SUPPLEMENTAL DATA

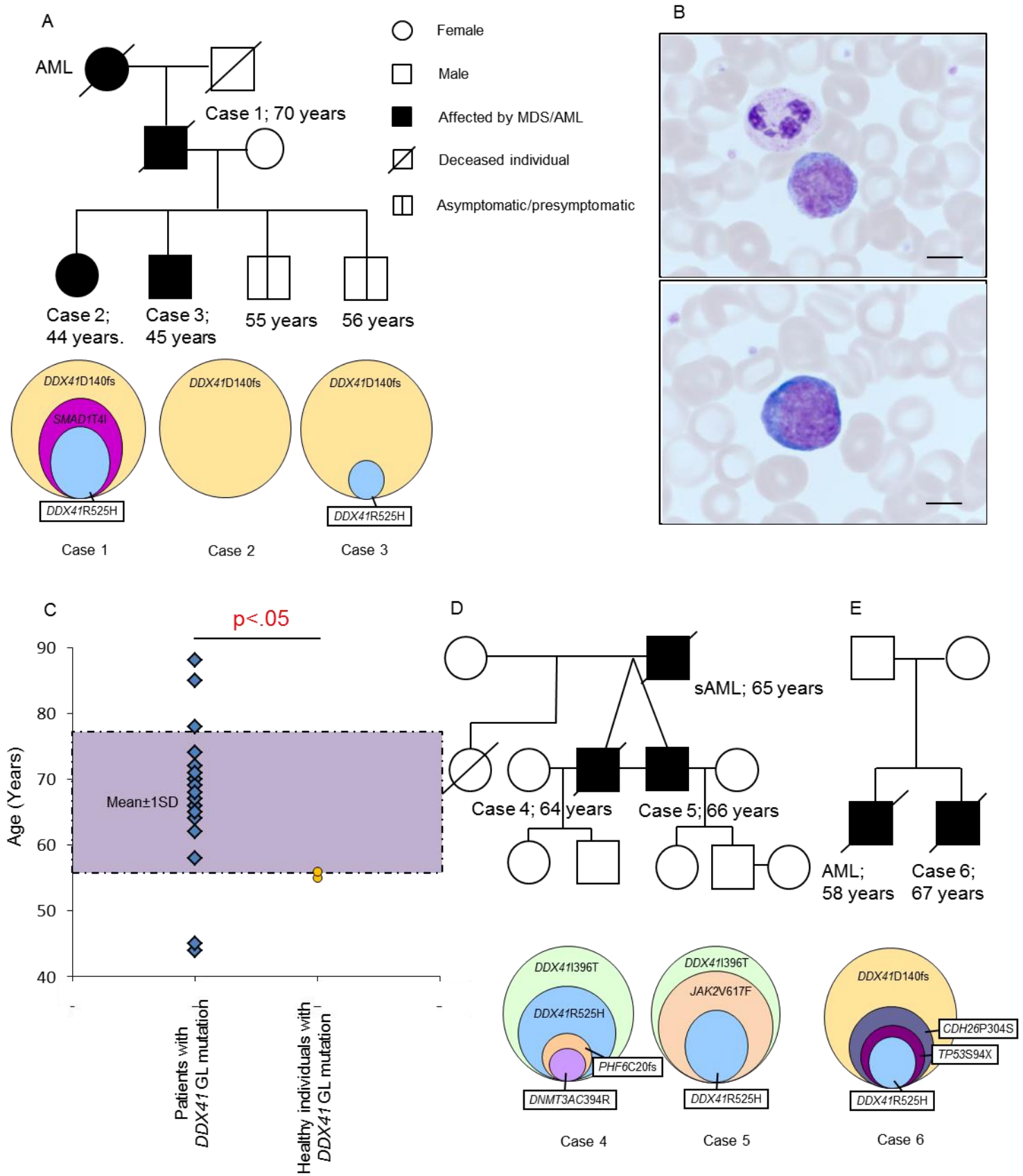


Figure S1: Related to Figure 1.

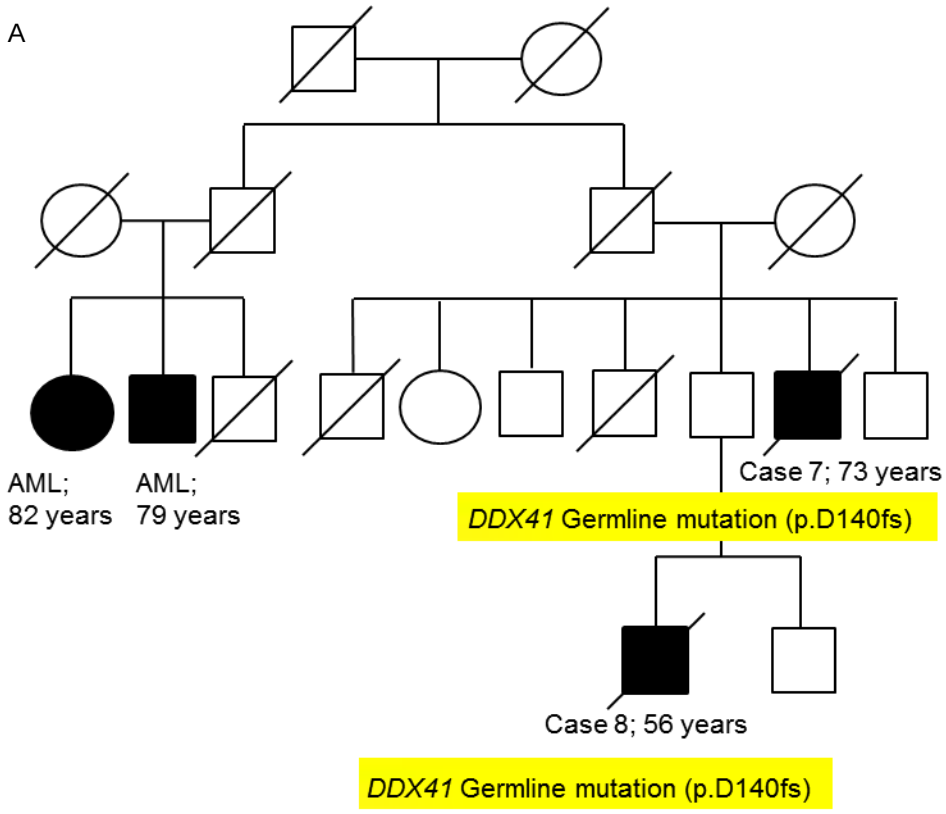
(A) Pedigree of family 1 depicts family members affected by AML or MDS and/or non-symptomatic or pre-symptomatic carriers. Case 1 and case 3 were affected by AML while case 2 was diagnosed with sAML. The paternal grandmother also died from AML. Case 1 harbored a germline mutation of *DDX41* (p.D140fs) and somatic mutations of *SMAD1* (p.T4I, 25%) and *DDX41* (p.R525H, 21%). Case 2 harbored germline mutation of *DDX41* (p.D140fs). Case 3 harbored germline mutation of *DDX41* (p.D140fs) and somatic mutation of *DDX41* (p.R525H, 12%). Frequencies of mutations are reflected by circle size. Case number is annotated according to Table 1. Asymptomatic/presymptomatic carrier-clinically unaffected at this time but could later exhibit symptoms.

(B) Peripheral blood smear of one healthy brother in family 1 contained immature-looking monocytes. Scale bar: 5 µm

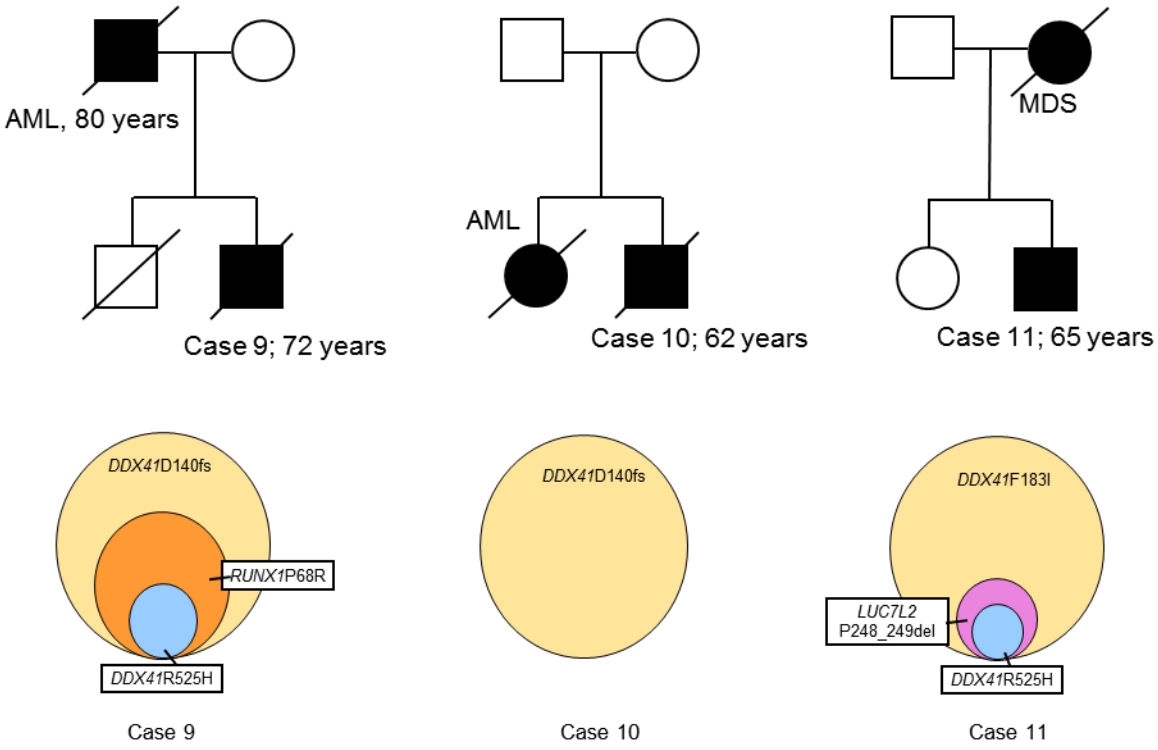
(C) Age of individuals with *DDX41* germline mutations. Pink area indicates the range (mean±1SD; 67±11 years) of age at disease presentation in patients with *DDX41* germline mutation (each blue diamond represents one patient). Orange circles indicate age of the healthy family members harboring the germline mutation at the time of sequencing. (GL=germline)

