) versus response (three parameters) was performed in GraphPad Prism v7.0 to obtain an IC50 values.

**mRNA isolation and analysis**

For patient samples, cells were resuspended in TRIzol (Thermo Fisher) and chloroform. RNA was then extracted using RNeasy mini spin columns (Qiagen) per manufacturer's instructions. The concentration of extracted RNA was determined via NanoDrop (Thermo Fisher) and the quality was assessed by TapeStation analysis (Agilent). Poly(A)-selected Illumina libraries were generated using the TruSeq RNA Library Kit v2 following the manufacturer's protocol with 100 ng of input RNA. Libraries were amplified by PCR (13 cycles) and samples were submitted for a second round of TapeStation analysis to determine the purity and abundance of the expected ~300 bp PCR amplicon. Samples that met these criteria were subjected to paired-end sequencing on the Illumina HiSeq at a depth of ~50M reads per sample.

For sorted mouse cell populations (live, lineage-negative c-Kit$^+$ cells), RNA was extracted using RNeasy columns (Qiagen) per manufacturer's instructions. RNA was then Poly(A)-selected, and unstranded Illumina libraries were prepared with the standard TruSeq protocol. To select for fragments <400 bp, 0.5× AMPure XP beads were added to the library followed by 1× AMPure XP beads to select for fragments >100 bp. These fragments were then amplified by PCR (15 cycles) and separated by gel electrophoresis (2% agarose). 300 bp DNA fragments were isolated and sequenced on the Illumina HiSeq 2000 at a depth of ~100M 2×49 bp reads per sample. For differential expression analysis, sequenced reads were mapped to mm10 using nf-core/rnaseq v3.0 pipeline[1] with star_rsem aligner. RSEM gene counts were normalized and differentially expressed genes were identified with adjusted p-values < 0.1 by DESeq2 v1.26.0,[2] where independent hypothesis weighting was applied.[3] Variance stabilizing transformed gene counts were used for hierarchical clustering and principal component

analysis. Hierarchical clustering was done using Euclidean distance and Complete linkage method. With differentially expressed gene sets, pathway analysis was performed by Enrichr.[4-6] In parallel, RNA-seq reads were mapped to the transcriptome annotations assembled as described above with RSEM v1.2.4[7], modified to invoke Bowtie v1.0.0[8] with the '-v 2' option. Remaining unaligned reads were then mapped to the genome as well as a database of possible splice junctions, consisting of all possible combinations of 5' and 3' splice sites annotated for each gene, using TopHat v2.0.8b.[9] The resulting read alignments generated by TopHat were then merged with the output from RSEM. Human patient samples were accessed via the NCBI Sequence Read Archive. Reads were aligned to Gencode annotation v25 for human using STAR v2.4.1d and quantified by QoRTs v1.1.8,[10] while alignment-free quantification was performed by Kallisto v0.43.0[11] accounting for hexamer bias and using 100 bootstrap iterations to estimate the uncertainty due to the finite depth of coverage. Differentially expressed genes were then identified from alignment-free data using Sleuth v0.28.1.[12] Splicing defects at known loci (exons etc.) were identified using SUPPA v1.[13]

**RT–PCR and quantitative RT–PCR**

For cDNA synthesis, total RNA was reverse transcribed to cDNA with the Verso cDNA Synthesis kit (AB-1453/B; Thermo Scientific). The resulting cDNA was diluted 10–20-fold before use. *EVI1* splice variants were detected via semiquantitative RT–PCR in the condition listed in **Supplementary Table 1**. Quantitative RT–PCR (RT–qPCR) was performed in 10-µl reactions with SYBR Green PCR Master Mix (Roche Life Science). All RT–qPCR analyses were performed on CFX Connect™ Real-Time PCR Detection System (BioRad). Relative gene expression levels were calculated using the comparative CT method, and the values were corrected with expression levels of the internal controls Gapdh. Primers used for RT–PCR are listed in **Supplementary Table 1**.

**Chromatin immunoprecipitation (ChIP) and ChIP-seq**

Chromatin fractions from HEK293T cells were prepared using the fanChIP method, as described previously.[14,15] Cells were suspended in CSK buffer (100 mM NaCl, 10 mM PIPES [pH 6.8], 3 mM $MgCl_2$, 1 mM EGTA, 0.3 M sucrose, 0.5% Triton X-100, 5 mM sodium butyrate, 0.5 mM DTT, and protease inhibitor cocktail) and centrifuged (400× g for 5 min, at 4°C) to remove the soluble fraction. The pellet was resuspended in MNase buffer (50 mM Tris-HCl [pH 7.5], 4 mM $MgCl_2$, 1 mM $CaCl_2$, 0.3 M sucrose, 5 mM sodium butyrate, 0.5 mM DTT, and protease inhibitor cocktail) and treated with MNase at 37°C for 3–6 min to obtain

oligonucleosomes. MNase reaction was then stopped by adding EDTA (pH 8.0) to a final concentration of 20 mM. An equal amount of lysis buffer (250 mM NaCl, 20 mM sodium phosphate [pH 7.0], 30 mM sodium pyrophosphate, 5 mM EDTA, 10 mM NaF, 0.1% NP-40, 10% glycerol, 1 mM DTT, and EDTA-free protease inhibitor cocktail) was added to increase solubility. The chromatin fraction was cleared by centrifugation (15,000 rpm for 5 min, 4°C) and subjected to immunoprecipitation with anti-HA antibody (3F10, Roche) and Protein-G magnetic microbeads (Invitrogen). Immunoprecipitates were then washed five times with washing buffer (1:1 mixture of lysis buffer and MNase buffer with 20 mM EDTA) and eluted in elution buffer (1% SDS and 50 mM $NaHCO_3$). The eluted DNA material was fragmented by DNA shearing system (M220 Covaris) and analyzed by deep sequencing, which was performed using a TruSeq ChIP Sample Prep Kit (illumina) and HiSeq2500 (illumina) at the core facility of Hiroshima University. Approximately 29 to 33 million single end reads were obtained and trimmed using cutadapt v1.2.1. 27 to 31 million of reads were mapped to the hg19 reference genome with BWA v0.7.5 and wer subjected to further. Peaks were called and differential peaks were detected using bdgdiff module in MACS2 software v2.0.10 with log likelihood ratio = 1. Enriched motifs in differential peaks were found using findMotifsGenome module with mask option and peaks were annotated using annotatePeaks module in HOMER software v4.11.[16] Percentages of annotated peaks were shown as Pie chart. Bigwig files were made from bam files by bamCoverage in deepTools software v3.5.1,[17] where counts in chrX were ignored and normalized in CPM. Bigwig files were converted to matrix files around the center of the peaks extended to 10 kb up-and-down stream regions by reference-point mode of computeMatrix software. ChIP peaks were then visualized as heatmap.

For ChIP-seq with anti-EVI1 antibody (2593; Cell signalling), we followed the protocol previously described.[18] Cells were cross-linked with 1% formaldehyde. Chromatin was isolated using lysis buffer A (50mM Tris pH 8, 10mM EDTA, 1% SDS). At least 30 million cells were double crosslinked with 2mM disuccinimidyl glutarate followed by 1% formaldehyde. Chromatin of double crosslinked cells was isolated using lysis buffer B (10mM Tris pH 7.5, 74mM NaCl, 3mM MgCl2, 1mM CaCl2, 4% NP40, 0.32% SDS). The chromatin was sonicated with a Bioruptor device (Diagenode) using the following settings: 10 cycles of 30 s on, 30 s off. Immunoprecipitation of cross-linked chromatin was performed with antibodies against EVI1. Chromatin bound antibody was precipitated with prot G Dynabeads (Thermo Fisher Scientific) and washed with low salt buffer (20mM Tris pH 8, 2 mM EDTA, 1% Triton, 150mM NaCl), high salt buffer (20mM Tris pH 8, 2 mM EDTA, 1% Triton, 500mM NaCl), LiCl buffer (10mM Tris, 1mM EDTA, 0.25mM LiCl, 0.5% IGEPAL, 0.5% Sodium-Deoxycholate) and TE (10mM Tris pH

8, 1mM EDTA). The ChIP chromatin was eluted in elution buffer (0.1M Sodiumhydrogencarbonate, 1% SDS). Crosslinks were reversed overnight at 65 °C in the presence of proteinase K (New England Biolabs). De-crosslinked material was purified using a QIAGEN PCR Purification Kit. The purified DNA was processed according to the Nextflex ChIP Sample Preparation Protocol (Perkin Elmer) or the Microplex library preparation kit V2 (Diagnode C05010013) and sequenced on the Illumina NovaSeq6000 platform.

### *EVI1* minigene construction

The *EVI1* minigene construct was generated by inserting the DNA fragment containing the human *EVI1* genomic sequence from exon 12 to exon 13 between the BamHI and XhoI restriction sites of pcDNA3.1(+) vector. The sequences of the inserted fragments were verified by sanger sequencing. Mutagenesis of minigene constructs was performed with the Agilent QuikChange II site-directed mutagenesis kits (Agilent) or Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's directions. Primers used in mutagenesis are listed in **Supplementary Table 1.**

### Lariat sequencing

Branchpoints of the alternative spliced intron of *MECOM*/*EVI1* were mapped by lariat RT-PCR amplifying branchpoint-spanning fragments from lariat RNAs arising from the splicing of this intron (intron 12 of **Figure 5**). Briefly, the RT reaction was performed with SuperScript IV first strain synthesis kit (Invitrogen) and a primer complementary to the intronic sequences downstream of the 5′ splice sites to generate cDNA from lariat RNAs. The branchpoint spanning fragments were amplified by nested PCR with pairs of outer primers (using RT primer as the reverse primer) and inner primers (listed in **Supplementary Table 1**), cloned into the pGEM-T vector (Promega) and sequenced by Sanger sequencing.

## References


1. Ewels PA, Peltzer A, Fillinger S, et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol.* 2020;38(3):276-278.
2. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
3. Ignatiadis N, Klaus B, Zaugg JB, Huber W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods.* 2016;13(7):577-580.
4. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* 2013;14:128.

5. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90-97.
6. Xie Z, Bailey A, Kuleshov MV, et al. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc.* 2021;1(3):e90.
7. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics.* 2011;12:323.
8. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology.* 2009;10(3):R25.
9. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105-1111.
10. Hartley SW, Mullikin JC. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics.* 2015;16:224.
11. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525-527.
12. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods.* 2017;14(7):687-690.
13. Trincado JL, Entizne JC, Hysenaj G, et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* 2018;19(1):40.
14. Miyamoto R, Yokoyama A. Protocol for fractionation-assisted native ChIP (fanChIP) to capture protein-protein/DNA interactions on chromatin. *STAR Protoc.* 2021;2(2):100404.
15. Miyamoto R, Okuda H, Kanai A, et al. Activation of CpG-Rich Promoters Mediated by MLL Drives MOZ-Rearranged Leukemia. *Cell Rep.* 2020;32(13):108200.
16. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576-589.
17. Ramirez F, Ryan DP, Gruning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160-165.
18. Ottema S, Mulet-Lazaro R, Erpelinck-Verschueren C, et al. The leukemic oncogene EVI1 hijacks a MYC super-enhancer by CTCF-facilitated loops. *Nat Commun.* 2021;12(1):5679.

**Supplementary Table 1**. Primer sequences and PCR conditions for mice genotyping, RT-PCR, and minigene assays.

| **Genotyping** | | **PCR protocol** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Target** | **Primer sequence** | **polymerase** | **initial denature** | **denature** | **annealing** | **extension** | **cycles** | **final extension** |
| inv (3) transgene | 5'-GCAGCTGCTCTGAGACAAGTCTG-3'<br>5'-GCACTGTGGTCTAATTGCCTGATC-3'<br>5'-CTAGGCCACAGAATTGAAAGATCT-3'<br>5'-GTAGGTGGAAATTCTAGCATCATCC-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 94˚C, 2 min | 94˚C, 30 sec | 60˚C, 30 sec | 72˚C, 1 min | 30 | 72˚C, 7 min |
| Sf3b1 floxed | 5'-TCATGGGCTGTGCTATCTTG-3'<br>5'-GCACTGATGGTCCGAACTTT-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 95˚C, 3 min | 95˚C, 30 sec | 59˚C. 30 sec | 72˚C. 1 min | 35 | 72˚C, 5 min |
| Mx1-Cre | 5'-CAAGTGACAGCAATGCTGTTTCAC-3'<br>5'-GGCTGGCCCTGTATTCCTGAT-3'<br>5'-CAGGTATCTCTGACCAGAGTCATC-3'<br>5'-TCTTCTGACCCTTCCCTACTGAGC-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 95˚C, 2 min | 95˚C, 30 sec | 68˚C, 90 sec | 72˚C. 1 min | 35 | 72˚C, 3 min |
| **RT-PCR for confirming Sf3b1-K700E expression** | | **PCR protocol** | | | | | | |
| mouse Sf3b1 | 5'-GCTGTGTGCAAAAGCAAGAAG-3'<br>5'-TCCTCTGTGTTGGCGGATAC-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 95˚C, 2 min | 95˚C, 30 sec | 55˚C, 30 sec | 72˚C, 1 min | 35 | 72˚C, 7 min |
| **RT-PCR for detecting EVI1+18** | | **PCR protocol** | | | | | | |
| human EVI1 exon 12 to 13 from human cell lines | 5'-CAACAAACCAATTTAGACAGACACC-3'<br>5'-GTCATCCAGAATCGCACCTG-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 95˚C, 2 min | 95˚C, 30 sec | 55˚C, 30 sec | 72˚C, 1 min | 35 | 72˚C, 7 min |
| human EVI1 exon 12 to 13 from mice with human inv(3) Tg | 5'-GCCATTTAAGTGTCACTTATGTGATAGG-3'<br>5'-CTTCTTTGTCATCCAGAATCGCA-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 95˚C, 2 min | 95˚C, 30 sec | 63˚C, 30 sec | 72˚C, 1 min | 35 | 72˚C, 7 min |
| **RT-PCR for lariat sequencing** | | **PCR protocol** | | | | | | |
| human EVI1 intron 12 outer | 5'-GCTAGTGAGAGAAGCACCTTC-3'<br>5'-CATAATCCAAATAAGGCCACTG-3' | DreamTaq Green master mix(K1081, Thermo Fisher Scientiffic) | 95˚C, 2 min | 95˚C, 30 sec | 55˚C, 30 sec | 72˚C, 30 sec | 35 | 72˚C, 5 min |
| human EVI1 intron 12 inner | 5'-GTACTTCATGATTTCAGTGGTTC-3'<br>5'-CCTGATTTTGGCGTCAAAATGG-3' | DreamTaq Green master mix(K1081, Thermo Fisher Scientiffic) | 95˚C, 2 min | 95˚C, 30 sec | 55˚C, 30 sec | 72˚C, 30 sec | 35 | 72˚C, 5 min |
| **RT-PCR for minigene assay** | | **PCR protocol** | | | | | | |
| minigene-derived human EVI1 | 5'-TGGTCAACAAACCAATTTAGACAGAC-3'<br>5'-ATCAGCGAGCTCTAGCATTTAGG-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 95˚C, 2 min | 95˚C, 30 sec | 58˚C, 30 sec | 72˚C, 1 min | 40 | 72˚C, 7 min |
| endogenous human EVI1 | 5'-TGGTCAACAAACCAATTTAGACAGAC-3'<br>5'-CTGGTCACCAAAGCCTTTTCATC-3' | 2xGoTaq GreenMaster Mix (M7123, Promega) | 95˚C, 2 min | 95˚C, 30 sec | 58˚C, 30 sec | 72˚C, 1 min | 40 | 72˚C, 7 min |
| **EVI1 minigene cloning** | | **PCR protocol** | | | | | | |
| EVI1 minigene, WT | 5'-AAAAAAGGATCCATGCAAATACTGTGACAGATC-3'<br>5'-AAAAAACTCGAGCTCTCCTCCACATTCCTGG-3' | Phusion High-Fidelity DNA Polymerase (M0530S, New England Biolabs) | 98˚C, 1 min | 98˚C, 10 sec | 60˚C, 30 sec | 72˚C, 90 sec | 35 | 72˚C, 10 min |
| MT1 | 5'-GATAAACAATAATCTTTGTCATAAACGGCTTTCTG CTCCACAC-3'<br>5'-GTGTGGAGCAGAAAGCCGTTTATGACAAAGATTA TTGTTTATC-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT2 | 5'-GCTAGATAAACAATAATCTTTGTCCTAAAAGGCTT TCTGCTCCAC-3'<br>5'-GTGGAGCAGAAAGCCTTTTAGGACAAAGATTATTG TTTATCTAGC-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT3 | 5'-GTTGCTAGATAAACAATACTCTTTGTCATAAAAGG CTTTC-3'<br>5'-GAAAGCCTTTTATGACAAAGAGTATTGTTTATCTA GCAAC-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT4 | 5'-GTTGCTAGATAAACAATAATCTTTTTCTTTTTAGG CTTTCTGCTCCACAC-3'<br>5'-GTGTGGAGCAGAAAGCCTAAAAAGAAAAAGATTAT TGTTTATCTAGCAAC-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT5 | 5'-TTGTTTATCTAGCAACTTATTTG-3'<br>5'-TAATCTTTGTCATTTTAGGCTTTCTGC-3' | Q5 Hot Start High-Fidelity 2X Master Mix (E0552, New England Biolabs) | 98˚C, 30 sec | 98˚C, 10 sec | 55˚C, 15 sec | 72˚C, 5 min | 25 | 72˚C, 2 min |
| MT6 | 5'-CAAAGATTATTGTTTATCTAGC-3'<br>5'-TCATAAAAGGCTTTCAAAACCACACAGGTA-3' | Q5 Hot Start High-Fidelity 2X Master Mix (E0552, New England Biolabs) | 98˚C, 30 sec | 98˚C, 10 sec | 55˚C, 15 sec | 72˚C, 5 min | 25 | 72˚C, 2 min |
| MT7 | 5'-CAAAGATTATTGTTTATCTAGC-3'<br>5'-TCATAAAAGGCTTTCTGAAAAACACAGGTACAGC-3' | Q5 Hot Start High-Fidelity 2X Master Mix (E0552, New England Biolabs) | 98˚C, 30 sec | 98˚C, 10 sec | 55˚C, 15 sec | 72˚C, 5 min | 25 | 72˚C, 2 min |
| MT8 | 5'-CTTTGTCATAAAAGGCTTACTGCTCCACACAGGT ACAG-3'<br>5'-CTGTACCTGTGTGGAGCAGTAAGCCTTTTATGAC AAAG-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT9 | 5'-CTTTGTCATAAAAGGCTTTATGCTCCACACAGGT ACAG-3'<br>5'-CTGTACCTGTGTGGAGCATAAAGCCTTTTATGAC AAAG-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT10 | 5'-CTTTGTCATAAAAGGCTTAATGCTCCACACAGGT ACAG-3'<br>5'-CTGTACCTGTGTGGAGCATTAAGCCTTTTATGAC AAAG-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT11 | 5'-CTTTGTCATAAAAGGCTAAATGCTCCACACAGGT ACAG-3'<br>5'-CTGTACCTGTGTGGAGCATTTAGCCTTTTATGAC AAAG-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT12 | 5'-CTTTGTCATAAAAGGCAAACTGCTCCACACAGGT ACAG-3'<br>5'-CTGTACCTGTGTGGAGCAGTTTGCCTTTTATGAC AAAG-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT13 | 5'-AATCTTTGTCATAAAAGGAAATCTGCTCCACACAG GTAC-3'<br>5'-GTACCTGTGTGGAGCAGATTTCCTTTTATGACAA AGATT-3' | Pfu Ultra DNA polymerase (200523, Agilent) | 95˚C, 30 sec | 95˚C, 30 sec | 55˚C, 1 min | 72˚C, 9 min | 18 | 72˚C, 15 min |
| MT14 | 5'-TTGCTAGATAAACAATcATCTTTGTCATAAAAGGC TTTC-3'<br>5'-CTTATTTGGAAACTCACTCCTTATGAATTTG-3' | Q5 Hot Start High-Fidelity 2X Master Mix (E0552, New England Biolabs) | 98˚C, 30 sec | 98˚C, 10 sec | 55˚C, 15 sec | 72˚C, 5 min | 25 | 72˚C, 2 min |
| MT15 | 5'-TCTTTGTCATgAAAGGCTTTCTGC-3'<br>5'-TTATTGTTTATCTAGCAACTTATTTGG-3' | Q5 Hot Start High-Fidelity 2X Master Mix (E0552, New England Biolabs) | 98˚C, 30 sec | 98˚C, 10 sec | 55˚C, 15 sec | 72˚C, 5 min | 25 | 72˚C, 2 min |
| MT16 | 5'-AAGTTGCTAGgTAAACAATAATCTTTGTC-3'<br>5'-ATTTGGAAACTCACTCCTTATGAATTTG-3' | Q5 Hot Start High-Fidelity 2X Master Mix (E0552, New England Biolabs) | 98˚C, 30 sec | 98˚C, 10 sec | 55˚C, 15 sec | 72˚C, 5 min | 25 | 72˚C, 2 min |

| **Sanger sequence** | |
|---|---|
| polylinker of pMYs, Fw | 5'-CCCTTGAACCTCCTCGTTCGACC-3' |
| polylinker of pcDNA3, Rv | 5'-TAGAAGGCACAGTCGAGG-3' |

| **qPCR** | | |
|---|---|---|
| mBcl11a | 5'-CACGTCCGCACTTGAACTTG-3' | 5'-AGATGAATTGTGGGAGAGCCG-3' |
| mCd34 | 5'-GGTAGCTCTCTGCCTGATGAG-3' | 5'-TGGTAGGAACTGATGGGGATATT-3' |
| mHes1 | 5'-CCAGCCAGTGTCAACACGA-3' | 5'-AATGCCGGGAGCTATCTTTCT-3' |
| mMeis1 | 5'-CATGATAGACCAGTCCAACCGA-3' | 5'-ATTGGCTGTCCATCAGGGTTA-3' |

