# A novel genetic and morphologic phenotype of *ARID2*-mediated myelodysplasia.

Hitoshi Sakai,[1,2*] Naoko Hosono,[1*] Hideyuki Nakazawa,[2] Bartlomiej Przychodzen,[1] Chantana Polprasert,[1] Hetty E. Carraway,[1] Mikkael A Sekeres,[1,3] Tomas Radivoyevitch,[1] Kenichi Yoshida,[4] Masashi Sanada,[5] Tetsuichi Yoshizato,[4] Keisuke Kataoka,[4] Masahiro M. Nakagawa,[4] Hiroo Ueno,[4] Yasuhito Nannya,[4] Ayana Kon,[4] Yusuke Shiozawa,[4] June Takeda,[4] Yuichi Shiraishi,[6] Kenichi Chiba,[6] Satoru Miyano,[6] Jarnail Singh,[7] Richard A. Padgett,[7] Seishi Ogawa,[4] Jaroslaw P. Maciejewski,[1] Hideki Makishima.[1,4]

1. Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA. 2. Division of Hematology, Department of Internal Medicine, Shinshu University, Matsumoto, Japan. 3. Leukemia Program, Department of Hematology and Medical Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA. 4. Department of Pathology and Tumor Biology, Kyoto University, Kyoto, and Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. 5. Clinical Research Center, Nagoya Medical Center, Nagoya, Japan. 6. Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. 7. Department of Molecular Genetics, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA.

Corresponding should be addressed to:
Hideki Makishima, MD., Ph.D., F.A.C.P.
Pathology and Tumor Biology, University of Kyoto, Yoshida-Konoe-cho, Sakyo-ku, Kyoto, 606-8501, Japan. Phone: +81-75-753-9285
E-mail: makishima.hideki.8x@kyoto-u.ac.jp.

No conflict of interest to disclose.

* These authors equally contributed to this study.

Running title: A distinct MDS subtype with *ARID2* deficiency.

**Abstract**

Clinical heterogeneity of myelodysplastic syndromes (MDS) and related myeloid neoplasms reflects molecular diversity. Most common genetic associations with distinct clinical or pathomorphologic phenotypes have been already reported, but many other common somatic lesions exist and their clinical contexts still remain elusive. By comprehensive genetic investigation of 1,473 cases with various myeloid neoplasms, we characterized here cases in which lesions mediate their clinical effects in MDS and other myeloid neoplasms via decreased expression of *ARID2*. We showed that insufficient *ARID2* expression mainly in MDS arose from *ARID2* mutations and deletions that yielded defective *ARID2* transcripts. Clonal architecture analyses showed that *ARID2* mutations and deletions occurred as initial events of MDS or myelodysplastic/myeloproliferative neoplasms, and not during progression to acute myeloid leukemia. Morphologically, progressive maturation in myeloid and erythroid lineages and hypolobated megakaryocytes were common in cases with *ARID2* mutations and deletions. Functionally, *in vitro* knockdown of *ARID2* expression led to myeloid and erythroid differentiation in hematopoietic cell lines. Missplicing of *ARID2* was identified in *U2AF1* mutant cases, resultant in a low expression of *ARID2* mRNA. In conclusion, *ARID2* is a MDS-suppressor gene whose expression is attenuated as it shapes the distinct morphological phenotype of a subset of myelodysplasia.

## INTRODUCTION

Heterogeneity of myelodysplastic syndromes (MDS) and related neoplasms[1-3] is a reflection of inherent diversity of molecular lesions and their combinations,[4, 5] which likely determine morphological and clinical features and/or the pace of evolution into secondary acute myeloid leukemia (sAML).[6, 7] Next generation sequencing and single nucleotide polymorphism arrays (SNP-A) have yielded increasingly defined spectra of somatic mutations and a growing understanding of their roles in pathogenesis and genotype-phenotype relationships.[8-12] Prominent examples include increased ring sideroblasts in cases with *SF3B1* mutations,[13, 14] proliferative monocytes with *SRSF2* mutations,[15, 16] and 5q- syndromes with recurrent *CSNK1A1* mutations.[17, 18] Similarly, in advanced MDS phenotypes with increasing blasts, one group of gene mutations were associated with sAML and another with high-risk MDS.[19] As such discoveries inch toward completion,[19-21] identification of the remaining relationships between driver mutations and clinical phenotypes may require resolution gains obtained by using multiple sources of information. Here we use mechanistic understanding to integrate data sources into our genetic analyses and to discover new disease mechanisms and principles.[19, 22]

AT rich interactive domain 2 (*ARID2*), which is located on chromosome 12q12, encodes a component of the SWI/SNF complex that is involved in chromatin remodeling.[23, 24] Among various SWI/SNF genes, multiple groups recently detected *ARID2* mutations or deletions in solid tumors, including hepatocellular carcinoma (HCC),[25] melanoma,[26] malignant mesothelioma,[27] urethral clear-cell adenocarcinoma,[28] and non-small cell lung carcinoma.[29] *ARID2* mutations were exclusively frameshift or nonsense in HCC and melanoma,[25, 26] and *ARID2* deletions were recurrently identified in urethral clear-cell adenocarcinoma[28] and non-small cell lung carcinoma.[29] These reports suggest that *ARID2* is a global tumor suppressor gene in various cancers. With regard to hematological neoplasms, *SMARCA4* (*BRG1*), which is a core component of both SWI/SNF-A and B, is essential for maintenance of stemness of AML.[30, 31] In addition, mutations of *ARID1A* and *ARID1B*, components of SWI/SNF-A, were identified in cases with acute promyelocytic leukemia (APL), in which *ARID1B* deficiency causes a block in differentiation.[32] These previous discoveries prompted us to investigate genetic defects in SWI/SNF components in MDS and its related neoplasms. Here, we report whole exome sequencing-guided identification of novel *ARID2* mutations in myeloid neoplasms. Our comprehensive analysis extends an approach we developed and applied successfully to *EZH2*.[33, 34] In addition to copy number analysis and targeted deep and exome sequencing, here we include RNA sequencing, for splicing analyses of the roles of spliceosomal mutations[35] in *ARID2*

missplicing and gene expression.

In this study, we aimed to elucidate the distinct impact of *ARID2* defects on myeloid neoplasms. Even though somatic mutations of this gene were relatively rare, integrated analysis of multiple genetic methodologies allowed us to find other mechanisms of *ARID2* deficiency and to shed light on a novel morphological phenotype of MDS associated with loss of *ARID2* function. Somatic mutations, deletions, and low expression may converge as alternate mechanisms of *ARID2* inactivation.