(D) Family 2: The twins were diagnosed as MDS-RCMD. Their father died from sAML. Case 4 harbored germline mutation of *DDX41* (p.I396T) and somatic mutations of *DDX41* (p.R525H, 28%), *PHF6* (p.C20fs, 15%) and *DNMT3A* (p.C394R, 9.6%). Case 5 harbored germline mutation of *DDX41* (p.I396T) and somatic mutations of *JAK2* (p.V617F, 46%) and *DDX41* (p.R525H, 32%). Mutational frequencies are reflected by circle size.

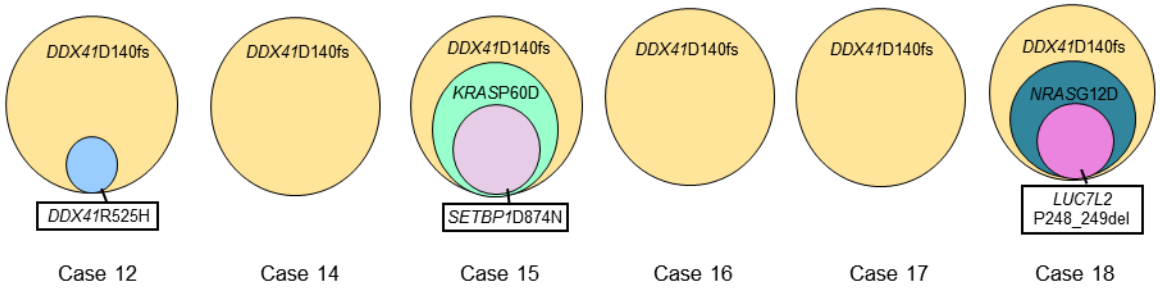
(E) Family 3: Case 6 was diagnosed as MDS-RAEB-I. His brother died from leukemia. Leukemic cells showed germline mutation of *DDX41* (p.D140fs) and somatic mutations of *CDH26* (p.P304S, 37%), *TP53* (p.S94X, 22%) and *DDX41* (p.R525H, 20%). Mutational frequencies are reflected by circle size.



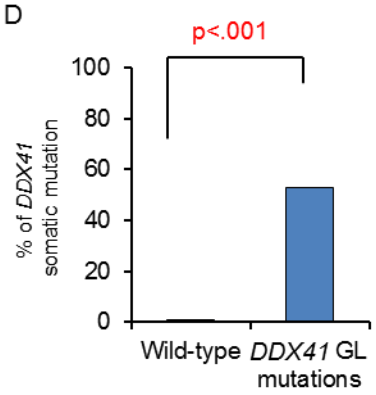
B



C



D



E

Variants	Population frequency
c.419insGATG, p.D140fs	1/12,518
c.T1187C, p.I396T	Not reported
c.C1200T	37%

F

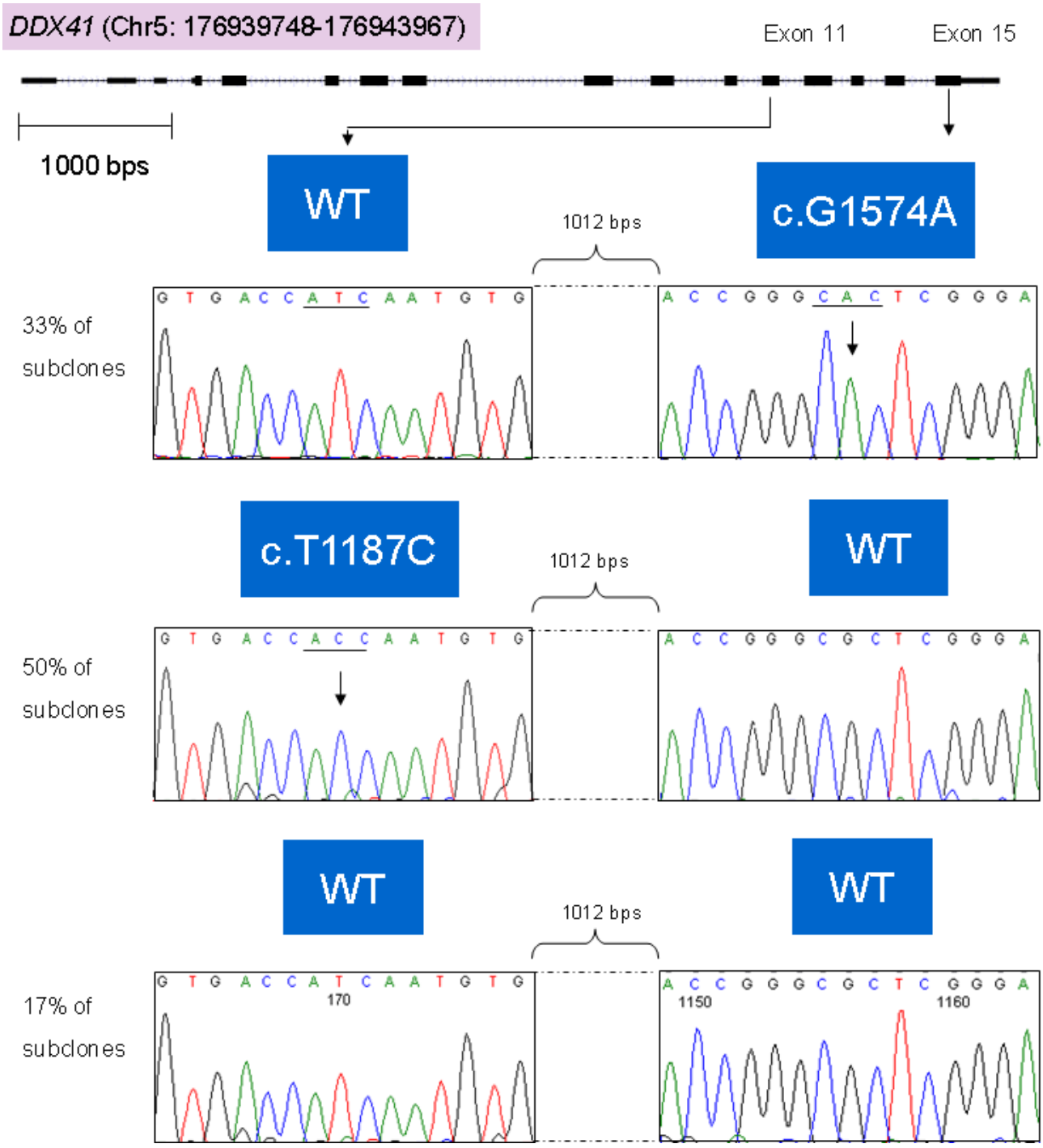

DDX41 (Chr5: 176939748-176943967)
Exon 11
Exon 15
1000 bps
WT
c.G1574A
1012 bps
33% of subclones
G T G A C C A T C A A T G T G
A C C G G G C A C T C G G G A
c.T1187C
WT
1012 bps
50% of subclones
G T G A C C A C C A A T G T G
A C C G G G C G C T C G G G A
WT
WT
1012 bps
17% of subclones
G T G A C C A T C A A T G T G
170
A C C G G G C G C T C G G G A
1150
1160

Figure S2: Related to Table 1.

(A) Family 4: Case 7 was diagnosed as sAML while his nephew (case 8) was diagnosed as pAML. Both cases harbored canonical *DDX41* germline mutation. Paternal cousins of the index case were also diagnosed with AML at the age of 79 and 89.

(B) Pedigrees indicate affected members of additional families. Case 9 was diagnosed with MDS-RAEB-I. His father died from leukemia. Leukemic cells showed germline mutation of *DDX41* (p.D140fs) and somatic mutations of *RUNX1* (p.P68R, 30%) and *DDX41* (p.R525H, 25.8%). Case 10 was diagnosed with MDS-RAEBII. His sister died from leukemia. Leukemic cells showed germline mutation of *DDX41* (p.D140fs). Case 11 was diagnosed with MDS-RAEB-I. His mother died from MDS. Leukemic cells showed germline mutation of *DDX41* (p.F183I) and somatic mutations of *LUC7L2* (p. 248_249del, 9%) and *DDX41* (p.R525H, 5%).