**Supplementary Table 2**. Patient characteristics of *EVI1* rearranged myeloid neoplasms with or without *SF3B1* mutations.

| | **Total cohort (n=46)** | ***SF3B1*-mut (n=18)** | ***SF3B1*-WT (n=28)** | **P-value*** |
|---|---|---|---|---|
| Age at diagnosis (mean, SD) | 64.3 (13.4) | 62.8 (12.6) | 65.3 (14.0) | 0.55 |
| Male sex (n, %) | 29 (63.0%) | 11 (61.1%) | 18 (64.3%) | 1.00 |
| Disease (n, %) | | | | |
| - AML | 25 (54.4%) | 10 (55.6%) | 15 (53.6%) | |
| - MDS | 16 (34.8%) | 7 (38.9%) | 9 (32.1%) | |
| - Other (blast-phase CML, CMML) | 5 (10.9%) | 1 (5.6%) | 4 (14.3%) | 0.76 |
| Therapy-related MN (n, %) | 7 (15.2%) | 1 (5.6%) | 6 (21.4%) | 0.22 |
| Number of lines of therapy (mean, SD) | 2.2 (1.8) | 2.6 (1.6) | 2.0 (1.9) | 0.24 |
| Best response to first line therapy | | | | |
| - CR/mCR/CRi/PR/MLFS | 9 (19.6%) | 2 (11.1%) | 7 (25.0%) | |
| - SD | 7 (15.2%) | 3 (16.7%) | 4 (14.3%) | |
| - Progressive disease/primary induction failure | 24 (52.3%) | 12 (66.7%) | 12 (42.9%) | |
| - not evaluable | 6 (13.0%) | 1 (5.6%) | 5 (17.9%) | 0.35 |
| Allogeneic hematopoietic cell transplant (n, %) | 13 (28.9%) | 7 (38.9%) | 6 (22.2%) | 0.32 |
| Concurrent cytogenetic abnormalities (n, %) | | | | |
| - Del(5q) | 6 (13.0%) | 0 | 6 (21.4%) | 0.07 |
| - Monosomy 7 | 22 (47.8%) | 9 (50.0%) | 13 (46.4%) | 1.00 |
| - 17p abnormality or monosomy 17 | 4 (8.7%) | 1 (5.6%) | 3 (10.7%) | 1.00 |
| - Complex karyotype | 12 (26.1%) | 4 (22.2%) | 8 (28.6%) | 0.74 |
| - Monosomal karyotype | 20 (43.5%) | 5 (27.8%) | 15 (53.6%) | 0.13 |
| - t(9;22) | 3 (6.5%) | 1 (5.6%) | 2 (7.1%) | 1.00 |

* Student's t-test for continuous variables and Pearson chi-square or Fisher's exact test for categorical variables

**Supplementary Table 3.** *EVI1*-rearranged leukemia cell lines.

| cell line | *SF3B1* | *EVI1* rearrangement | age | sex | diagnosis |
|---|---|---|---|---|---|
| HNT-34 | K700E | t(3;3)(q21;q26) | 45 | F | CMMoL overt AML |
| MUTZ-3 | K666N | inv(3)(q21q26) | 29 | M | AML |
| YCU-AML1 | K700E | t(3;3)(q21;q26.2) | 62 | M | AML-MRC (MDS/AML) |
| OCI-AML-20 | wild-type | inv(3)(q21q26.2) | 34 | M | AML |
| MOLM-1 | wild-type | inv(3)(q21q26) | 41 | M | CML-BC |
| Kasumi-3 | wild-type | t(3:7) (q27;q22) | 57 | M | AML (M0) |
| Kasumi-4 | wild-type | inv(3)(q21q26) | 6 | F | CML-BC |
| UCSD-AML1 | wild-type | t(3;3)(q21;q26) | 73 | F | AML |

CMMoL, Chronic myelomonocytic leukemia; AML-MRC, acute myeloid leukemia with myelodysplasia-related changes (AML-MRC); CML-BC, chronic myeloid leukemia-blast crisis.