## MATERIALS AND METHODS

### Patient population

Bone marrow aspirates or blood samples were collected from 1,473 patients with MDS (n=455), myelodysplastic/myeloproliferative neoplasms (MDS/MPN) (n=201), myeloproliferative neoplasms (MPN) (n=56), sAML (n=221), and *de novo* AML (n=540) at the Cleveland Clinic and The University of Tokyo; the registered data at The Cancer Genome Atlas (TCGA) was also included (**Supplementary Table S1**). Diagnoses were classified according to World Health Organization (WHO) criteria. Myelodysplasia in bone marrow of each case was evaluated by multiple pathologists. MDS with excess blasts (MDS-EB)-1 and MDS-EB-2 were assigned to high-risk MDS, and MDS with single lineage dysplasia (MDS-SLD), MDS with multilineage dysplasia (MDS-MLD), MDS with ring sideroblasts (MDS-RS), 5q- syndrome, and MDS unclassifiable (MDS-U) to low-risk MDS (**Supplementary Table S1**). To study the germ line genotype as the control, immunoselected CD3+ T cells from each patient were analyzed. Cytogenetic analyses were performed using standard banding techniques on 20 metaphases. Clinical parameters, including age, sex, bone marrow blast counts, and clinical outcome, were collected. Informed consent for sample collection was obtained according to a protocol approved by each Institutional Review Board in accordance with the Declaration of Helsinki.

### Single-nucleotide polymorphism array (SNP-A)

SNP-A samples were processed as previously described.[36, 37] Briefly, karyotyping was performed using Affymetrix 250K and 6.0 SNP arrays (Affymetrix, Santa Clara, CA, USA). A stringent algorithm was used to identify SNP-A lesions. Patients with SNP-A lesions concordant with metaphase cytogenetics or typical lesions known to be recurrent required no further validation. Changes reported in our internal or publicly-available copy number variation databases were considered non-somatic and excluded.[37] Results were analyzed using CNAG (v3.0) or Genotyping Console (Affymetrix). All other lesions were deemed either somatic or germline in comparative

analysis of CD3+ T cells.

**Whole exome sequencing**

Whole exome sequencing was performed according to the manufacturer's protocol, as previously described.[15, 19] In short, tumor DNAs were obtained from patients' bone marrow or peripheral blood mononuclear cells. For germline controls, DNA was extracted from immunoselected CD3+ T cells. Using liquid phase hybridization of sonicated genomic DNA having 150-200bp of mean length, the exomes attached to the bait cRNA library synthesized on magnetic beads (SureSelect ver.3 or 4, Agilent Technology, Santa Clara, CA, USA) were captured. The obtained targets were subjected to massive parallel sequencing using Illumina HiSeq 2000 with the pair end 75-108bp read option. The raw sequence data were processed through the in-house pipeline constructed for whole exome analysis of paired cancer genomes at the Human Genome Center, Institute of Medical Science, The University of Tokyo, as summarized previously.[15, 19] Generation of bam files with its preprocessing and detection of somatic point mutations or insertions and deletions was done as previously described.[15] In addition, for detailed analyses, exome sequencing data (n=189) on AML patients obtained through TCGA data portal (http://cancergenome.nih.gov/) were used.

**Re-sequencing**

To validate mutations, Sanger sequencing and amplicon deep sequencing were applied as previously described (**Supplementary Table S2**).[19, 38] Using an Illumina MiSeq sequencer, deep sequencing was performed for the measurement of variant allele frequency (VAF). To confirm the somatic nature of the mutations, DNA derived from CD3+ T cells was also subjected to re-sequencing.[15, 39]

**RNA sequencing and confirmatory RT-PCR**

To analyze *ARID2* exon splicing patterns as a function of *U2AF1* mutation status, RNA sequencing was performed as previously described.[35] Briefly, total RNA was extracted from bone marrow mononuclear cells and 1–2 mg of cDNA was generated from 100 ng of total RNA. cDNA was fragmented and subjected to Illumina library preparation. After the quality and quantity and the size distribution of the Illumina libraries were determined using an Agilent Bioanalyzer, the libraries were then submitted to Illumina HiSeq2000 sequencing using standard operating procedures. Paired-end 90-bp reads were generated and subjected to data analysis using the platform provided by DNAnexus (Mountain View, CA, USA; https:// dnanexus.com), which allowed visualization of reads derived from spliced mRNA and those that completely match the

genome, including both sense and antisense. To identify significantly enriched pathways, Gene Set Enrichment Analysis (GSEA)[40] was performed using the MSigDB-curated gene sets. Confirmatory RT-PCR for abnormal splicing (missplicing) of *ARID2* exon 8 was applied to *U2AF1* mutated (p.S34F and p.Q157P) and wild-type cases and also to healthy donors. Normal transcripts with expected intron splicing (348 base pairs [bps]) and shorter variants (180 bps and 97 bps) due to missplicing were subjected to direct Sanger sequencing or subcloning to confirm newly selected splice sites.

**Quantitative RT-PCR**

Total RNA was extracted from bone marrow mononuclear cells. cDNA was synthesized from 500ng total RNA using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, USA). Quantitative gene expression levels were detected using real-time PCR with the ABI PRISM 7500 Fast Sequence Detection System and FAM dye labeled TaqMan MGB probes (Applied Biosystems). TaqMan assay was performed according to the manufacturer's instructions. The expression level of target genes was normalized to *GAPDH* mRNA.

**Cell culture and knockdown gene expression**

HL60 and K562 cell lines were cultured using IMDM with 10% fetal bovine serum (FBS). Two independent pLKO.1_*ARID2*-shRNA (TRCN0000166359 and TRCN0000166264) and the control non-target shRNA were purchased from Sigma-Aldrich. In brief, 293T cells were transfected with shRNA targeting *ARID2* or non-target shRNA control plasmid together with packing plasmid pCMVdR8.2 and envelope plasmid containing VSV-G. Viral supernatants were harvested at 48, 72, and 96 hours posttransfection, and target cells were infected in the presence of 8 mg/ml Polybrene for 24 hours and selected with puromycin for K562 and HL60. Cell lines were tested for mycoplasma contamination and found negative before used for experiments.

**Proliferation assay**

Cell growth was determined by counting cell numbers in culture. Briefly, 1 mL of 5 x $10^4$ cells were cultured in IMDM containing 10% FBS in the six-well plates at day 0, and the cell numbers were scored by Trypan blue exclusion at 24, 48, 72, and 96 hours.