(C) Other cases were identified to harbor the germline mutation (p.D140fs): Case 12 (sAML) showed germline mutation of *DDX41* (p.D140fs) and somatic mutation of *DDX41* (p.R525H, 11%). Case 14 (RAEB-I), case 16 (RAEB-I) and case 17 (pAML) showed germline mutation of *DDX41* (p.D140fs). Case 15 (CMML-1) showed germline mutation of *DDX41* (p.D140fs) and somatic mutation of *KRAS* (p.P60D, 30%) and *SETBP1* (p.D874N, 45%). Case 18 (sAML) showed germline mutation of *DDX41* (p.D140fs) and somatic mutation of *NRAS* (p.G12D, 47%) and *LUC7L2* (p. 248_249del, 11.5%). Mutational frequencies are reflected by circle size.

(D) Germline alterations (p.D140fs, p.I396T, p.F183I and p.Q52fs) associated with somatic *DDX41* mutations (53%; 9/17), compared to those who showed wild-type germline configuration *DDX41* (0.8%; 8/1026, p<.001). (GL=germline)

(E) Summary of population frequencies of the three *DDX41* variants

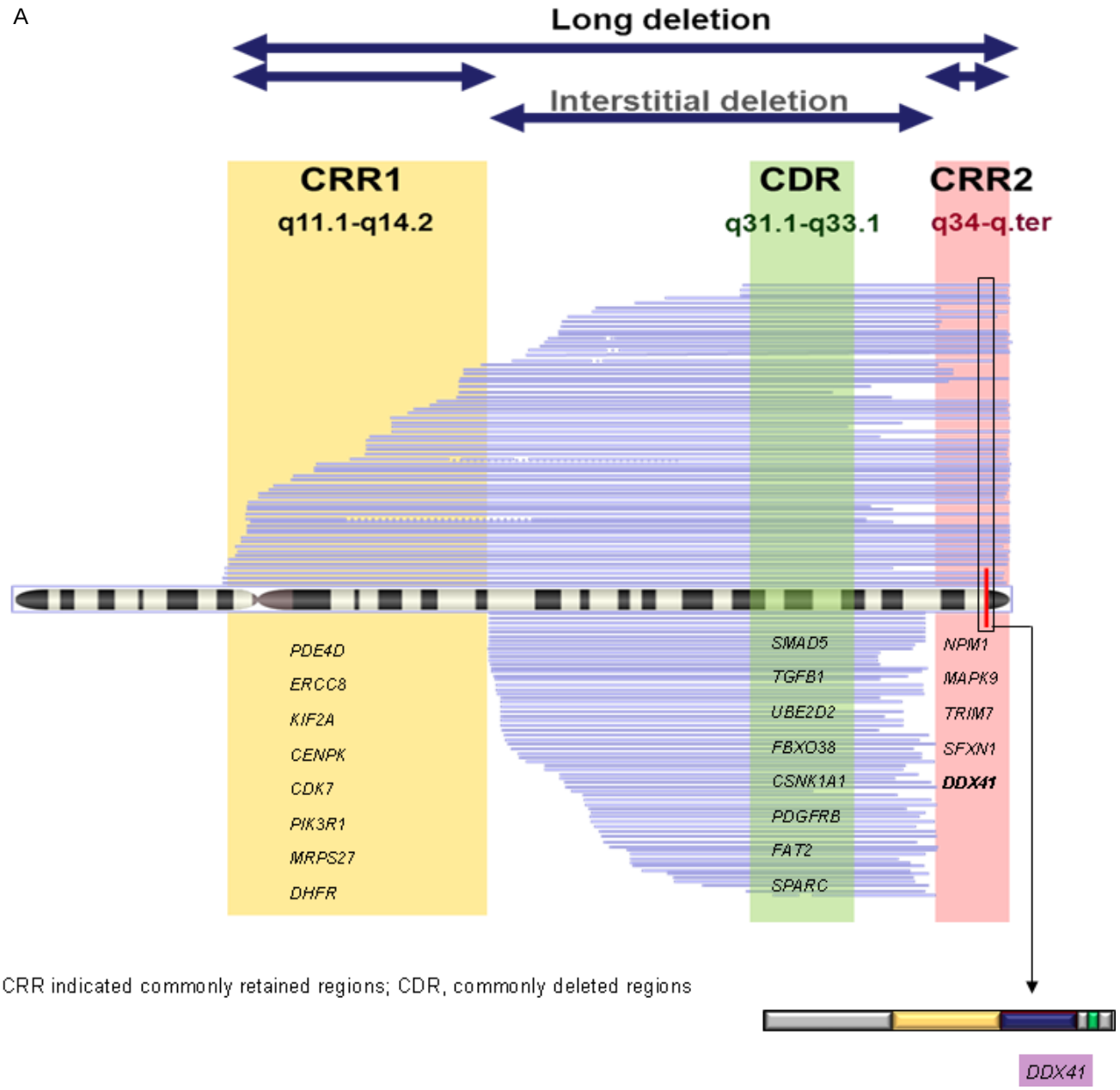
(F) Subcloning analysis of *DDX41* germline and somatic mutations (Family 2). Each *DDX41* allele was amplified by primers on exon 11 and 15 to include both loci of germline (p.I396T) and somatic (p.R525H) mutations. Sanger sequencing was performed for each subclone after TA cloning of amplicons into pCR2.1 vectors (n=6).

Table S1: Related to Table 1. Classification of myeloid neoplasms in the analyzed cohort.

Diseases	WES	Targeted sequencing
MDS		
RCUD,RCMD,5q-,MDS-U,RARS	48	188
RAEBI/II	27	115
MDS/MPN		
CMML	25	68
JMML,aCML,MDS/MPN-U (RARS-T)	15	53
MPN		
PV,PMF,ET	5	45
AML		
Primary AML	199	103
Secondary AML	22	132
Total	341	704

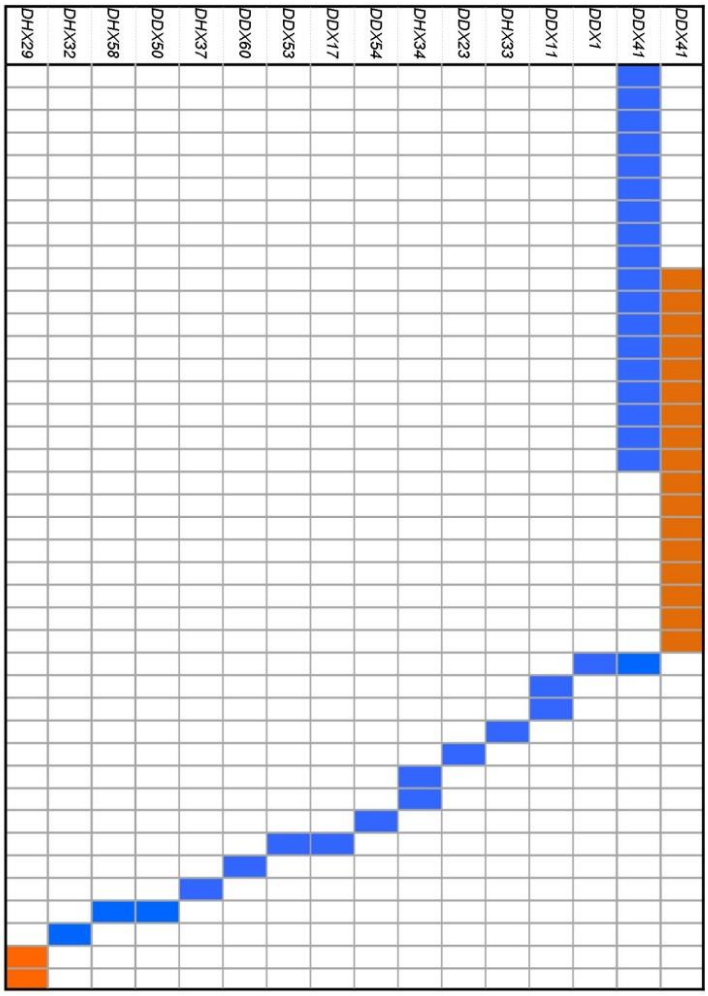
MDS indicated myelodysplastic syndrome; RCUD, refractory cytopenia with unilineage dysplasia; RCMD, refractory cytopenia with multilineage dysplasia; MDS-U, MDS unclassifiable; RARS, refractory anemia with ring sideroblasts; RAEB, refractory anemia with excess blasts; MDS/MPN, MDS/myeloproliferative neoplasms; CMML, chronic myelomonocytic leukemia; aCML, atypical chronic myeloid leukemia; JMML, juvenile myelomonocytic leukemia; RARS-T, RARS associated with marked thrombocytosis; PV, polycythemia vera; PMF, primary myelofibrosis; ET, essential thrombocythemia; and AML, acute myeloid leukemia.