**Supplementary Table 4.** Mutational analysis of *EVI1*-rearranged leukemia cell lines.

| Cell line | Gene | Gene panel | Protein Change | Annotation | Chromosome | Start Pos | End Pos | Ref | Var | Allele Freq |
|---|---|---|---|---|---|---|---|---|---|---|
| HNT-34 | SF3B1 | N/A | K700E | OncoKB: Likely Oncogenic, level_4, resistance NA;CIViC: Predictive: 1;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 2 | 198266834 | 198266834 | T | C | 0.509043928 |
| HNT-34 | PTPN11 | N/A | T468M | OncoKB: Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: yes | 12 | 112926270 | 112926270 | C | T | 0.509186352 |
| HNT-34 | PTPN11 | N/A | A72V | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: yes | 12 | 112888199 | 112888199 | C | T | 0.492900609 |
| HNT-34 | RAD21 | N/A | T28Ifs*11 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 8 | 117878886 | 117878887 | - | AA | 0.350584307 |
| HNT-34 | BCOR | N/A | S1446Vfs*15 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 23 | 39921485 | 39921486 | - | T | 0.526315789 |
| MUTZ-3 | SF3B1 | N/A | K666N | OncoKB: Likely Oncogenic, level_4, resistance NA;CIViC: Predictive: 1;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 2 | 198267359 | 198267359 | C | A | 0.462167689 |
| MUTZ-3 | RECQL4 | N/A | X766_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 8 | 145738768 | 145738768 | G | C | 0.744186047 |
| MUTZ-3 | ARID1A | N/A | Y1101Cfs*3 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 1 | 27097708 | 27097709 | TC | - | 0.465397924 |
| MUTZ-3 | KRAS | N/A | G10dup | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 12 | 25398287 | 25398288 | - | CTC | 0.451219512 |
| MUTZ-3 | CREBBP | N/A | R589Kfs*15 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 16 | 3830787 | 3830790 | TTCC | - | 0.334801762 |
| MUTZ-3 | CTCF | N/A | X613_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 16 | 67670591 | 67670591 | A | G | 0.551378446 |
| OCI-AML-20 | WT1 | N/A | L1? | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 11 | 32456890 | 32456891 | AG | GA | 0.061452514 |
| OCI-AML-20 | NRAS | N/A | Q61K | OncoKB: Oncogenic, level NA, resistance NA;CIViC: Predictive: 20, Predisposing: 2;MyCancerGenome: present;CancerHotspot: yes;3DHotspot: yes | 1 | 115256530 | 115256530 | G | T | 0.457905544 |
| OCI-AML-20 | MAX | N/A | R36W | OncoKB: Predicted Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 14 | 65560491 | 65560491 | T | A | 0.028225806 |
| OCI-AML-20 | ATRX | N/A | X199_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 23 | 76940496 | 76940496 | A | G | 0.056701031 |
| OCI-AML-20 | DICER1 | N/A | M1? | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 14 | 95599795 | 95599795 | T | A | 0.078431373 |
| OCI-AML-20 | MED12 | N/A | S63N | OncoKB: Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 23 | 70339311 | 70339311 | G | A | 0.067510549 |
| OCI-AML-20 | KMT2D | N/A | G1628Vfs*94 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 12 | 49438607 | 49438607 | C | - | 0.057894737 |
| OCI-AML-20 | CDK12 | N/A | F1376Sfs*5 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 17 | 37687220 | 37687220 | C | - | 0.078549849 |
| OCI-AML-20 | SMAD4 | N/A | X263_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 18 | 48584615 | 48584709 | GTATGTACATACTTTAAAAAATCTTTTAAATAGTTGAGAAAAAAGTAGGCAGCCTTTATAAAAGCAAATTAACCCATGTGGGCCTTAATTTTTAG | - | 0.376556017 |
| OCI-AML-20 | SUZ12 | N/A | X625_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 17 | 30325677 | 30325677 | G | A | 0.062656642 |
| OCI-AML-20 | ARID2 | N/A | Q1835* | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 12 | 46298856 | 46298856 | C | T | 0.082051282 |
| OCI-AML-20 | KMT2D | N/A | X5508_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 12 | 49415653 | 49415653 | T | C | 0.053097345 |
| OCI-AML-20 | BRCA2 | N/A | 7_I2628delinsSIWVANHYF | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 13 | 32936705 | 32936736 | AATTTGGGTTTATAATCACTATAGATGGATCA | CATCTGGGTTGCTAACCATTACAGGTGGATTG | 0.055555556 |
| OCI-AML-20 | NF1 | N/A | X1574_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 17 | 29588873 | 29588873 | T | C | 0.070953437 |
| OCI-AML-20 | PTPR5 | N/A | X1256_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 19 | 5219476 | 5219476 | G | T | 0.06302521 |
| OCI-AML-20 | PPP2R1A | N/A | X308_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 19 | 52719147 | 52719259 | GTTGGTGCTGGCAGCCGGAACACAGCAAGTGGGGTGGGTATCCAAGGGGCTGGAGGTGGAACTAGCACATCAGGTCTCACTTCCCTTTGCCTCCCTCTCCCTGCCACAGAGT | AGG | 0.078393881 |
| OCI-AML-20 | FOXP1 | N/A | X629_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 3 | 71015043 | 71015043 | G | A | 0.077568134 |
| OCI-AML-20 | STAG2 | N/A | X14_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 23 | 123156519 | 123156519 | A | G | 0.070967742 |
| OCI-AML-20 | RBM10 | N/A | D20* | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 23 | 47028754 | 47028807 | GACCGCTCGCAGGATGATGGTGGGGAGAACCGCAGCCGAGACCACGACTACCGG | TGATCGTTCACAGGATGATAGTGGAGAGAACCACAGCTGGGACCATGATTATAGA | 0.078947368 |
| OCI-AML-20 | KLF4 | N/A | X422_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 9 | 110248206 | 110248206 | A | G | 0.129310345 |
| OCI-AML-20 | FOXP1 | N/A | X4_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 3 | 71247548 | 71247548 | T | G | 0.117370892 |
| OCI-AML-20 | PMS2 | N/A | X117_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 7 | 6043323 | 6043323 | C | T | 0.071895425 |
| OCI-AML-20 | MAP2K4 | N/A | X172_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 17 | 12011109 | 12011109 | T | C | 0.052238806 |
| OCI-AML-20 | PIK3R3 | N/A | X339_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 1 | 46511760 | 46511760 | C | T | 0.057636888 |
| OCI-AML-20 | SMAD4 | N/A | X142_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 18 | 48575230 | 48575663 | GGTAAGTAGACTTTGCTTTCATCCTAAGAAACATAAAGGGAAAAGGATCTCAATAGTGTTTCAATTTTTGTAAGTTAAATTATAAATTTGGAAGATAGCCCGCGACTTTAAATAAGGTTAAAGAATCAGACATTTTTAATATACGTATTTAAATTTGAATTTAGGAACAAACTTATATTTTGACTGCTAAAACCGAGTATTGTAGTTGATATTTTGCCCCTTTAGAACATTGTTTTTGAGAAATAAGATGGAAAGCATTATAAATTTGAATACTTAGTTATGTATTGTTAGATAGCGTTTATGCTACTTCTGAATTGAAATGGTTCATGAACTTTTGTAATCTTTGGTTTAAATTTACATTCTCTGTTTTTAAATGGAAAATACTTTCATTGTAATGATTAATGTTTCATTTGTTTTCCCCTTTAAACAATTAA | - | 0.066308244 |
| OCI-AML-20 | NF2 | N/A | X81_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 22 | 30035081 | 30035082 | AC | GT | 0.065502183 |
| OCI-AML-20 | HIST1H1B | N/A | A164Gfs*42 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 6 | 27834817 | 27834818 | - | CCGCGGGCTTCTTCGCCTTCTTCGGAGTCTTCTTCGGCCGCCTTTTTGGGTGTGGCAGCACCAGTCGCCTTCTTAGGCTTCTTGGCTGCTC | 0.217391304 |
| OCI-AML-20 | KMT2D | N/A | S849Lfs*30 | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 12 | 49444921 | 49444922 | - | TTCCTCAGGCCGGGGGGACAGGCATGGCTCCTCAGACTGGGGGGACAGGCGGCAGCTCCTTAGGTGCAGAGCTCTGGCTTGGTTCCTCAGGAGGCTCTGCAGGCTGAGGAGGTCCATTTAG | 0.064159292 |
| OCI-AML-20 | KMT2C | N/A | X2480_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 7 | 151876921 | 151876921 | A | G | 0.057142857 |
| OCI-AML-20 | RBM10 | N/A | X731_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 23 | 47045185 | 47045461 | TGGAGGTGAGGTGTGACCTGACCCTGGGCTCCCTCCCCTGTGTCCCTTCCAAGTCCCCATCACCTCAGGCAGCTGGAGTTCAATTTCTCACGTGTTAACCCCGTGAGCCTTCTTCCCCATTTTCCTATCCTGGCTGCTTGTGCCTGTGATGGCCTGACCCTCAGGCACCTGAGCCTCATTAGCCAGATGGCATGCCCTGTGGGACCCCACTCCCCATGCCTGTGTTGCCTGAGAAAAAGAGACCAGCTCCCCAACATTCACATACACATACAAACTTTC | CGG | 0.06557377 |
| MOLM-1 | TP53 | N/A | R196* | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 17 | 7578263 | 7578263 | G | A | 0.997409326 |
| MOLM-1 | RECQL4 | N/A | X766_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 8 | 145738768 | 145738768 | G | C | 0.787234043 |
| MOLM-1 | HLA-B | N/A | E69G | OncoKB: Predicted Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 6 | 31324602 | 31324602 | T | C | 0.400479616 |
| Kasumi-3 | TP53 | N/A | X261_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 17 | 7577498 | 7577498 | C | T | 0.489583333 |
| Kasumi-3 | TP53 | N/A | I162F | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 17 | 7578446 | 7578446 | T | A | 0.495639535 |
| Kasumi-3 | GRIN2A | N/A | X723_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 16 | 9892321 | 9892321 | C | T | 0.482758621 |
| Kasumi-4 | HLA-B | N/A | E69G | OncoKB: Predicted Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: no | 6 | 31324602 | 31324602 | T | C | 0.259958071 |
| Kasumi-4 | PTPN11 | N/A | Y62D | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 12 | 112888168 | 112888168 | T | G | 0.55165692 |
| UCSD-AML1 | KIT | N/A | D816V | OncoKB: Oncogenic, level NA, resistance NA;CIViC: Prognostic: 2, Predictive: 12;MyCancerGenome: present;CancerHotspot: yes;3DHotspot: no | 4 | 55599321 | 55599321 | A | T | 0.518382353 |
| UCSD-AML1 | PTPN11 | N/A | D61V | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: yes;3DHotspot: yes | 12 | 112888166 | 112888166 | A | T | 0.496389892 |
| UCSD-AML1 | TERT | N/A | Promoter | OncoKB: Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 5 | 1295228 | 1295228 | G | A | 0.492753623 |
| UCSD-AML1 | MRE11 | N/A | X36_splice | OncoKB: Likely Oncogenic, level NA, resistance NA;CIViC: NA;MyCancerGenome: not present;CancerHotspot: no;3DHotspot: no | 11 | 94226952 | 94226952 | C | T | 0.504823151 |

**Supplementary Table 5.** Fraction of *EVI1*+18 variant in *SF3B1*-mutated and wildtype cases.

| | *SF3B1* | *SF3B1* -mutation VAF | *EVI1* Canonical | *EVI1* Novel | Fraction of *EVI1* novel (%) |
|---|---|---|---|---|---|
| *SF3B1* - mutated | p.G740E | 35.90% | 80 | 54 | 40.30% |
| | p.G740E | 44.00% | 785 | 483 | 38.09% |
| | p.G740E | 49.12% | 33 | 20 | 37.74% |
| | p.R625C | 47.00% | 381 | 186 | 32.80% |
| | p.H662Q | 51.27% | 217 | 73 | 25.17% |
| | p.K700E | 54.31% | 66 | 19 | 22.35% |
| | p.K700E | 49.00% | 282 | 68 | 19.43% |
| | p.K700E | 45.82% | 180 | 37 | 17.05% |
| | p.K700E | 22.92% | 258 | 50 | 16.23% |
| | p.K700E | 44.00% | 158 | 29 | 15.51% |
| | p.K700E | 50.00% | 1759 | 305 | 14.78% |
| | p.K700E | 43.75% | 172 | 29 | 14.43% |
| | p.K700E | 55.67% | 65 | 6 | 8.45% |
| | p.K666N | 42.01% | 41 | 3 | 6.82% |
| | p.K666T | 43.60% | 197 | 11 | 5.29% |
| | p.K666N | 48.77% | 107 | 2 | 1.83% |
| *SF3B1* - wild-type | wild-type | (-) | 72 | 1 | 1.37% |
| | wild-type | (-) | 82 | 1 | 1.20% |
| | wild-type | (-) | 178 | 2 | 1.11% |
| | wild-type | (-) | 107 | 1 | 0.93% |
| | wild-type | (-) | 253 | 2 | 0.78% |
| | wild-type | (-) | 140 | 1 | 0.71% |
| | wild-type | (-) | 185 | 1 | 0.54% |
| | wild-type | (-) | 263 | 1 | 0.38% |
| | wild-type | (-) | 264 | 1 | 0.38% |
| | wild-type | (-) | 542 | 2 | 0.37% |
| | wild-type | (-) | 567 | 2 | 0.35% |
| | wild-type | (-) | 1282 | 2 | 0.16% |
| | wild-type | (-) | 18 | 0 | 0.00% |
| | wild-type | (-) | 132 | 0 | 0.00% |
| | wild-type | (-) | 781 | 0 | 0.00% |
| | wild-type | (-) | 147 | 0 | 0.00% |
| | wild-type | (-) | 298 | 0 | 0.00% |
| | wild-type | (-) | 27 | 0 | 0.00% |
| | wild-type | (-) | 384 | 0 | 0.00% |
| | wild-type | (-) | 37 | 0 | 0.00% |
| | wild-type | (-) | 55 | 0 | 0.00% |
| | wild-type | (-) | 162 | 0 | 0.00% |
| | wild-type | (-) | 132 | 0 | 0.00% |
| | wild-type | (-) | 73 | 0 | 0.00% |
| | wild-type | (-) | 146 | 0 | 0.00% |
| | wild-type | (-) | 85 | 0 | 0.00% |
| | wild-type | (-) | 139 | 0 | 0.00% |
| | wild-type | (-) | 114 | 0 | 0.00% |
| | wild-type | (-) | 15 | 0 | 0.00% |
| | wild-type | (-) | 117 | 0 | 0.00% |
| | wild-type | (-) | 130 | 0 | 0.00% |
| | wild-type | (-) | 168 | 0 | 0.00% |
| | wild-type | (-) | 2 | 0 | 0.00% |
| | wild-type | (-) | 467 | 0 | 0.00% |
| | wild-type | (-) | 277 | 0 | 0.00% |
| | wild-type | (-) | 127 | 0 | 0.00% |

**Supplementary Figure 1. The genetic characteristics of *EVI1*-rearranged (*EVI1*-r) myeloid neoplasms.** (**A**) Schematic image of inv(3) resulting in *EVI1* expression by *GATA2* distal hematopoietic enhancer (G2DHE). The bacterial artificial chromosome (BAC) we utilized in the murine model is also shown. (**B**) Diagram of location of 18 *SF3B1* mutations identified in *EVI1*-r myeloid neoplasms of MSKCC cohort (n=46). HD, HEAT-repeat domain. (**C**) Correlations between driver mutations in entire MDS/AML cohorts (left) and *EVI1*-r myeloid neoplasms (right). Significantly co-occurring and mutually exclusive mutations are shown in red and blue circles, respectively. Odds ratio and associated -$\log_{10}$(Q-value) are indicated by the color gradient and size of circles, respectively. Q-values were calculated by Benjamini-Hochberg (BH) adjustment from p-values obtained from Fisher's exact test. (**D**) Overall survival (OS) from the time of diagnosis and from the time of inv(3) detection in inv(3) AML with (red) or without (blue) *SF3B1* mutations.