**Apoptosis assay**

For analysis of apoptosis, the cells were stained with APC-conjugated anti-AnnexinV and propidium iodide (eBioscience, catalog no. 88-8007-74) as per the manufacturer's

protocol. Sample analysis was performed on a flow cytometer (Beckman Coulter FC500).

**Database and URL**

Validation of expression array, SNP-A, RNA sequencing, and whole exome sequencing results was sought using TCGA (http://cancergenome.nih.gov/) and Oncomine (https://www.oncomine.org/). Analytic tools were obtained from URLs such as GSEA (http://www.broadinstitute.org/gsea/) and Integrative Genomics Viewer (IGV), (https://www.broadinstitute.org/igv/). Sequencing data were deposited to SRA (PRJNA203580) ( https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA203580).

**Statistical analysis**

Pairwise comparisons were performed by Wilcoxon test for continuous variables and by two-sided Fisher's exact test for categorical variables. Whisker plot boxes display medians, 25th and 75th percentiles, and minimum and maximum values (whisker ends). The Kaplan-Meier method was used to analyze overall survival by the log-rank test. Statistical analyses were performed with R (https://www.r-project.org) or JMP9 software (SAS). Significance was determined at a two-sided α level of 0.05, except for P values in multiple comparisons, in which multiple testing was adjusted using the Benjamini and Hochberg method.[42]

**RESULTS**

**Loss-of-function mutations of *ARID2.***

We examined *ARID2* (NM_152641; 1,835 amino acids) mutational status of a cohort of patients (n=393) with various myeloid malignancies, including low-risk (n=58) and high-risk (n=35) MDS, MDS/MPN (n=37), MPN (n=21), sAML (n=33), and *de novo* AML (n=209) (**Supplementary Table S1**) by whole exome sequencing and resequencing. *ARID2* mutations were identified in 6 cases with MDS (n=2), MDS/MPN (n=1), MPN (n=1), and *de novo* AML (n=2), (**Table 1**). All mutations were confirmed to be somatic using germ line DNA (**Figure 1a**). Among the 6 mutations, 5 (83%) were nonsense or frame shift, suggesting that loss of *ARID2* function is pertinent to myeloid neoplasms. Amplicon deep sequencing and SNP-A karyotyping revealed that all of the mutations were heterozygous (**Figure 1a, b**). No homozygous or hemizygous mutations were identified. Although some cases showed mutations in germ line DNA, the somatic nature of all tested mutations was supported by significantly higher ratios of mutant sequence reads in tumor samples than in paired CD3+ T cells; the small fraction of

mutant DNA in CD3+ T cells may also reflect *ARID2* mutations arising in a hematopoietic stem cell capable of some differentiation into CD3+ T cells. In the tumor fraction, VAF of *ARID2* mutations were all very high (27-48%, median=43%), suggesting *ARID2* mutations might be relatively early events (**Figure 1b**). In 11 cell lines of myeloid leukemia, *ARID2* mutations were not identified (**Supplementary Table S3**).

**Haploinsufficiency of *ARID2.***

Frequent identification of loss-of-function mutations in *ARID2* prompted us to investigate copy number variations of this gene locus. For this purpose, a separate cohort of patients (n=1,080) with myeloid malignancies, including MDS (n=362), MDS/MPN (n=164), MPN (n=35), sAML (n=188), and *de novo* AML (n=331) was subjected to SNP-A karyotyping (**Supplementary Table S1**). Copy number losses (deletions) of *ARID2* locus (chr12q12) were identified in 8 cases with MDS (n=5), sAML (n=1), and *de novo* AML (n=2), whereas focal gain lesions were not detected (**Figure 1a, Table 1**). Relative expression of *ARID2* evaluated by quantitative RT-PCR showed lower expression in cases with *vs.* without deletions, suggesting haploinsufficiency (**Figure 1c**). Taken together, in our cohort, *ARID2* defects were present in approximately 1% of cases (14 out of 1,473; **Table 1**).

**Genetic defects associated with *ARID2* defects.**

According to our findings, we then comprehensively investigated *ARID2*-mediated myeloid neoplasms by clarifying additional genetic events in cases with *ARID2* defects. Most common additional genetic events were mutations and deletions in genes harboring histone repressive mark, including *EZH2* and *JARID2*, which were identified in 43% (6/14) of cases (**Figure 2a, Supplementary Figure S1a**), suggesting synergetic effects of histone modification together with defective chromatin remodeling. Receptor tyrosine kinase pathway and RAS pathway genes were affected by mutations as secondary events in cases with *ARID2*-mediated myeloid neoplasms (**Figure 2b**). Mutations and deletions of *TP53* (n=5), *DNMT3A* (n=2), and *TET2* (n=1) were also identified. While 3 (21%) cases were cytogenetically normal, complex karyotypes and trisomy 8 were found in 7 and 2 cases, respectively (**Figure 2a**).

**Clonal architecture of *ARID2* defects.**

In 5 cases with *ARID2* defects, serial sample sequencing or SNP-A analysis was performed to investigate clonal architecture. According to mutated clone size and cell fraction, *ARID2* mutations and deletions tended to be acquired as early events followed by other mutations and copy number abnormalities (**Figure 2b and Supplementary**

**Figure S1b**). These findings suggest that defective *ARID2* is likely to be an initial clonal event.

**Clinical phenotype associated with *ARID2* defects.**

We next investigated genotype-phenotype associations of defective *ARID2*. Relative expression of wild-type *ARID2* was significantly lower in MDS than in healthy donors (*P*=0.02; **Figure 3a**). A mean value of *ARID2* relative expression showed 25% reduction in patients with MDS (95% confidence interval (CI); 0.67-0.80) compared to that in healthy donors (95% CI; 0.82-1.12). These findings suggest that loss of *ARID2* function plays an important role in the early stage of MDS, compatible with our results that *ARID2* genetic defects are in fact initially acquired as clonal events (**Figure 2b**). We next assessed the impact of *ARID2* deficiency on clinical outcomes. *ARID2* mutations, deletions, and less expression had no effect on overall survival (**Supplementary Figure S2**). We could not find any correlation between low expression of *ARID2* and any other clinical parameters studied, including age, sex, counts of white blood cells, neutrophils, and platelets, hemoglobin levels, and bone marrow blast percentage (data not shown).

**Megakaryocytic dysplasia associated with *ARID2* deficiency.**

To further assess the features of *ARID2* deficiency, an independent pathologist examined bone marrow cell morphologies of the cases with *ARID2* mutation and deletion (n=8). In addition to the megaloblastoid erythorid and hypogranular myeloid cells, progressive maturation was identified in myeloid and erythroid lineages. Most prominent findings were megakaryocytic dysplasias, which were seen in all of 8 cases with *ARID2* defects, including hypolobated forms in 7 cases (**Table 2, Figure 3b**).