A



CRR indicated commonly retained regions; CDR, commonly deleted regions

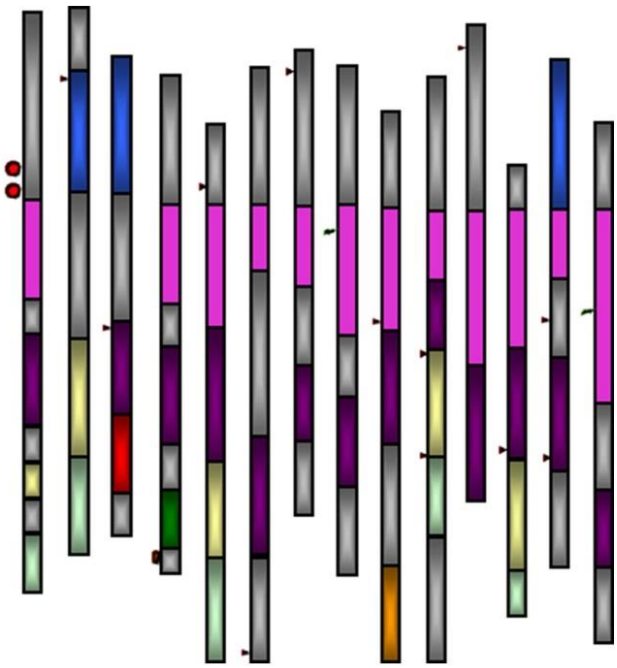
B



Somatic mutation

Germline mutations

Genes	Events
DDX1	Splice site (Exon15-2)
DDX11	p.A607P, p.A848V
DDX23	p.R54Q
DHX34	p.T556A, p.V648I
DHX33	p.A440V
DDX54	p.T280M
DDX17	Splice site (Exon5+1)
DDX53	p.P41L
DDX60	p.I1676V
DHX37	p.D180E
DDX50	p.R726*
DHX58	p.S412T
DHX32	p.V81A
DHX29	p.V543M, p.E521K



Mutation at splice site

Missense mutation

Non-sense mutation

Germline mutation

Helicase ATP binding

DEAD/H box

ATP dependent helicase, C-terminal

Helicase associated domain

Domain of unknown function

DBP

GUCT

C-terminal domain

Zinc finger

Figure S3: Related to Figure 2.

(A) Shown are del(5q) cases in the analyzed cohort. Using SNP-A karyotyping, deletions of 5q were identified in 133 cases among myeloid neoplasms. The blue bars visualize deletions of 5q. Depicted are genes located in CRR1, CRR2 (commonly retained regions) and CDR (commonly deleted regions). Deletions of 5q involving the *DDX41* locus, which was located in CRR2, were present in 26% of del5q cases (n=35/133)

(B) Mutations and domain structure of other DEAD/H box RNA helicases in the analyzed cohort. We identified 15 somatic mutations of other members of DEAD/H-box RNA helicase family in the cohort with myeloid neoplasms (n=342), which were *DDX23*-R54Q, *DDX11*-A607P and A848V, *DHX33*-A440V, *DHX34*-T556A and V648I, *DDX1* (splice site; exon15-2), *DDX17* (splice site; exon5+1), *DDX50*-R726*, *DDX53*-P41L, *DDX54*-T280M, *DDX60*-I1676V, *DHX32*-V81A, *DHX37*-D180E, *DHX58*-S412T. We also identified 2 rare germline events; *DHX29* c.G1627A (p.V543M) and c.G1561A (p.E521K), not found in the SNP database (ESP). Lower panel shows domain structures of DEAD/H box RNA helicase members.

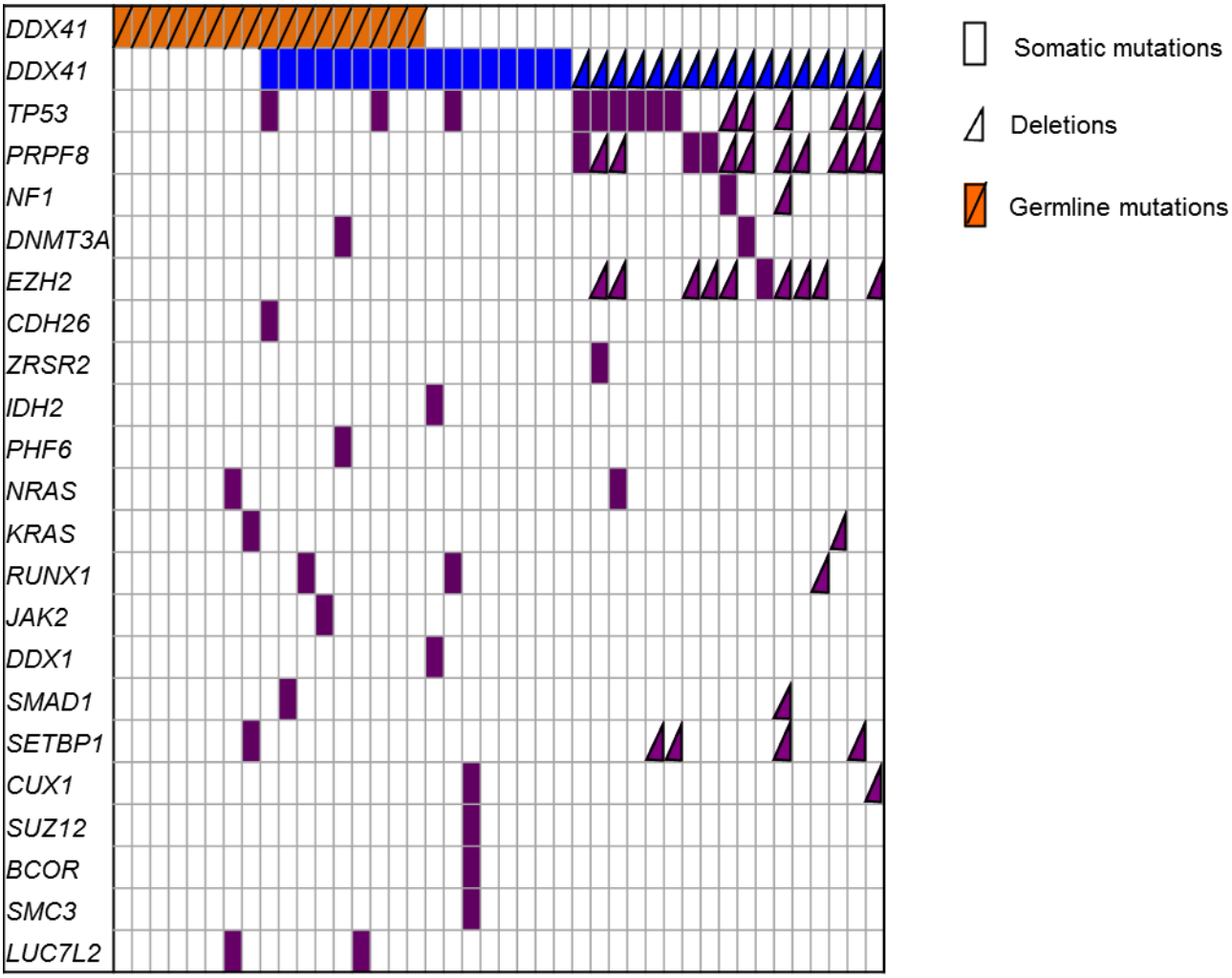


Figure S4: Related to Figure 3.

DDX41 mutations, deletions and associated mutations. Seventeen cases harbored *DDX41* germline mutations (p.Q52fs, p.D140fs, p.M155I and p.I396T; indicated in orange), while somatic mutation of *DDX41* (indicated in blue) was found in 17 cases. Other mutations are indicated in purple. Seventeen deletion cases are shown as triangle.

Table S2: Related to Figure 3. Characteristics of patients diagnosed with myeloid neoplasms and treated with lenalidomide.

	Responders % (n=63)	Non-responders % (n=48)	p value
Age (years)	66	67	
Females	41	29	
Median IPSS score	0.5	0.5	
Median R-IPSS score	3.5	3.5	
Cytogenetics at treatment			
Normal	48	48	1
Complex Karyotype	13	27	0.1
Del5/5q	31	27	0.8
Del7/7q	8	14	0.5
Trisomy8	13	14	1
Del20/20q	4	17	0.1
Diagnosis at treatment			
Low-risk MDS	57	56	1
High-risk MDS	25	17	0.4
sAML	2	0	1
MDS/MPN	11	17	0.4
Primary Myelofibrosis	5	6	1

Table S3: Related to Figure 3. Characteristics of 19 patients treated with lenalidomide showing *DDX41* mRNA expression level.

Cases	*DDX41* expression*	Lenalidomide response**	Disease	Cytogenetics
1	0.27	Non-response	RAEBI	Normal
2	0.97	Non-response	RARS	47,XY,+8
3	0.58	Non-response	RAEBI	Normal
4	1.6	Non-response	MDS/MPN	Normal
5	2.56	Non-response	RAEBII	Complex
6	0.75	Non-response	RARS	Normal
7	0.54	Non-response	RCMD	Complex
8	1.42	Non-response	CMML1	Normal
9	0.63	Non-response	RCUD	Normal
10	0.5	Non-response	MDS/MPN	46,XY,del(5)(q12q33),del(13)(q12q14)
11	0.38	HI-E	PMF	Normal
12	0.62	BM-CR	RAEBI	Normal
13	0.43	BM-CR	RAEBI	Normal
14	0.27	HI-E,HI-N,HI-P	RAEBII	Complex
15	0.42	BM-CR	RAEBII	Normal
16	0.42	HI-E	5q-syndrome	46,XX,del(5)(q13q33)
17	0.78	HI-E	MDS/MPN	Normal
18	0.44	BM-CR	RAEBII	Complex
19	0.51	BM-CR	RAEBII	Normal

*Level of *DDX41* is measured relatively to healthy bone marrow cells (1.00).