**Supplementary Figure 2. The impact of combined inv(3) and *Sf3b1*$^{K700E}$ mutations on hematopoiesis and leukemogenesis. (A)** Mean number of colonies derived from bone marrow mononuclear cells from *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ mice and controls. Pre-B colonies are shown on left and BFU-E are on right. Mean ± SD. **(B)** Peripheral blood counts overtime following transplantation. **(C)** As in Figure 2E, but for LT-HSC, ST-HSC, GMP, and MEP. **(D)** BM cytospins of MDS-derived AML mice indicating immature blasts and dysplastic cells (black and red arrows, respectively) within the *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ group. Scale Bars, 20 μm; x1,000 magnification. **(E)** The cause of death within the *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ group. WBC, white blood cell count; Hb, hemoglobin; MCV, mean corpuscular volume; PLT, platelet count; LT-HSC, long-term hematopoietic stem cells, CD150$^{+}$CD48$^{-}$LSK; ST-HSC, short-term hematopoietic stem cells, CD150$^{-}$CD48$^{-}$LSK; MEP; megakaryocyte-erythrocyte progenitor, CD34$^{-}$FcγR$^{-}$Lin$^{-}$Kit$^{+}$ScaI$^{-}$; GMP, granulocyte-monocyte progenitor, CD34$^{+}$FcγR$^{+}$Lin$^{-}$Kit$^{+}$ScaI$^{-}$. P values were calculated by two-sided t-test, *P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001.

**Supplementary Figure 3. Bone marrow cytomorphology of inv(3) *Sf3b1* double mutant recipient mice and controls. (A)** *Hematoxylin and eosin stain of bone marrow sections of 24-week-old recipient CD45.1 mice transplanted with bone marrow cells of mice with each of the indicated genotypes. The samples were collected 12 weeks after pIpC injection. The BM of the Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ animals exhibited hypercellularity and a monomorphic cell population. Scale Bars, 100 μm; ×100 magnification. **(B)** BM cytospins of mice from (A) indicating blasts (red arrows) within the *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ group. Scale Bars, 20 μm; ×600 magnification.

**Supplementary Figure 4. The inv(3) allele rescues the self-renewal defect of mutant SF3B1. (A)** Schema of competitive transplantation of CD45.2 *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$ mice and single mutant controls. **(B)** % of CD45.2$^{+}$ peripheral blood cells in primary (1$^{o}$) and secondary competitive transplantation. Mean ± standard deviation shown. **(C)** % of donor-derived (CD45.2$^{+}$) B220$^{+}$, CD11b$^{+}$Gr1$^{+}$, and LSK cells in bone marrow (BM) and/or spleen following 5 months of transplantation. **(D)** As in (C), but for stem and progenitor fractions. MPP, multipotent progenitors, CD150$^{-}$CD48$^{+}$LSK; CMP, common myeloid progenitor, CD34$^{+}$FCγR$^{+}$Lin$^{-}$Kit$^{+}$ScaI$^{-}$. P values were calculated by two-sided t-test, *P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001.

**Supplementary Figure 5. Combined impact of mutations in *SF3B1* and *EVI1* rearrangement on gene expression and splicing. (A)** Significantly dysregulated Gene Ontology (GO) pathways in the transcripts of *Mx1*-Cre inv(3), *Mx1*-Cre *Sf3b1*$^{K700E/WT}$, and *Mx1*-Cre inv(3) *Sf3b1*$^{K700E/WT}$, compared to those of *Mx1*-Cre control. Three samples were

independently collected in each group. $\log_{10}$(P-values) are color-coded. **(B)** As in Figure 3F, but for the murine model. Each circle shows the number of aberrant splicing events of the indicated model.

**Supplementary Figure 6. The roles of novel EVI1+18 variant. (A)** As in Figure 4C**,** but for K562 cells expressing SF3B1 wild-type (WT), K666N, K700E, and G740E. **(B)** As in Figure 4D, sanger sequencing of cDNA showing that the same +18 nucleotides were inserted between exon 12 and 13 in the transcript of *Mx1*-Cre *Sf3b1*$^{K700E/WT}$; inv(3) mice. **(C)** The predicted structure of the 2$^{nd}$ ZF domain in the EVI1+18 variant. A superposition of AlphaFold2 models of three tandem ZF domains in the C-terminus (residues 909 - 1229) of human MECOM (UniProt accession Q03112) wild-type (grey) and mutant (orange). The 6 amino acid residues (FLLHTG) inserted in the mutant were highlighted by magenta. **(D)** The relative mRNA expression evaluated by qRT-PCR. **(E)** Representative FACS plot at day0 and day14 in the competition assay between HSPCs transduced with *EVI1* (GFP$^+$) and *EVI1*+18 (tdTomato$^+$). **(F)** The chimerism evaluated by flow cytometry (GFP$^+$ vs tdTomato$^+$) in the replating assay in M3434. **(G)** GSEA plots for the comparison of the transcripts derived from K562 cells expressing EVI1+18 (left) and EVI1 wild-type (right). P values were calculated by two-sided t-test, **P<0.01, ***P<0.001, and ****P<0.0001.

**Supplementary Figure 7. Spliceosomal disruption may be therapeutically effective against *EVI1*-rearranged AML with *SF3B1* mutations. (A)** The relative growth of each cell line under the condition with the indicated dose (μM) of indisulam for 72 hours. Each growth was calculated in comparison with that of DMSO. **(B)** IC50 (μM) of indisulam in each cell line.

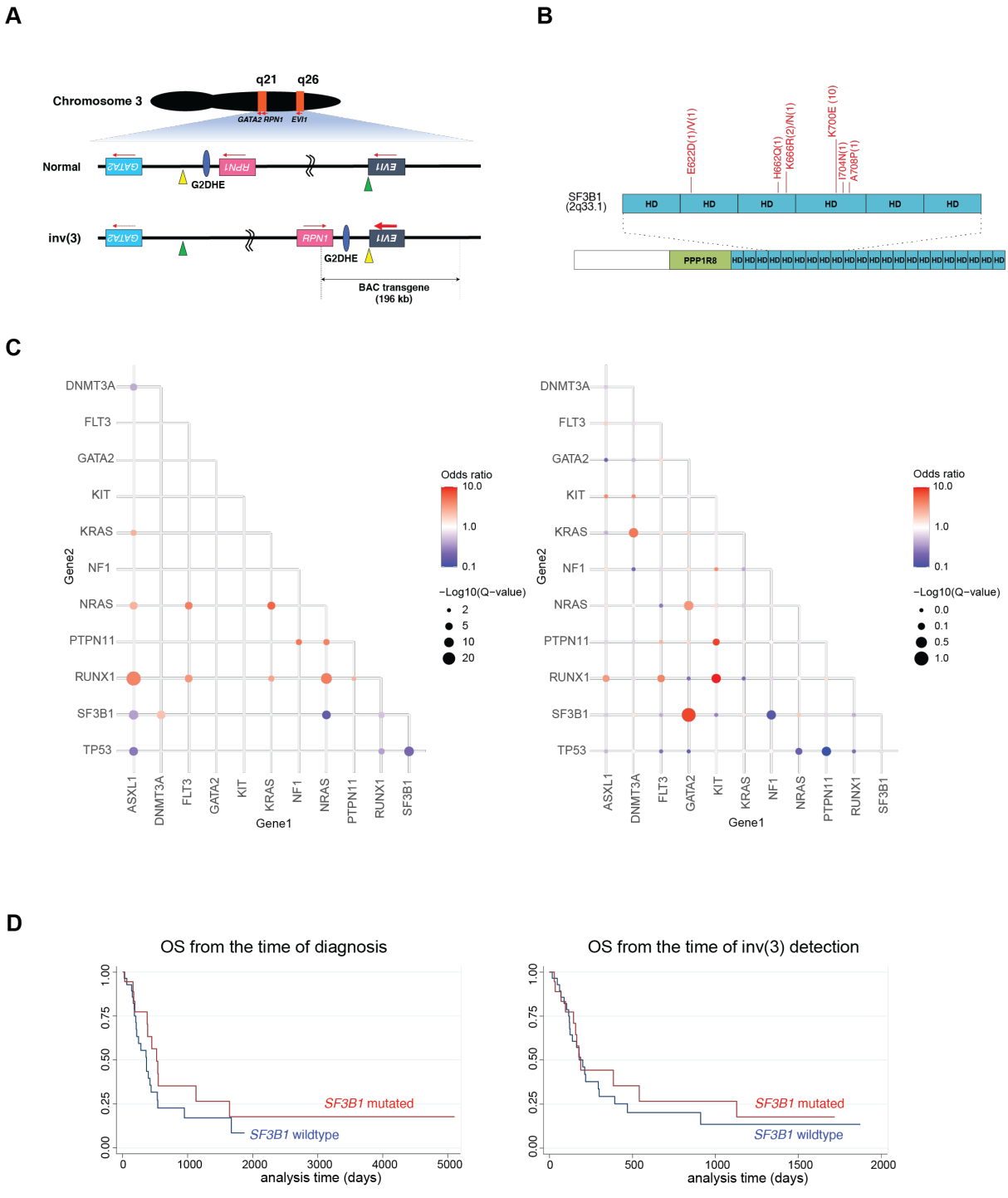
**Supplementary Figure 8. Anti-EVI1 ChIP-seq of human *EVI1*-rearranged leukemia cells. (A)** Venn diagram of EVI1 peaks from MUTZ-3, MOLM-1, and HNT-34 cell lines. **(B)** Transcription factor enrichment analysis in the genes associated with 5,698 peaks (orange section in (A)). **(C)** Coverage tracks showing EVI1 ChIP-seq occupancy at the indicated genomic loci.

**Supplementary Figure 9. Anti-HA ChIP-seq of 293T cells expressing HA-tagged EVI1 wild-type and +18 mutant. (A)** Immunoprecipitation (IP) with HA followed by immunoblotting with HA in K562 cells that express 3×HA tagged EVI1 wild-type and EVI1+18 cDNA. **(B)** Pie charts representing the distribution of HA (EVI1 and EVI1+18) binding sites on the genome of 293T cells**. (C)** Metaplots and heatmaps illustrating log2 ratios of HA peak counts per million (CPM) values of wildtype EVI1 and EVI1+18 against mock ChIP-seq data. **(D)** Metaplots and heatmaps illustrating CPM values of EVI1 and EVI1+18 ChIP-seq data. **(E)** As in (D), but for where the peaks were merged and used for illustrating both panels.