**Functional significance of *ARID2* deficiency.**

To investigate the functional significance of *ARID2* deficiency in myelodysplasia, we utilized *in vitro* knockdown models of *ARID2* expression in hematopoietic cell lines and bone marrow mononuclear cells. No homozygous deletion or mutation of *ARID2* was identified (**Figure 1, Table 1**), and relative *ARID2* expression is moderately reduced in MDS (**Figure 3a**). Accordingly, we transduced shRNA in neoplastic hematopoietic cells to obtain disease models with partial reduction of *ARID2* expression (**Supplementary Figure S3**). Two myeloid cell lines (HL60 and K562) in which *ARID2* expression was knocked down showed significantly lower cell counts compared to those with normal *ARID2* expression (**Figure 4a, b, Supplementary Figure S4**), compatible with more apoptotic cells in knockdown experiments (**Figure 4c**). Flow cytometric

analysis of the cell lines with reduced *ARID2* expression revealed increased cell-surface maturation markers, CD11b and glycophorin A (GPA), suggesting that reduced expression of *ARID2* resulted in more differentiation in myeloid and erythroid lineages (**Figure 4d, e**). These results indicate that reduced *ARID2* expression might induce more differentiation in myeloid/erythroid lineages and more apoptosis to reduce cell populations and reduction of proliferation capacity.

**Enrichment of gene pathways associated with *ARID2* deficiency.**

To investigate the effects of *ARID2* defects on transcription, "Hallmark gene sets" were used for GSEA of cases with *ARID2* low expression (n=58) compared to those with normal expression (n=58). Hallmark gene sets downloaded from Broad institute include specific well-defined biological states or processes which condense information from over 4,000 original overlapping gene sets from v4.0 The Molecular Signatures Database (MSigDB).[7] To assess the self-renewal and differentiation potential of defective *ARID2*, we applied GSEA using gene sets expressing in hematopoietic stem cells and in myeloid differentiated cells which were downloaded from MSigDB. A gene set of 305 genes involved in maintenance of hematopoietic stem cells was negatively enriched in the low-expression cases (NES=-1.79, P=0.0076). The other gene set expressing in myeloid differentiated cells (49 genes) was positively enriched (NES=1.77, P=0.0041) (**Supplementary Figure S5**). Accordingly, *ARID2* defects are most likely to induce myeloid differentiation rather than stem cell potential. Overall, GSEA clearly confirmed functional implications of *ARID2*-knockdown experiments in hematopoietic cells.

**Missplicing of *ARID2* defects.**

Previously, we and other groups reported abnormal splicing (missplicing) due to spliceosomal mutations[33, 34, 43, 44]. *U2AF1* mutations associate with spliced-out exons (cassette exons) with specific motifs in 3' splice sites,[35, 45, 46] and such a motif is present in the 3' splice site of *ARID2* exon 8 (**Supplementary Figure S6**). We therefore investigated the transcriptional patterns of this exon, in search of abnormalities. According to Sashimi plots created by applying IGV software to our RNA sequencing data, either partial (180bp) or complete (97bp) skipping of exon 8 was more evident in the case with *U2AF1* mutation than with wild-type *U2AF1* (**Figure 5a, Supplementary Figure S7**). RT-PCR using multiple mutant and wild-type cases revealed 2 types of abnormal transcripts that were compatible with our RNA sequencing results and confirmed by direct sequencing (**Figure 5b**). The longer abnormal transcript due to partial skipping of *ARID2* exon 8 had lost the conserved LXXLL motif, which is predicted

to be involved in binding to a nuclear receptor. The shorter transcript variant with complete skipping of exon 8, which was identified only in cases with a *U2AF1* mutation, resulted in frameshift sequences at a glutamic acid residue at position 258 (p.E258fs) of *ARID2*. Relative expression of *ARID2* was lower in the cases with *U2AF1* mutations than in individuals with wild-type *U2AF1* (**Figure 5b**). Regarding genetic association between *ARID2* and *U2AF1*, *ARID2* defects and *U2AF1* mutations were mutually exclusive except for one case. However, the exclusivity was not statistically significant due to low frequency of *ARID2* mutations and deletions.

## DISCUSSION

Our comprehensive study by whole exome sequencing, RNA sequencing, and SNP-A, elucidated a role for a novel tumor suppressor gene, *ARID2*, in myeloid neoplasms. Loss of *ARID2* function due to its genomic defects is frequently acquired in early stage of myelodysplasia. Intriguingly, *ARID2* deficiency represents differentiation defect in megakaryocytes which results in hypolobated forms, while differentiation is remarkable in myeloid and erythroid lineages with *ARID2* deficiency. Functional analysis using *in vitro* models of *ARID2* knockdown confirmed such genetic and pathologic phenotypes with findings of increased myeloid and erythroid differentiation. Overall, these common features are shared by various genetic defects leading to *ARID2* deficiency. To overcome survival disadvantage, cells losing *ARID2* function likely require additional hits, and indeed, many additional alterations were identified in *ARID2*-mutant cases, suggesting that some of them may play an important role in the clonal advantage of *ARID2*-mediated myelodysplastic phenotypes.

*ARID2* encodes a subunit of SWI/SNF-B (PBAF) complex which is involved in chromatin remodeling and transcription regulation. Together with the other SWI/SNF complex, SWI/SNF-A (BAF), the SWI/SNF system is essential for transcriptional activation and repression in mammalian cells[47]. While homeostasis of this cellular machinery is strictly regulated for normal physiology, various deregulations of SWI/SNF due to genetic events can cause cancers in many tissues[23]. According to previous studies of AML, *SMARCA4* is required for self-renewal capacity[30, 31]. Most recently, loss-of-function of *ARID1A* or *ARID1B*, was reported in APL, in which *ARID1B* deficiency causes a block in differentiation[32]. Despite the lack of *ARID2* mutations in APL[32], we now identified multiple mutations of this gene in MDS and MDS/MPN cases. In contrast, *ARID1A* and *ARID1B* mutations were not identified in MDS or MDS/MPN. These findings suggest that mutations of the different components in SWI/SNF-A and -B complexes result in the presentation of distinct phenotypes with different functional consequences probably depending on downstream transcriptional targets. In fact,

candidate downstream pathways associated with defective *ARID2* were extracted by gene enrichment analysis in this study. For example, more enrichment of myeloid differentiation pathway genes and less enrichment of self-renewal pathway genes are compatible to both cytopenia in initial MDS presentation and cell-count reduction in functional *ARID2*-knockdown analysis.