**Response criteria are used according to the international working group response criteria for myelodysplastic syndromes. (Cheson et al., 2000)

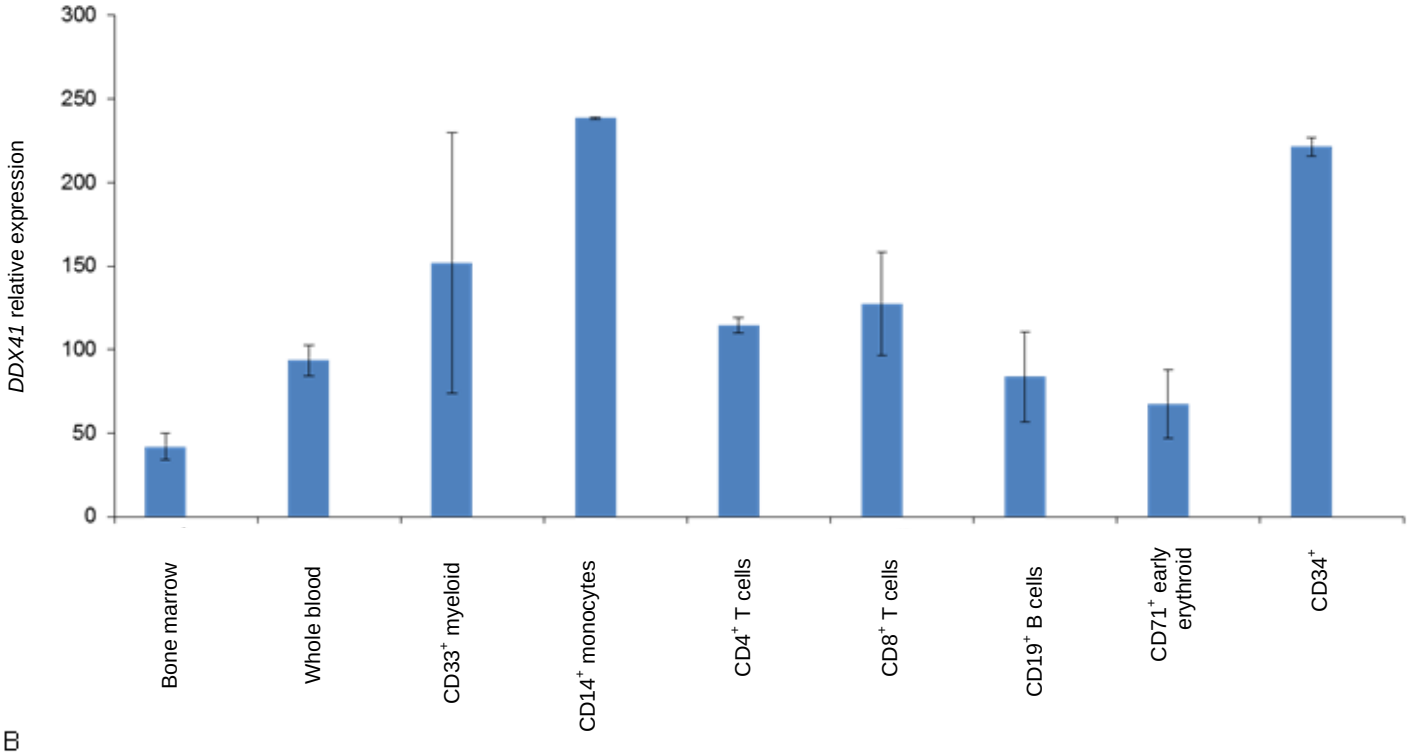
HI-E: hematologic improvement-erythroid response, HI-N: hematologic improvement-neutrophil response, HI-P: hematologic improvement-platelet response, BM-CR: bone marrow evaluation-complete remission,

Table S4: Related to Figure 3. Characteristics of *DDX41* mutants and wild-type.

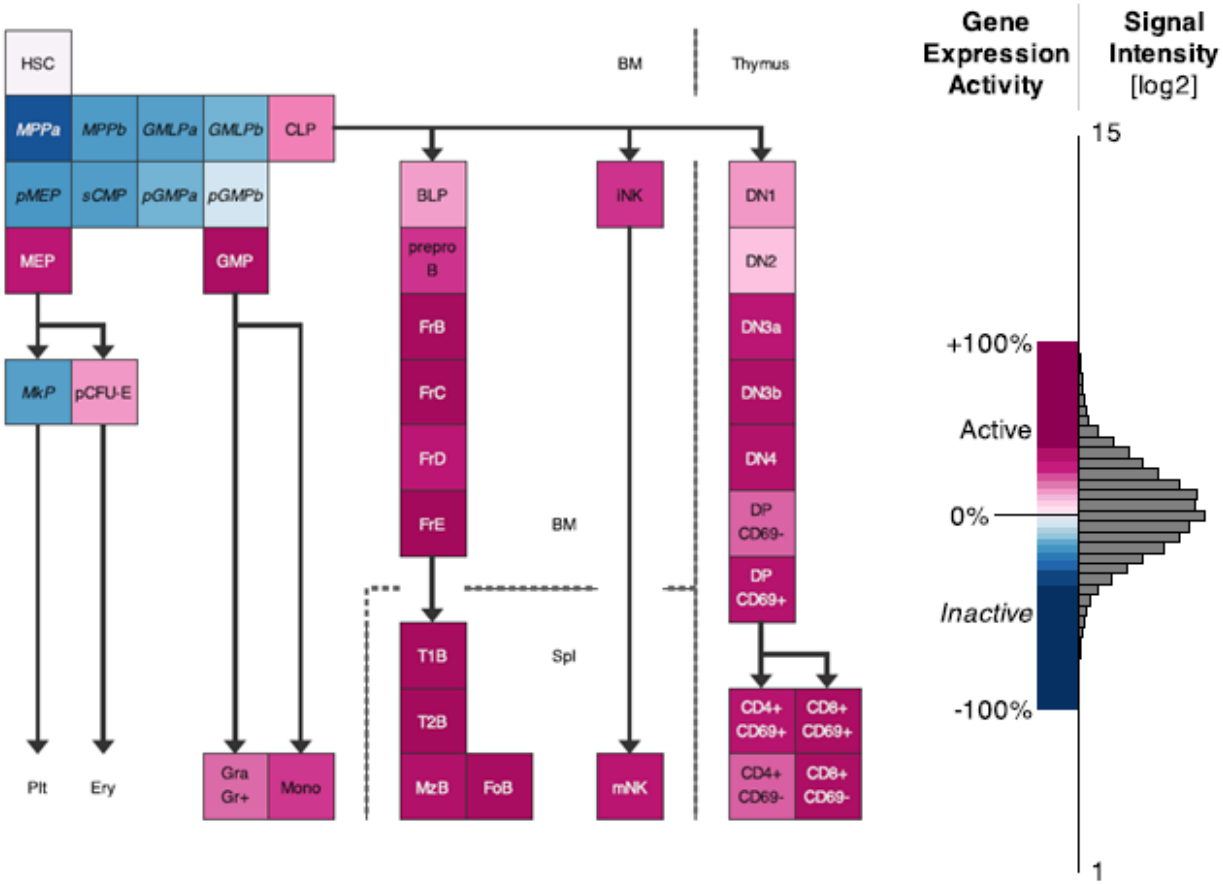
Variables	*DDX41* mutation	*DDX41* wild-type	p value*
Number	27	1018	
Age (years)	67	65	ND
Male sex (%)	92	61	<0.05
Cytogenetics (%)			
Abnormal	30	53	0.0045
Del5/5q	8	12	1
Trisomy 8	4	9	0.8
Diagnosis (%)			
High-risk MDS	37	17	0.009
Low-risk MDS	22	29	1
pAML	22	11	0.2
sAML	19	18	1