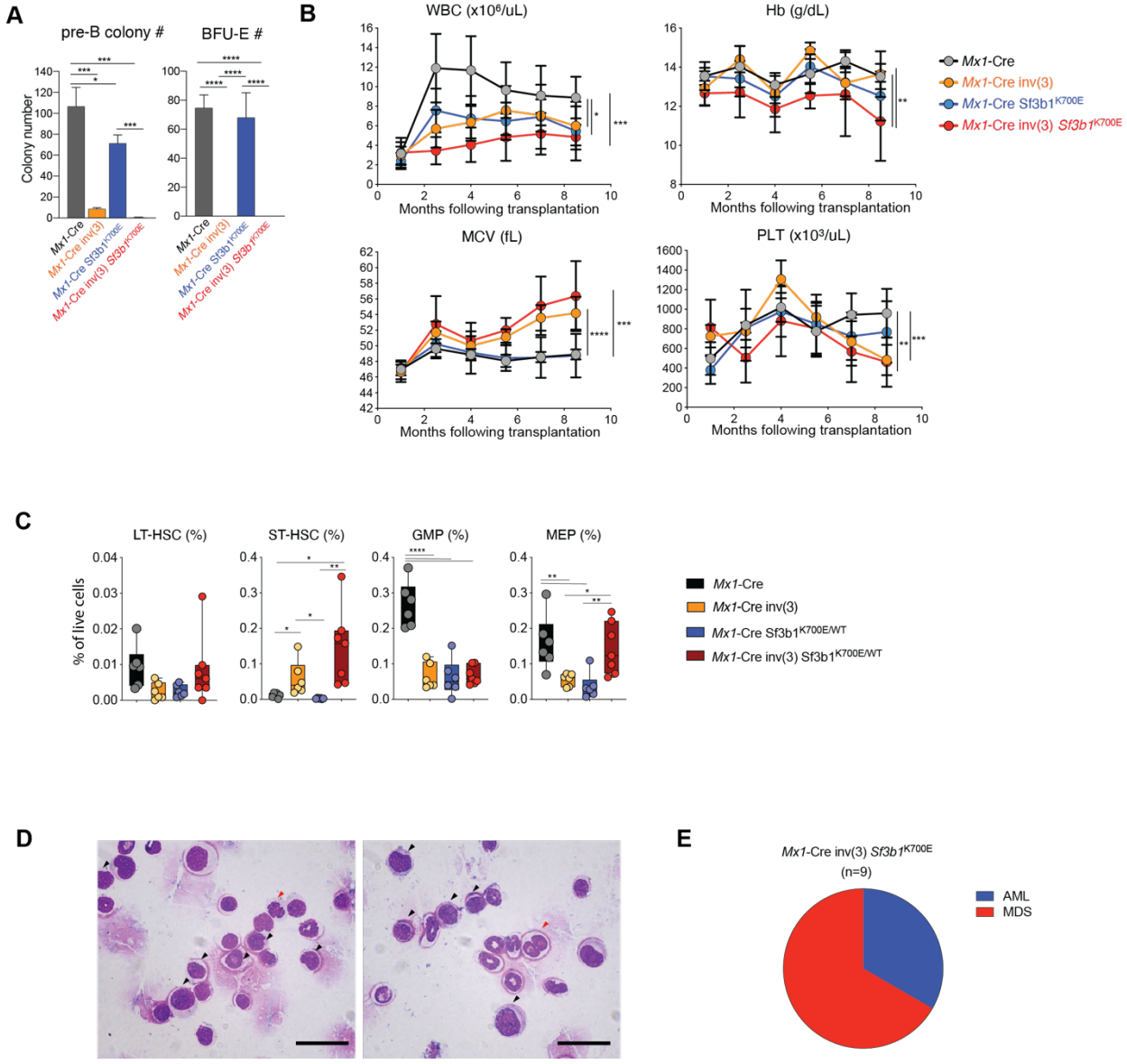
**Supplementary Figure 10. *Cis* elements within *EVI1* required for generation of the *EVI* +18 splice variant by mutant SF3B1. (A)** Schematic image of lariat RT-PCR using RNA derived from *SF3B1*-K700E knocked-in K562 cells and mapped branch sites by lariat RNA. For detailed methods, see Supplemental Methods**. (B)** Mapped branch sites by lariat-sequencing RNA derived from K562 cells with endogenous knock-in of the *SF3B1*-K700E mutation. **(C)** RT-PCR analysis of the +18 nucleotides inclusion in a minigene (top) or endogenous (bottom) context following transfection of minigenes with the illustrated mutation (MT3-1) into *SF3B1*-K666N knocked-in K562 cells. **(D)** RT-PCR analysis of the +18 nucleotides inclusion in a minigene (top) or endogenous (bottom) context following transfection of minigenes with the illustrated mutations into *SF3B1*-K666N knocked-in K562 cells and *SF3B1* wild-type K562 cells. Each individual minigene construct is shown in Figure 5A. **(E)** *BRD9* transcripts when the poison exon

(14a) is excluded (top) or included (bottom). Green and Red A indicate the branchpoint for canonical and aberrant transcripts, respectively. The exonic splicing enhancer (ESE) in the poison exon is indicated by the blue square. The mutant minigene which alters the normal AG dinucleotide (to AA, top green box) and results in addition of +18 nucleotides (blue rectangle; bottom) and use of an aberrant, intron proximal AG (orange box, bottom). Green and Red A indicate the branchpoint for canonical and aberrant transcripts, respectively. **(F)** As in (C), but for K562 parental (wild-type) cells with the minigene mutagenized at 3′ splice site (AG to AA). **(G)** Sanger sequencing of cDNA arising from the top band in (F). The nucleotide sequences are indicated. **(H)** As in (C), but for MT14 and MT15 described in Figure 5A.