To regulate differentiation genes, SWI/SNF complexes work with various histone modification enzymes, including methyltransferases and acetyltransferases[48, 49]. Among these, polycomb repressive complex 2 (PRC2) is involved in trimethylation of histone H3 lysine 27 (H3K27), chromatin condensation, and transcriptional repression. Genes for components of PRC2 were frequently affected by loss-of-function mutations and deletions in *ARID2*-deficient MDS cases. In contrast, PRC2 is frequently activated in solid tumors[23]. Accordingly, correlation between *ARID2* deficiency and PRC2 dysfunction in myelodysplasia is likely tissue-specific. Further studies of expression profiles associated with unique patterns of histone modification and DNA remodeling will be needed to clarify which target genes are responsible for their synergistic effects in MDS genesis.

While *U2AF1* mutations are frequently observed in MDS, misspliced gene targets have yet to be identified[45, 50-52]. In this study, we showed MDS genesis due to *U2AF1* mutations through cassette exon 8 of *ARID2*. Indeed, a report of less colony formation in a *U2AF1* mutant model[15] was consistent with less proliferative potential due to *ARID2* deficiency identified in this study. This supports the view that *ARID2* might be a candidate gene in *U2AF1*-mediated MDS genesis.

Our main conclusion is that *ARID2* defects attenuated by multiple mechanisms are early events in a subset of MDS cases characterized by distinct morphological features.

**Acknowledgments**

This work was supported by National Institutes of Health (Bethesda, MD; NIH) grants RO1HL-082983 (J.P.M.), U54 RR019391 (J.P.M.), K24 HL-077522 (J.P.M.), a grant from the AA & MDS International Foundation (Rockville, MD; J.P.M. and Cleveland, OH; H.M.), the Robert Duggan Charitable Fund (Cleveland, OH; J.P.M.), a grant from Edward P. Evans Foundation (Cleveland, OH; J.P.M and M.A.S), Scott Hamilton CARES grant (Cleveland, OH; H.M.), Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan and KAKENHI (23249052, 22134006, and 21790907) (Tokyo; S.O.), (15km0305018h0101, 16km0405110h0004, 16H05338) (Kyoto; H.M.), project for development of innovative research on cancer therapies (p-direct) (Tokyo; S.O.), the Japan Society for the Promotion of Science (JSPS) through the 'Funding Program for World-Leading Innovative R&D on Science and Technology', initiated by the Council for

Science and Technology Policy (CSTP) (Tokyo; S.O.), project for cancer research and therapeutics evolution (P-CREATE) from Japan Agency for Medical Research and Development (16cm0106501h0001 (Kyoto; S.O.) and 16cm0106422h0001 (Kyoto; T.Y.), National Aeronautics and Space Administration grant NNJ13ZSA001N (Cleveland, OH, T.R.), and a grant from The YASUDA Medical Foundation (Kyoto; H.M.). The results published here were partly based on data generated by The Cancer Genome Atlas pilot project established by the National Cancer Institute and National Human Genome Research Institute.

## Conflict of Interest

No conflict of interest to disclose.

## References

1. Corey SJ, Minden MD, Barber DL, Kantarjian H, Wang JC, Schimmer AD. Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nat Rev Cancer* 2007; **7:** 118-129.
2. Bejar R, Stevenson KE, Caughey BA, Abdel-Wahab O, Steensma DP, Galili N*, et al.* Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. *J Clin Oncol* 2011; **30:** 3376-3382.
3. Della Porta MG, Travaglino E, Boveri E, Ponzoni M, Malcovati L, Papaemmanuil E*, et al.* Minimal morphological criteria for defining bone marrow dysplasia: a basis for clinical implementation of WHO classification of myelodysplastic syndromes. *Leukemia* 2015; **29:** 66-75.
4. Hosono N, Makishima H, Jerez A, Yoshida K, Przychodzen B, McMahon S*, et al.* Recurrent genetic defects on chromosome 7q in myeloid neoplasms. *Leukemia* 2014; **28:** 1348-1351.
5. Jerez A, Gondek LP, Jankowska AM, Makishima H, Przychodzen B, Tiu RV*, et al.* Topography, clinical, and genomic correlates of 5q myeloid malignancies revisited. *J Clin Oncol* 2012; **30:** 1343-1349.
6. Heaney ML, Golde DW. Myelodysplasia. *N Engl J Med* 1999; **340:** 1649-1660.
7. Tefferi A, Vainchenker W. Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. *J Clin Oncol* 2011; **29:** 573-582.
8. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G*, et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* 2011; **364:** 2496-2506.
9. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult

de novo acute myeloid leukemia. *N Engl J Med* 2013; **368:** 2059-2074.

10. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND*, et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016; **374:** 2209-2221.
11. Rose D, Haferlach T, Schnittger S, Perglerova K, Kern W, Haferlach C. Subtype-specific patterns of molecular mutations in acute myeloid leukemia. *Leukemia* 2017; **31:** 11-17.
12. Tiu RV, Gondek LP, O'Keefe CL, Elson P, Huh J, Mohamedali A*, et al.* Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies. *Blood* 2011; **117:** 4552-4560.
13. Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D*, et al.* Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* 2011; **365:** 1384-1395.
14. Visconte V, Makishima H, Jankowska A, Szpurka H, Traina F, Jerez A*, et al.* SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia* 2012; **26:** 542-545.
15. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R*, et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 2011; **478:** 64-69.
16. Meggendorfer M, Roller A, Haferlach T, Eder C, Dicker F, Grossmann V*, et al.* SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood* 2012; **120:** 3080-3088.
17. Schneider RK, Adema V, Heckl D, Jaras M, Mallo M, Lord AM*, et al.* Role of Casein Kinase 1A1 in the Biology and Targeted Therapy of del(5q) MDS. *Cancer Cell* 2014; **26:** 509-520.
18. Heuser M, Meggendorfer M, Cruz MM, Fabisch J, Klesse S, Kohler L*, et al.* Frequency and prognostic impact of casein kinase 1A1 mutations in MDS patients with deletion of chromosome 5q. *Leukemia* 2015; **29:** 1942-1945.
19. Makishima H, Yoshizato T, Yoshida K, Sekeres MA, Radivoyevitch T, Suzuki H*, et al.* Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet* 2017; **49:** 204-212.
20. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G*, et al.* Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 2014; **28:** 241-247.
21. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P*, et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013; **122:** 3616-3627.
22. Sallman DA, Komrokji R, Vaupel C, Cluzeau T, Geyer SM, McGraw KL*, et al.* Impact of TP53 mutation variant allele frequency on phenotype and outcomes in myelodysplastic

syndromes. *Leukemia* 2016; **30:** 666-673.

23. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* 2011; **11:** 481-492.

24. Masliah-Planchon J, Bieche I, Guinebretiere JM, Bourdeaut F, Delattre O. SWI/SNF chromatin remodeling and human malignancies. *Annu Rev Pathol* 2015; **10:** 145-171.

25. Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA*, et al.* Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat Genet* 2011; **43:** 828-829.

26. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP*, et al.* A landscape of driver mutations in melanoma. *Cell* 2012; **150:** 251-263.

27. Yoshikawa Y, Sato A, Tsujimura T, Otsuki T, Fukuoka K, Hasegawa S*, et al.* Biallelic germline and somatic mutations in malignant mesothelioma: multiple mutations in transcription regulators including mSWI/SNF genes. *Int J Cancer* 2015; **136:** 560-571.

28. Mehra R, Vats P, Kalyana-Sundaram S, Udager AM, Roh M, Alva A*, et al.* Primary urethral clear-cell adenocarcinoma: comprehensive analysis by surgical pathology, cytopathology, and next-generation sequencing. *Am J Pathol* 2014; **184:** 584-591.

29. Manceau G, Letouze E, Guichard C, Didelot A, Cazes A, Corte H*, et al.* Recurrent inactivating mutations of ARID2 in non-small cell lung carcinoma. *Int J Cancer* 2013; **132:** 2217-2221.

30. Buscarlet M, Krasteva V, Ho L, Simon C, Hebert J, Wilhelm B*, et al.* Essential role of BRG, the ATPase subunit of BAF chromatin remodeling complexes, in leukemia maintenance. *Blood* 2014; **123:** 1720-1728.

31. Shi J, Whyte WA, Zepeda-Mendoza CJ, Milazzo JP, Shen C, Roe JS*, et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev* 2013; **27:** 2648-2662.

32. Madan V, Shyamsunder P, Han L, Mayakonda A, Nagata Y, Sundaresan J*, et al.* Comprehensive mutational analysis of primary and relapse acute promyelocytic leukemia. *Leukemia* 2016; **30:** 1672-1681.

33. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SC, Ramakrishnan A*, et al.* SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell* 2015; **27:** 617-630.

34. Khan SN, Jankowska AM, Mahfouz R, Dunbar AJ, Sugimoto Y, Hosono N*, et al.* Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. *Leukemia* 2013; **27:** 1301-1309.

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

539 36. Tiu R, Gondek L, O'Keefe C, Maciejewski JP. Clonality of the stem cell compartment
540 during evolution of myelodysplastic syndromes and other bone marrow failure
541 syndromes. *Leukemia* 2007; **21:** 1648-1657.
542 37. Tiu RV, Gondek LP, O'Keefe CL, Huh J, Sekeres MA, Elson P*, et al.* New lesions
543 detected by single nucleotide polymorphism array-based chromosomal analysis have
544 important clinical impact in acute myeloid leukemia. *J Clin Oncol* 2009; **27:** 5219-5226.
545 38. Huang D, Nagata Y, Grossmann V, Radivoyevitch T, Okuno Y, Nagae G*, et al.* BRCC3
546 mutations in myeloid neoplasms. *Haematologica* 2015.
547 39. Makishima H, Yoshida K, Nguyen N, Przychodzen B, Sanada M, Okuno Y*, et al.* Somatic
548 SETBP1 mutations in myeloid malignancies. *Nat Genet* 2013; **45:** 942-946.
549 40. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J*, et al.*
550 PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately
551 downregulated in human diabetes. *Nat Genet* 2003; **34:** 267-273.
552 41. Polprasert C, Schulze I, Sekeres MA, Makishima H, Przychodzen B, Hosono N*, et al.*
553 Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell* 2015; **27:**
554 658-670.
555 42. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful
556 approach to multiple testing. *J R Stat Soc B* 1995; **57:** 289–300.
557 43. Dolatshad H, Pellagatti A, Liberante FG, Llorian M, Repapi E, Steeples V*, et al.* Cryptic
558 splicing events in the iron transporter ABCB7 and other key target genes in
559 SF3B1-mutant myelodysplastic syndromes. *Leukemia* 2016; **30:** 2322-2331.
560 44. Mupo A, Seiler M, Sathiaseelan V, Pance A, Yang Y, Agrawal AA*, et al.*
561 Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in
562 human MDS but develop anemia without ring sideroblasts. *Leukemia* 2016.
563 45. Shirai CL, Ley JN, White BS, Kim S, Tibbitts J, Shao J*, et al.* Mutant U2AF1 Expression
564 Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell* 2015; **27:** 631-643.
565 46. Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y*, et al.* Mechanisms for U2AF to define
566 3' splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol*
567 2014; **21:** 997-1005.
568 47. Roberts CW, Orkin SH. The SWI/SNF complex--chromatin and cancer. *Nat Rev Cancer*
569 2004; **4:** 133-142.
570 48. Neely KE, Workman JL. Histone acetylation and chromatin remodeling: which comes
571 first? *Mol Genet Metab* 2002; **76:** 1-5.
572 49. Stanton BZ, Hodges C, Calarco JP, Braun SM, Ku WL, Kadoch C*, et al.* Smarca4
573 ATPase mutations disrupt direct eviction of PRC1 from chromatin. *Nat Genet* 2016.
574 50. Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J*, et al.* Recurrent
575 mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet* 2011;

**44:** 53-57.

51. Okeyo-Owuor T, White BS, Chatrikhi R, Mohan DR, Kim S, Griffith M*, et al.* U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia* 2015; **29:** 909-917.

52. Fei DL, Motowski H, Chatrikhi R, Prasad S, Yu J, Gao S*, et al.* Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival. *PLoS Genet* 2016; **12:** e1006384.

**Figure legends**

**Figure 1. Identification of deletions and mutations in *ARID2*.**

a. Upper panels: a SNP-A chromosome 12 karyogram of a representative case shows partial deletion of chromosome 12q (del12q) encompassing the *ARID2* locus (highlighted by dotted lines); 8 red horizontal lines indicate the sites of the del12q lesions in all 8 cases. Lower panels: somatic mutations are identified and confirmed by Sanger sequencing of paired tumor and germline (CD3+ T cells) samples. Distribution and types of mutations are shown by triangles on a schema of ARID2 (1,835 amino acids). b. An upper panel: variant allele frequencies of the mutations in paired tumor and germline samples measured by deep sequencing in 4 cases. A lower panel: depth of coverage of independent reads. * indicates $P < 0.0001$. c. Relative fold expression of *ARID2* are measured by quantitative RT-PCR in patients with *ARID2* deletion and intact *ARID2*.