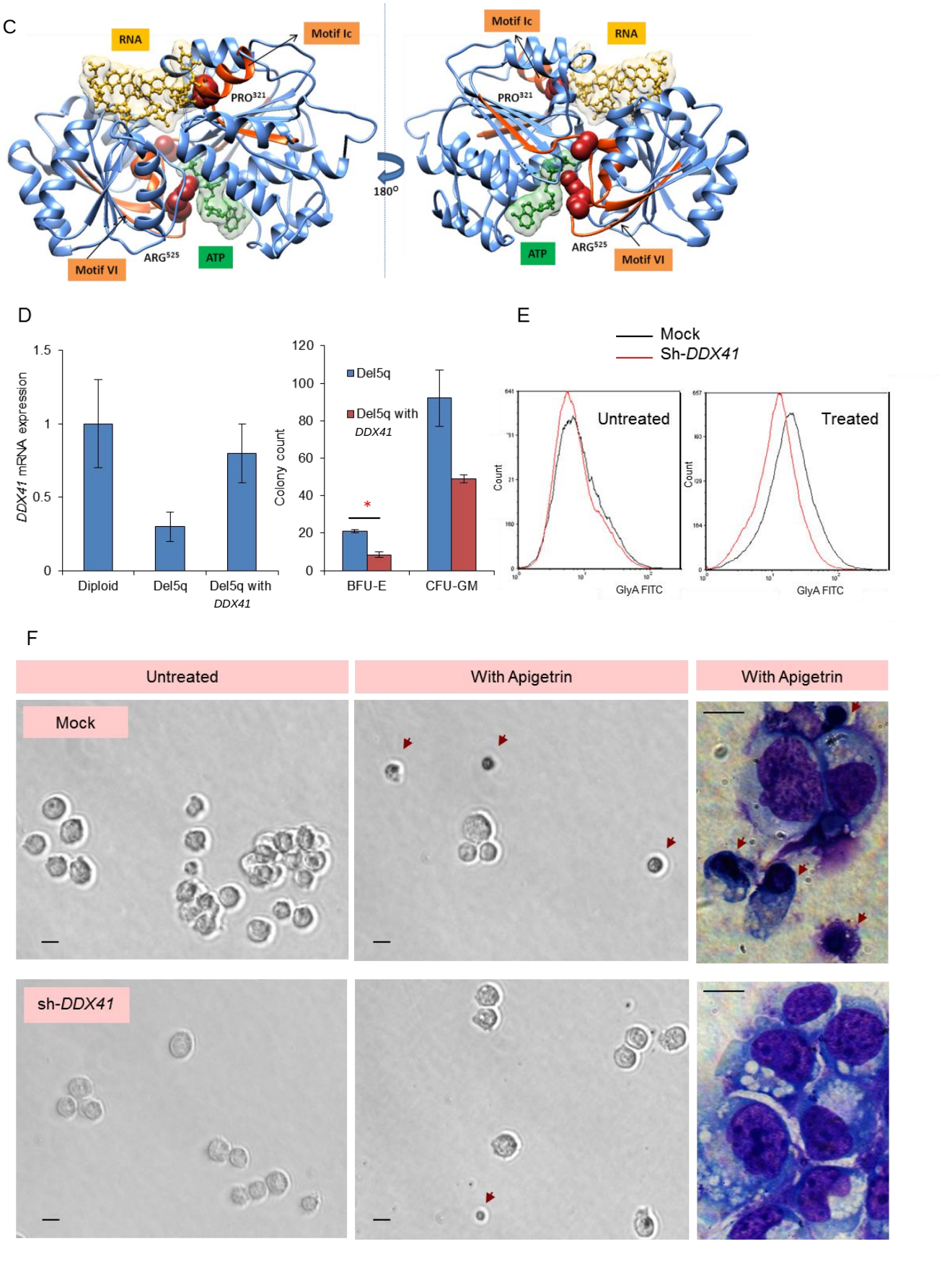
*Fisher's exact test was used to determine p values, except where otherwise indicated. p values in multiple comparisons were evaluated by Bonferroni correction.

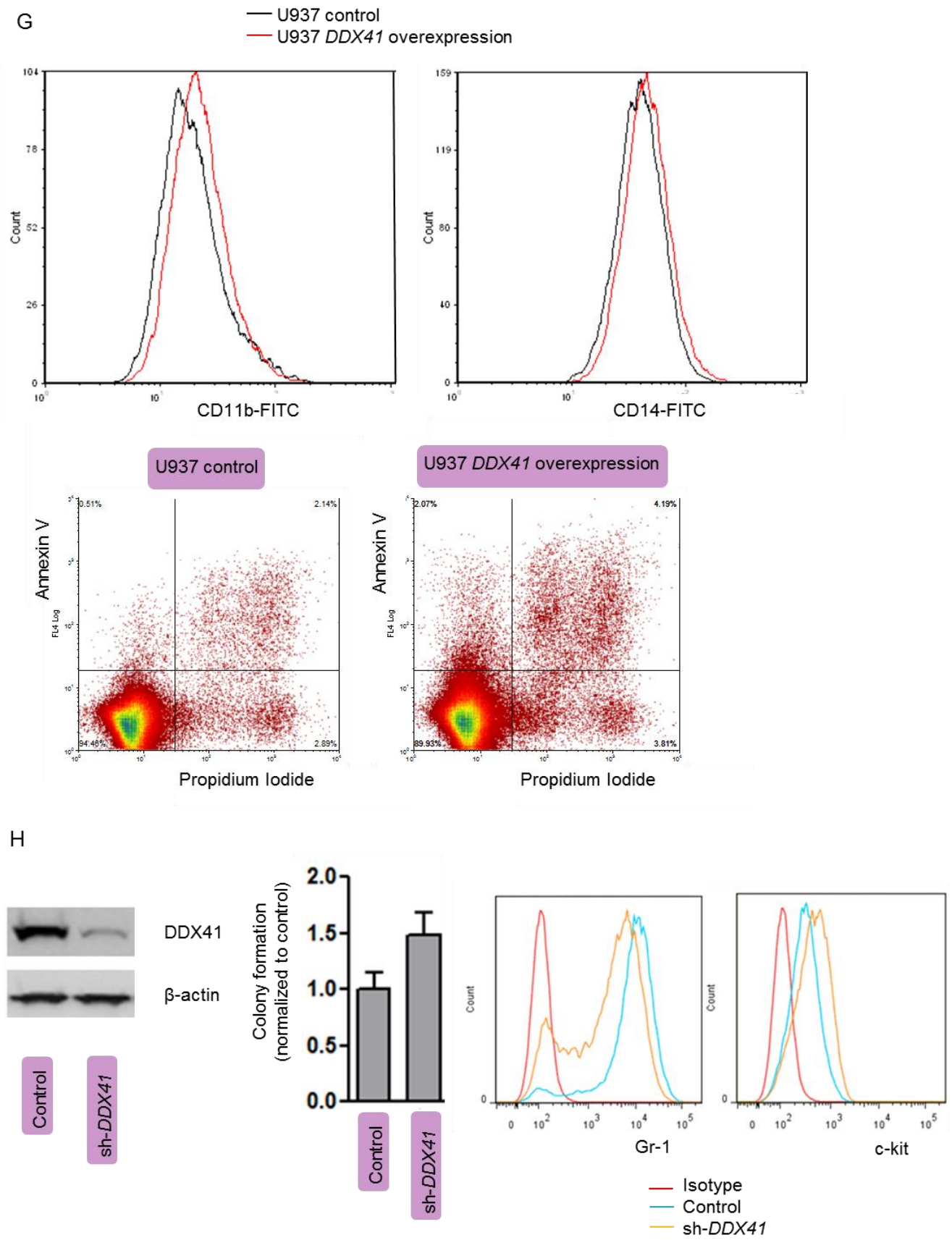
A


DDX41 relative expression
300
250
200
150
100
50
0
Bone marrow
Whole blood
CD33+ myeloid
CD14+ monocytes
CD4+ T cells
CD8+ T cells
CD19+ B cells
CD71+ early erythroid
CD34+

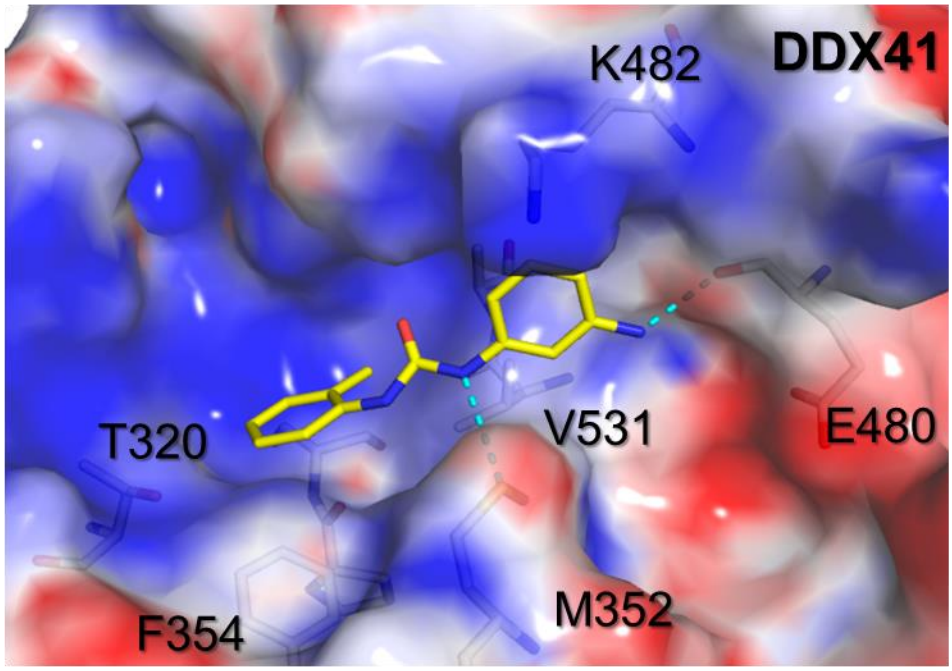
B


HSC
MPPa
MPPb
GMLPa
GMLPb
CLP
pMEP
sCMP
pGMPa
pGMPb
MEP
GMP
M&P
pCFU-E
Plt
Ery
Gra
Gr+
Mono
BM
Thymus
BLP
prepro B
FrB
FrC
FrD
FrE
BM
T1B
T2B
MzB
FoB
Spl
iNK
mNK
DN1
DN2
DN3a
DN3b
DN4
DP CD69-
DP CD69+
CD4+ CD69+
CD8+ CD69+
CD4+ CD69-
CD8+ CD69-
Gene Expression Activity
Signal Intensity [log2]
15
+100%
Active
0%
Inactive
-100%
1