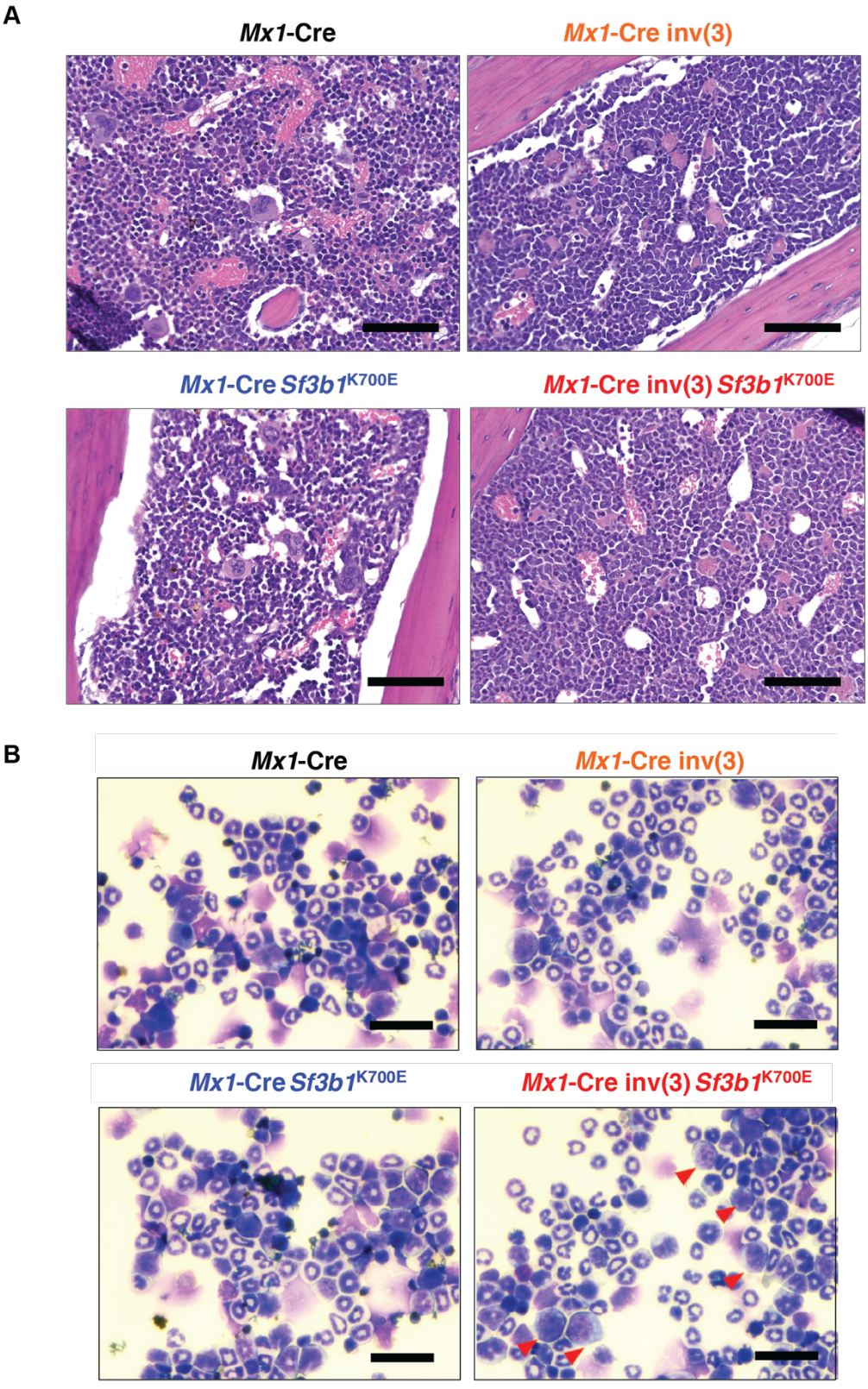
# Supplementary Figure 1



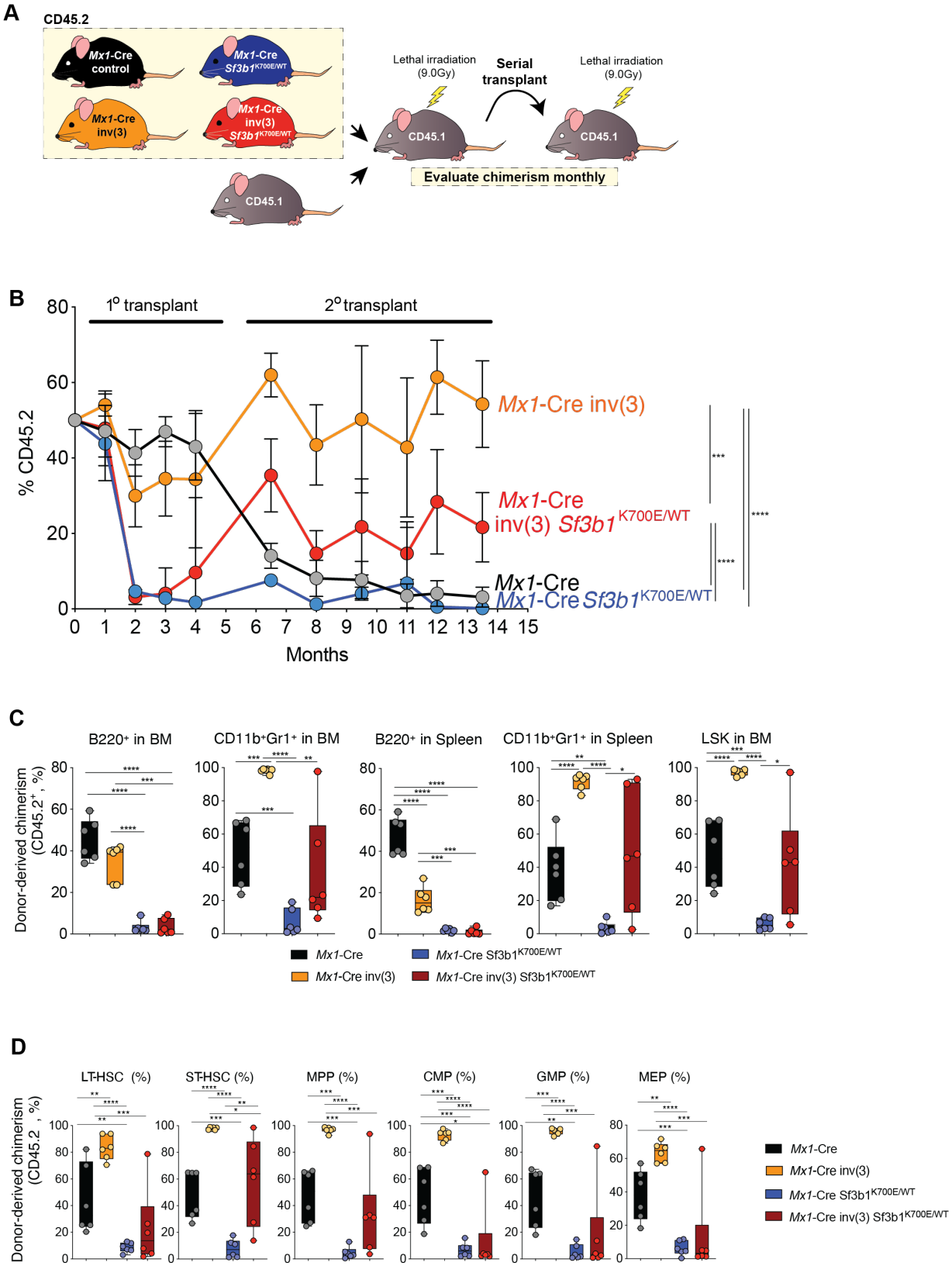
# Supplementary Figure 2



## Supplementary Figure 3

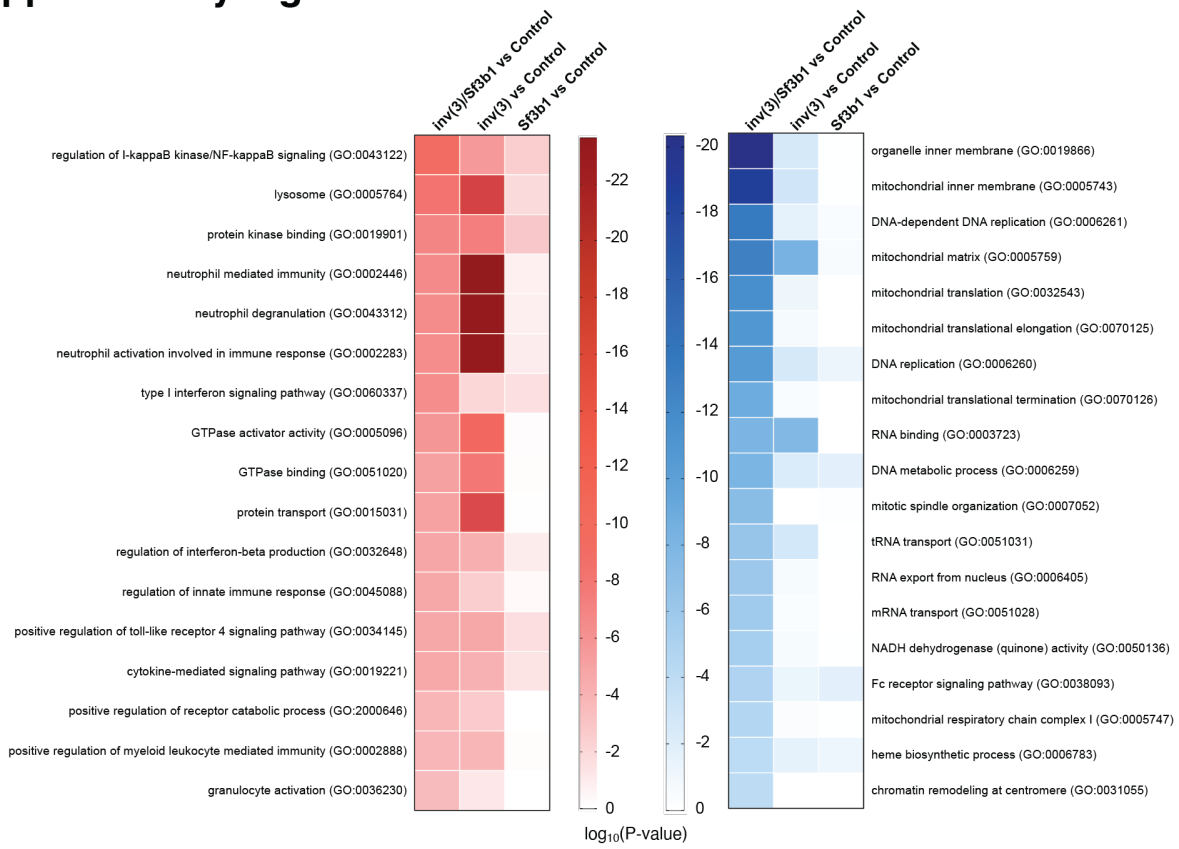


# Supplementary Figure 4

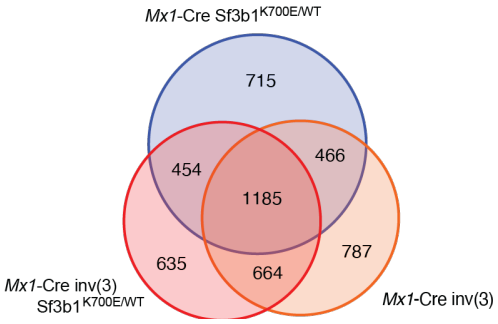


# Supplementary Figure 5

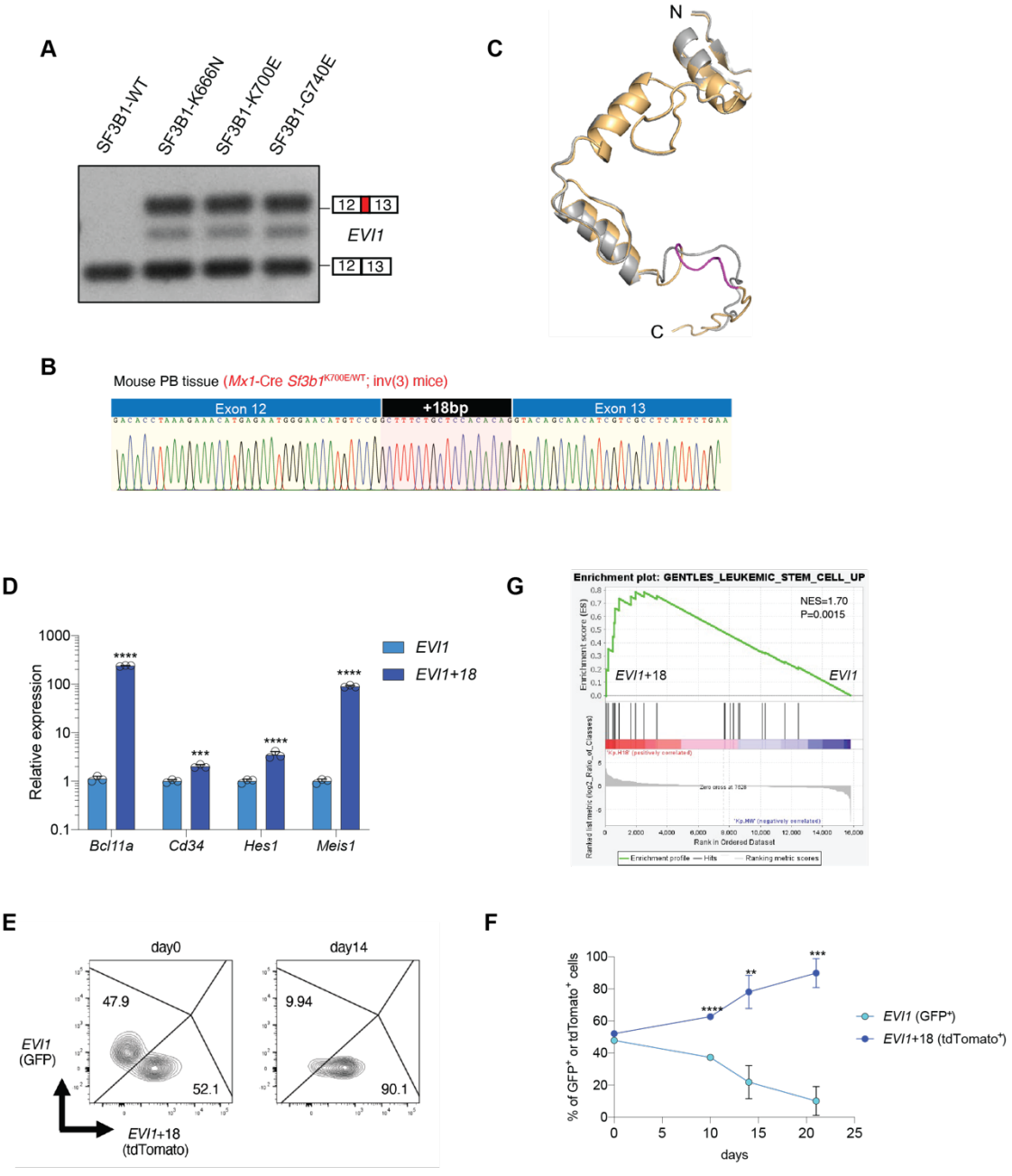