**Figure 2. Mutation spectrum and clonal architecture associated with *ARID2* defects.**

a. Left panel: mutation spectrum in cases with *ARID2* mutation and deletion. b. Clonal architecture of *ARID2* and associated mutations evaluated by serial sequencing in 3 cases.

**Figure 3. Disease phenotype of *ARID2* and *U2AF1* defects.**

a. *ARID2* expression values extracted from Oncomine data and compared between healthy controls and MDS patients. b. Morphology of bone marrow smears in cases with *ARID2* and *U2AF1* mutations (Wright-Giemsa stain; X400). Arrow heads indicate hypolobated megakaryocytes.

**Figure 4. Functional impacts of *ARID2* defects.**

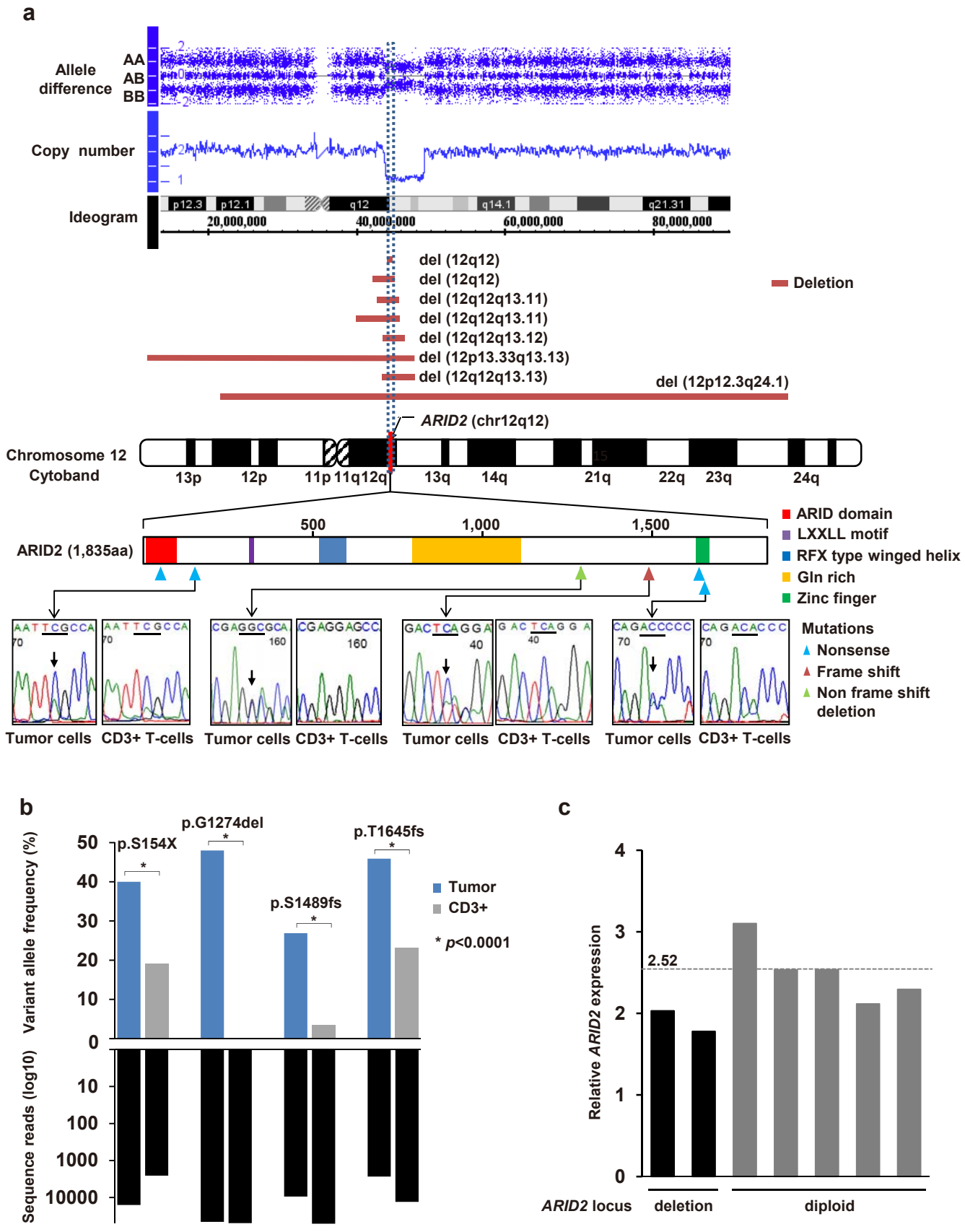
a, b. Expansion of myeloid cell lines (HL60 in a and K562 in b) treated with *ARID2* shRNA (TRCN0000166359). For each cell line, triplicate experiments were performed. c. Apoptosis assay indicates proportion of apoptotic cells in K562 cells with and without *ARID2* knockdown. d, e. Flow cytometry analysis shows effects of *ARID2* knockdown on surface expression of CD11b in HL60 (d) and GPA in K562 (e).