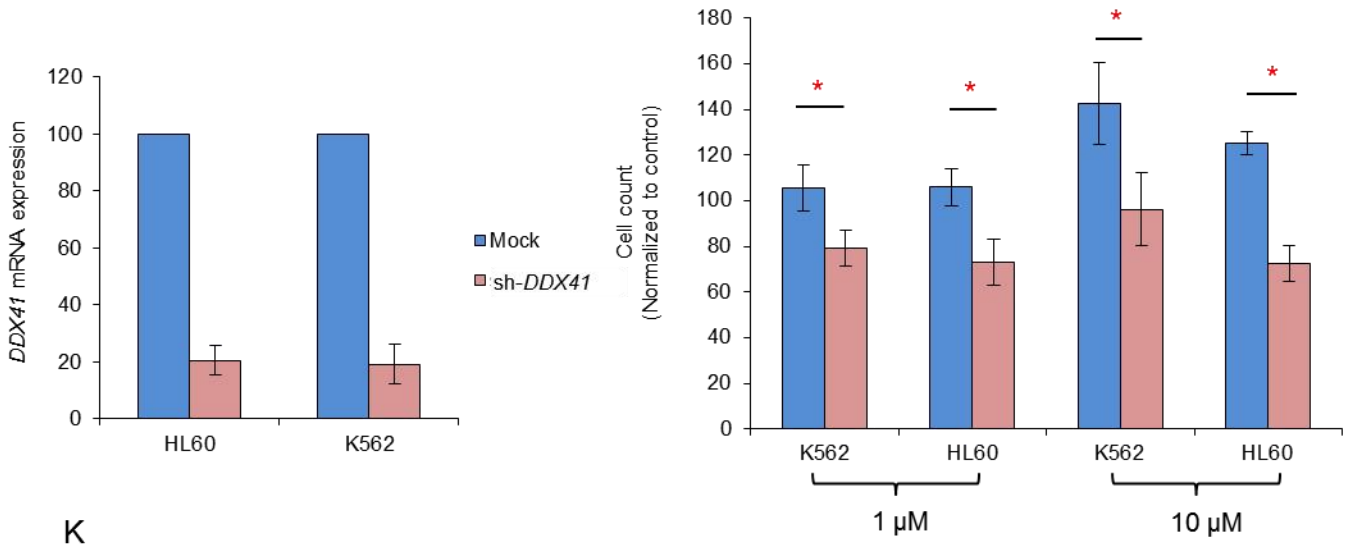

C
RNA
Motif Ic
PRO321
ARG525
ATP
Motif VI
180°
D
DDX41 mRNA expression
Diploid
Del5q
Del5q with DDX41
Colony count
Del5q
Del5q with DDX41
BFU-E
CFU-GM
E
Mock
Sh-DDX41
Untreated
Treated
Count
GlyA FITC
F
Untreated
With Apigetrin
With Apigetrin
Mock
sh-DDX41


G
U937 control
U937 DDX41 overexpression
Count
CD11b-FITC
CD14-FITC
U937 control
U937 DDX41 overexpression
Annexin V
Propidium Iodide
10.51%
2.14%
94.46%
2.89%
2.07%
4.19%
89.93%
3.81%
H
DDX41
β-actin
Control
sh-DDX41
Colony formation (normalized to control)
2.0
1.5
1.0
0.5
0.0
Gr-1
c-kit
Isotype
Control
sh-DDX41

I


K482
DDX41
T320
V531
E480
F354
M352

J


DDX41 mRNA expression
120
100
80
60
40
20
0
HL60
K562
Mock
sh-DDX41
Cell count
(Normalized to control)
180
160
140
120
100
80
60
40
20
0
*
*
*
*
K562
HL60
K562
HL60
1 μM
10 μM

K

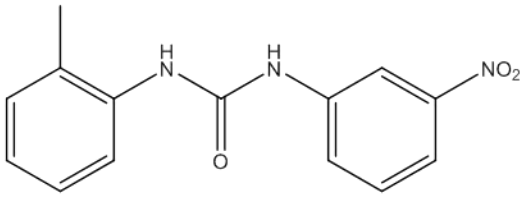

H
N
H
N
O
NO2

Figure S5: Related to Figure 4.

(A) Expression of *DDX41* in various tissues was determined by normalization to *GAPDH.* Bone marrow cells: 41.9±8.1; Whole blood: 93.55±9.15; $CD33^+$ Myeloid cells: 151.85±77.95; $CD14^+$ Monocytes: 238.7±0.6; $CD4^+$ T cells: 114.55±4.85; $CD8^+$ T cells: 127.25±30.85; $CD19^+$ B cells: 83.75±27.05; $CD71^+$ early erythroid cells: 67.35±20.45; $CD34^+$ cells: 221.4±5.6. (http://biogps.org) Each bar represents the mean±SEM of 2 independent experiments.

(B) Gene expression values of *DDX41* in blood cells were analyzed via Gene Expression Commons; HSC (Hematopoietic stem cells): 5.63%; MPPa (multipotent progenitor subset A): -81.59%; MPPb (multipotent progenitor subset B): -55.69%; GMLPa (granulo/macrophage/lymphoid progenitor subset A): -53.15; GMLPb (granulo/macrophage/lymphoid progenitor subset B): -43.08; CLP (common lymphoid progenitor): 58.45%; pMEPa (pre megakaryocyte/erythroid progenitor): -57.52%; sCMP (strict common myeloid progenitor): -54.23%; pGMPa (pre granulocyte/macrophage progenitor subset A): -43.7%; pGMPb (pre granulocyte/macrophage progenitor subset B): -18.11%; MEP (megakaryocyte/erythroid progenitor): 85.62%; GMP (granylocyte/macrophage progenitor): 93.19%; MkP (megakaryocyte progenitor): -53.23%); pCFU-E (preCFU-E): 50.64%; Gra,Gr^+ (granulocyte): 63.47%; Mono (monocyte): 73.69%; BPLP (earliest B lymphocyte progenitor): 48.11%; preproB (pre pro B cells): 77.88%; FrB (fraction B B cells): 95.18%; FrC (Fraction C B cells): 96.21%; FrD (Fraction D B cells): 85.96%; FrE (Fraction E B cells): 97.26%; T1B (T1 B cells): 95.91%; T2B (T2 B cells): 96.16%: MzB (marginal zone B cells): 90.22%; FoB (follicular B cells): 94.58%; iNK (immediate natural killer cells): 77.48%: mNK (mature natural killer cell): 86.09%; DN1 (double negative T cell 1): 50.34%; DN2 (double negative T cell 2): 35.23%; DN3a (double negative T cell 3a): 85.25%; DN3b (double negative T cell 3b): 91.13%; DN4 (double negative T cell 4): 89.62%; DP $CD69^-$ (double positive $CD69^-$ T cell): 65.44%; DP $CD69^+$ (double positive $CD69^+$): 88.51%; $CD4^+CD69^+$ ($CD4^+CD69^+$ T cell): 87.58%; $CD4^+CD69^-$ ($CD4^+CD69^-$ T cell): 66.64%; $CD8^+CD69^+$ ($CD8^+CD69^+$ T cell): 93.73%; $CD8^+CD69^-$ ($CD8^+CD69^-$ T cell): 92.90%. (https://gexc.stanford.edu)

(C) Structural model of the helicase core of DDX41. Mutations in the helicase core (p.R525H and p.P321L) are highlighted in orange. The structure model was created by combining the existing structure of the partial helicase domain of DDX41 and the structure of Drosophila Vasa (PDB ID: 2DB3). The modeled structure was energy minimized using a combination of steepest descent and conjugate gradient methods. The final structure was refined for steric clashes. Structure was visualized in UCSF Chimera molecular extensible environment. Color scheme: RNA atoms: Gold; ATP: Green; Conserved motifs: Orange ribbons, significant residues: Red sphere.

(D) *DDX41* expression levels in del(5q) primary sAML cells and del(5q) primary cells with *DDX41* rescue were determined by RT-PCR (left). Differential colony counting of del(5q) primary cells compared to *DDX41* overexpressing del(5q) primary cells plated in methylcellulose. Colony numbers were assessed after 10-14 days. (*p<.05). Each bar represents the mean±SEM of 3 independent experiments performed in duplicates (right).

(E) Flow cytometry analysis of Glycophorin A expression in *DDX41* knockdown K562 cells compared to control after treatment with apigetrin for 9 days. Mean fluorescence intensity (MFI) in mock cells untreated vs. treated = 8.3 vs. 18.5. MFI in sh-*DDX41* untreated vs. treated = 7.4 vs. 11.7.

(F) Differentiation state was evaluated by morphological changes using light microscopy (left and middle, scale bar: 20 µm) and Wright-Giemsa staining (right, scale bar: 10 µm.) of *DDX41* knockdown K562 cells compared to mock transduced cells after inducing erythroid differentiation with apigetrin. Arrows indicate mature erythroid cells.