**A**



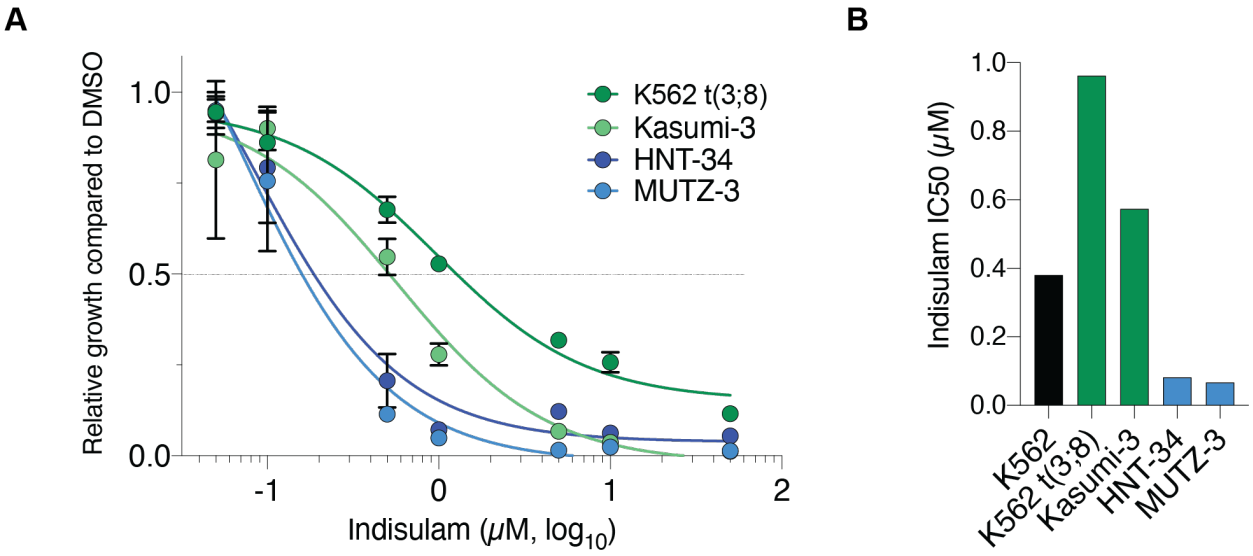

**B**



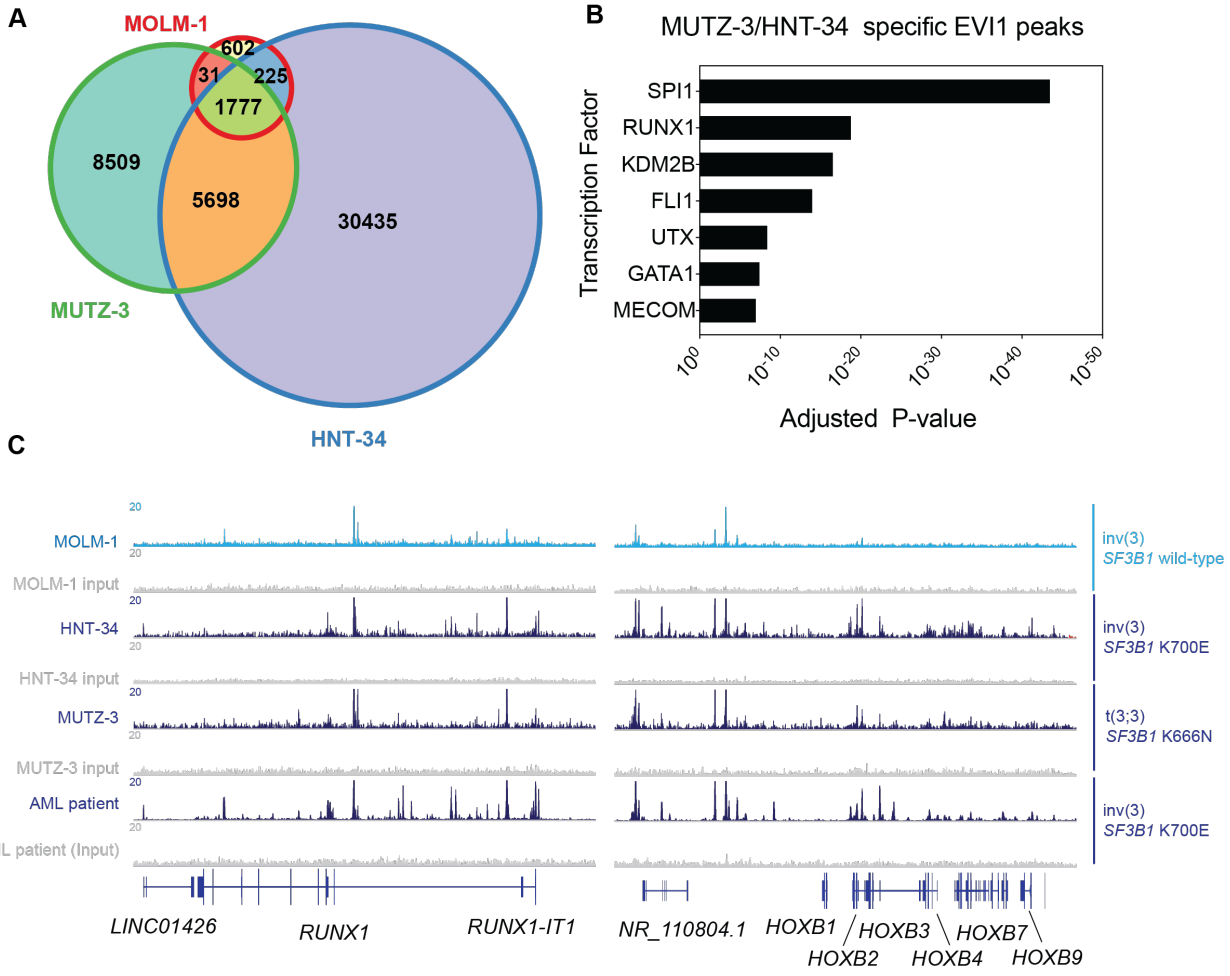
## Supplementary Figure 6



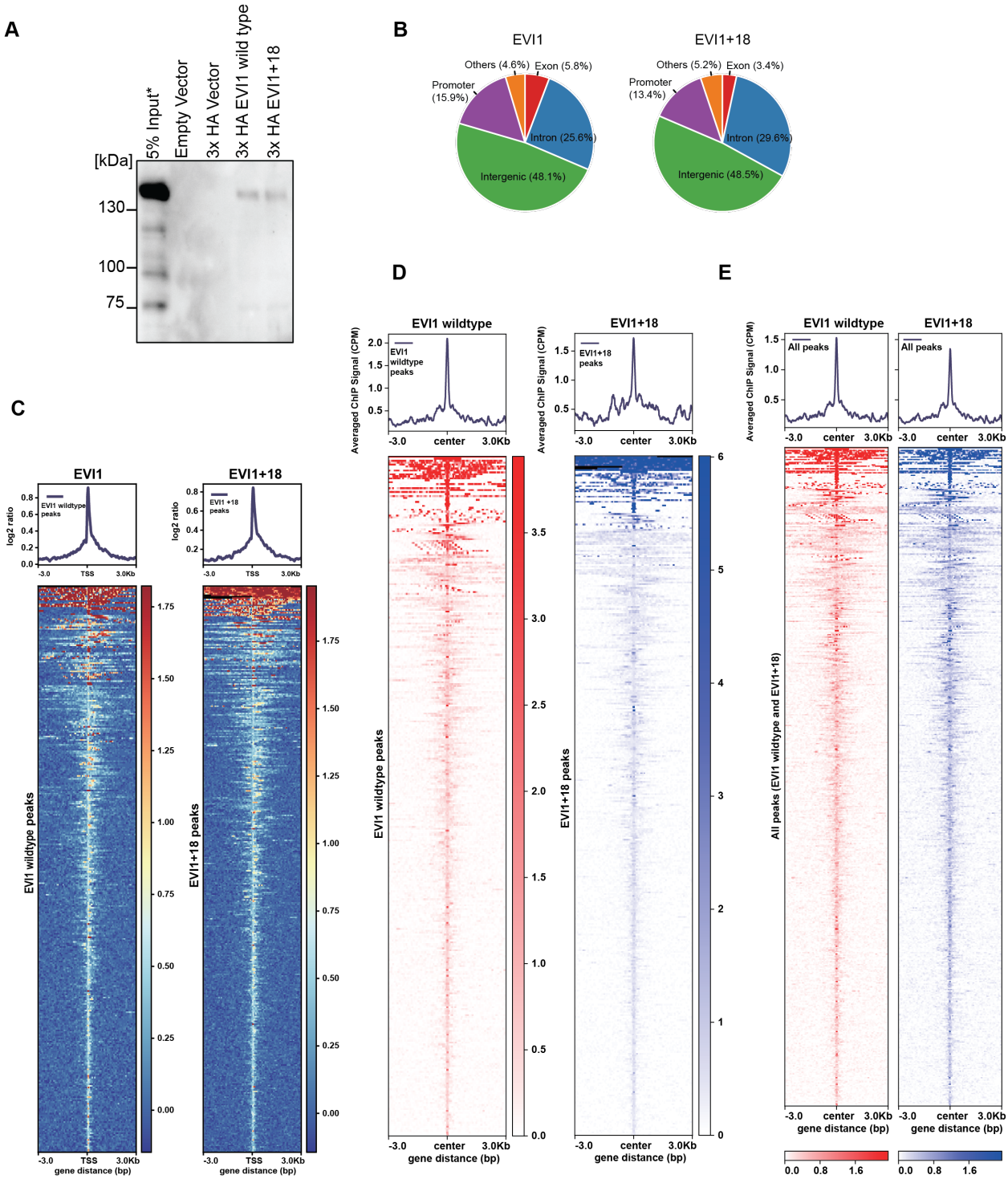
## Supplementary Figure 7



## Supplementary Figure 8



## Supplementary Figure 9



## Supplementary Figure 10