**Figure 5. *ARID2* defects in *U2AF1* mutants.**

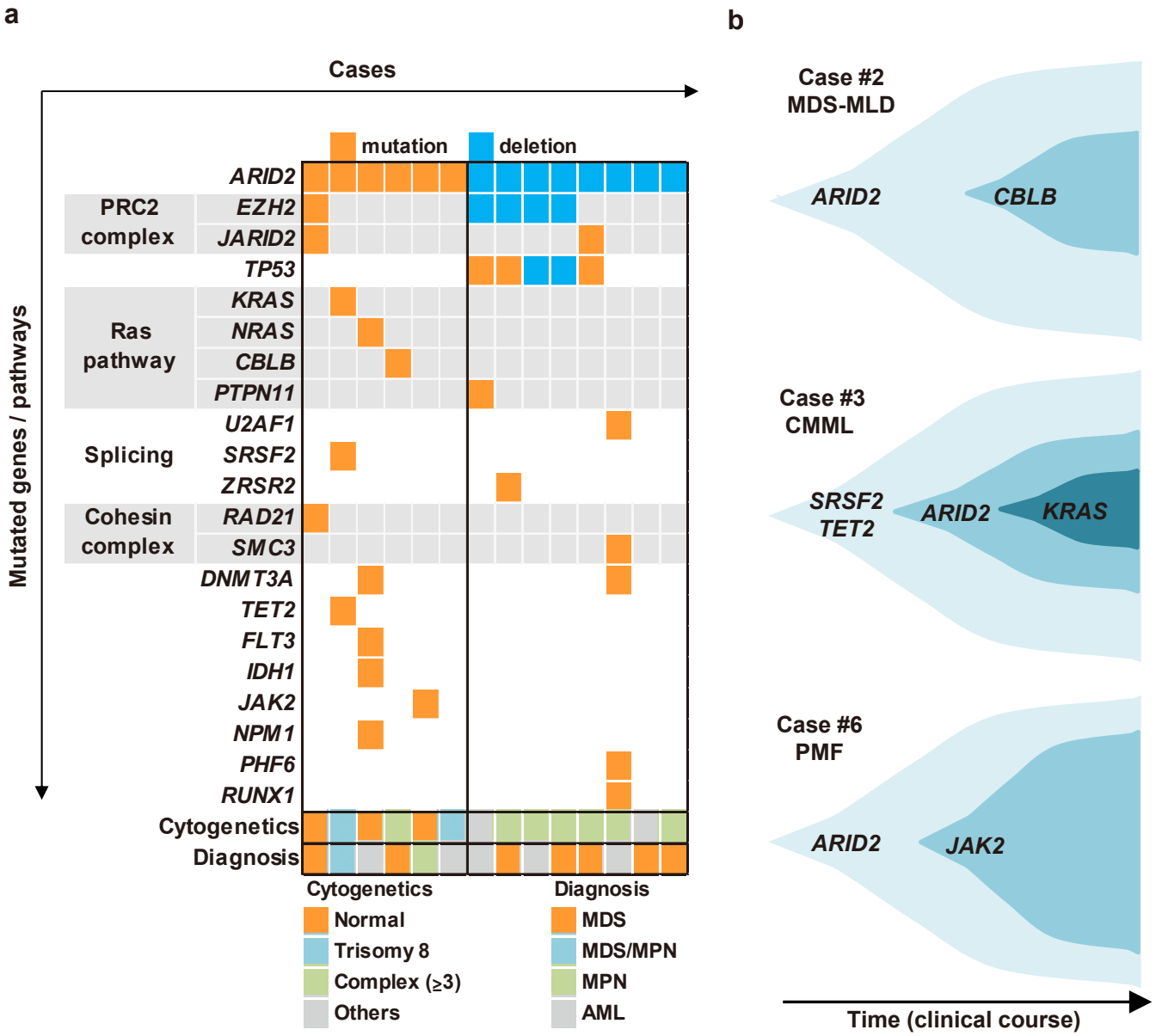
a. Normal and misspliced transcripts of exons 7, 8, and 9 are presented by Sashimi plots in cases with and without *U2AF1* mutations. A bar plot indicates percentages of sequence reads supporting complete and partial skipping of *ARID2* exon 8. b.

Confirmatory RT-PCR shows normal (348bp) and abnormal (180bp and 97bp) transcripts. Transcripts were validated by Sanger sequencing. A bar plot indicates relative fold expression of *ARID2* measured by quantitative RT-PCR in patients with and without *U2AF1* mutation and a healthy donor.

# Figure 1

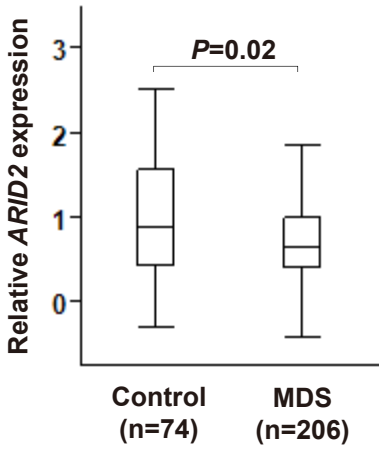

a
Allele difference
AA
AB
BB
Copy number
Ideogram
p12.3
p12.1
q12
q14.1
q21.31
20,000,000
40,000,000
60,000,000
80,000,000
del (12q12)
del (12q12)
del (12q12q13.11)
del (12q12q13.11)
del (12q12q13.12)
del (12p13.33q13.13)
del (12q12q13.13)
del (12p12.3q24.1)
Deletion
ARID2 (chr12q12)
Chromosome 12 Cytoband
13p
12p
11p
11q12q
13q
14q
21q
22q
23q
24q
ARID2 (1,835aa)
500
1,000
1,500
ARID domain
LXXLL motif
RFX type winged helix
Gln rich
Zinc finger
Mutations
Nonsense
Frame shift
Non frame shift deletion
Tumor cells
CD3+ T-cells
Tumor cells
CD3+ T-cells
Tumor cells
CD3+ T-cells
Tumor cells
CD3+ T-cells
b
Variant allele frequency (%)
p.S154X
p.G1274del
p.S1489fs
p.T1645fs
Tumor
CD3+
* p<0.0001
Sequence reads (log10)
c
Relative ARID2 expression
2.52
ARID2 locus
deletion
diploid

# Figure 2

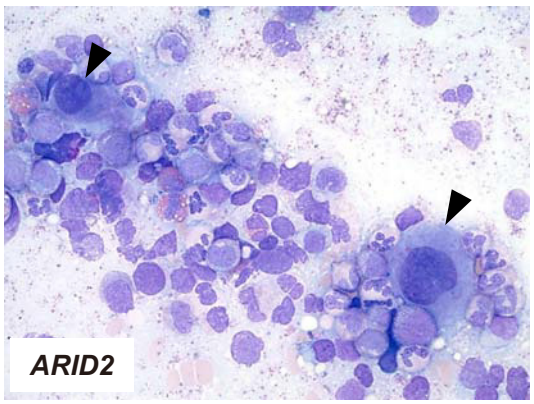

a
Cases
mutation
deletion
Mutated genes / pathways
PRC2 complex
Ras pathway
Splicing
Cohesin complex
ARID2
EZH2
JARID2
TP53
KRAS
NRAS
CBLB
PTPN11
U2AF1
SRSF2
ZRSR2
RAD21
SMC3
DNMT3A
TET2
FLT3
IDH1
JAK2
NPM1
PHF6
RUNX1
Cytogenetics
Diagnosis
Cytogenetics
Normal
Trisomy 8
Complex (≥3)
Others
Diagnosis
MDS
MDS/MPN
MPN
AML
b
Case #2
MDS-MLD
ARID2
CBLB
Case #3
CMML
SRSF2
TET2
ARID2
KRAS
Case #6
PMF
ARID2
JAK2
Time (clinical course)

## Figure 3

a


Relative ARID2 expression
3
2
1
0
P=0.02
Control
(n=74)
MDS
(n=206)

b


ARID2