(G) Expression of CD11b and CD14 in U937 with *DDX41* rescue and control cells was determined by FACS [MFI: 22.1 vs. 19 and 45.1 vs. 39.2, respectively (upper panel)]. Staining of apoptotic cells was performed in *DDX41* overexpressing

U937 cells compared to control cells: early apoptotic cells (Annexin V^+/PI^-: 2.07% vs. 0.51%), necrotic cells (Annexin V^-/PI^+: 3.81% vs. 2.89%) and dead cells (Annexin V^+/PI^+: 4.19% vs. 2.14%).

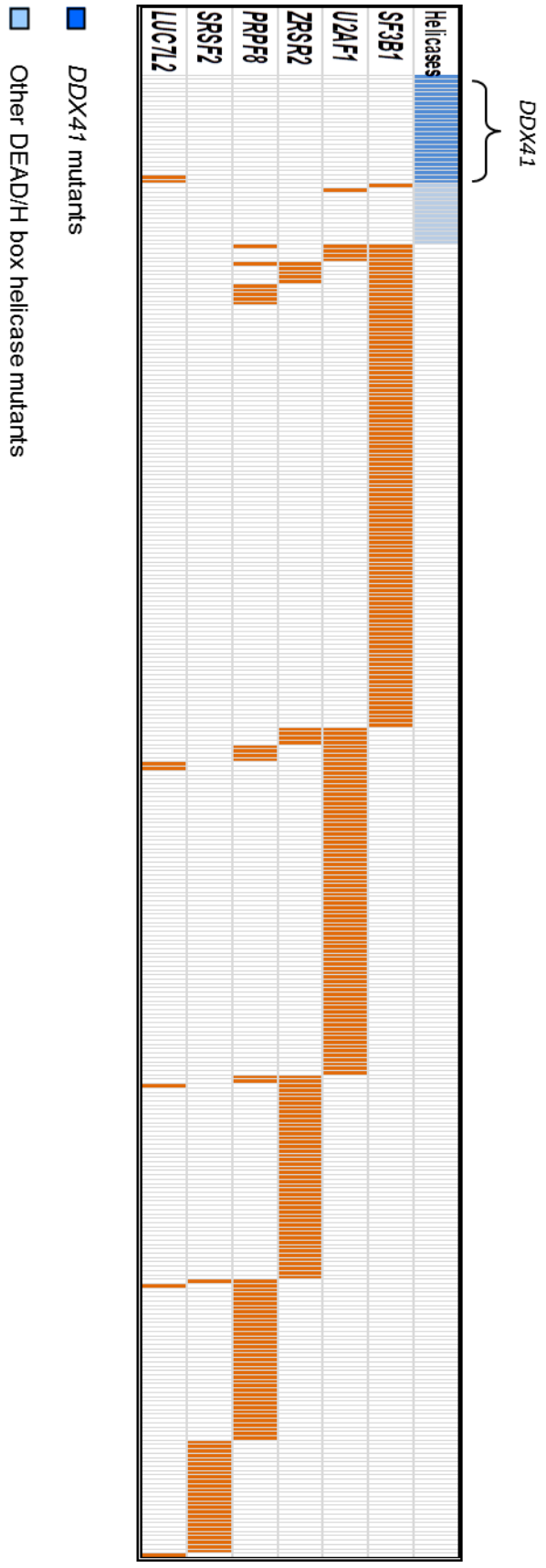
(H) *DDX41* deficiency in murine $lin^-/sca\text{-}1^+/c\text{-}kit^+$ (LSK) LSK cells. Left panel, representative Western blot analysis of DDX41 and β-Actin protein in infected cells 72 hours after infection. Colony formation potential (mean±SD) of purified murine LSK cells in the presence of SCF (100 ng/ml), IL-6 (10 ng/ml) and IL-3 (6 ng/ml) at 48 hours after infection with lentiviral *DDX41*-specific shRNA or control shRNA (NC-sh) constructs (middle panel). Right panel shows FACS analysis of c-kit and Gr-1 expression on murine LSK cells infected with *DDX41*-sh or NC-sh lentiviral shRNA constructs and passaged in culture in the presence of SCF (100 ng/ml), IL-6 (10 ng/ml) and IL-3 (6 ng/ml) for 7 days. Each bar represents the mean±SEM of 3 independent experiments performed in duplicates.

(I) Docking of the DDX3 helicase inhibitor (diaryl urea 1) in the RNA binding site of our computational model of DDX41. For the best docking pose of urea 1, the aniline group docks to a negatively charged pocket and forms a hydrogen bond with E480, whereas the methyl phenyl group interacts with a positively charged site of the binding pocket. The binding model indicates that urea 1 does not interact with DDX41 optimally but suggests several opportunities for structural modifications that may improve its potency against DDX41.

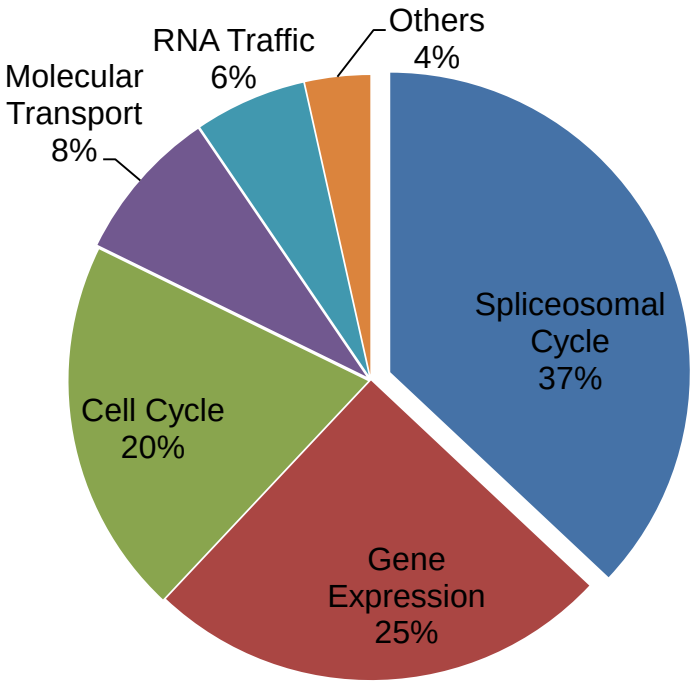
(J) Left panel; expression level of *DDX41* in leukemic cell lines K562 and HL60 as determined by normalization to *GAPDH* upon knockdown of *DDX41*.Right panel; sensitivity to helicase inhibitor, compound 8 ($C_{14}H_{15}N_3O$) of *DDX41* deficient cell lines (K562 and HL60) at 1 μM and 10 μM concentration was determined by *DDX41* mRNA expression levels and cell counts (*p<.05). Each bar represents the mean±SEM of 3 independent experiments performed in duplicates.

(K) Chemical structure of small molecule DDX3 inhibitor.

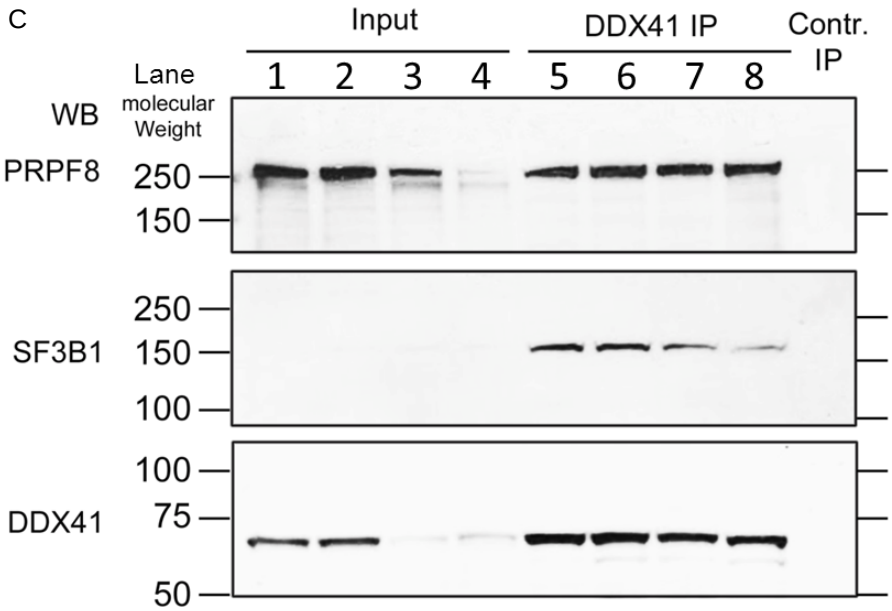
A


Helicases
SF3B1
U2AF1
ZRSR2
PRPF8
SRSF2
LUC7L2
DDX41
DDX41 mutants
Other DEAD/H box helicase mutants

B


Others
4%
RNA Traffic
6%
Molecular
Transport
8%
Spliceosomal
Cycle
37%
Cell Cycle
20%
Gene
Expression
25%

C


Input
DDX41 IP
Contr.
IP
Lane
1 2 3 4 5 6 7 8
WB
molecular
Weight
PRPF8
250
150
SF3B1
250
150
100
DDX41
100
75
50

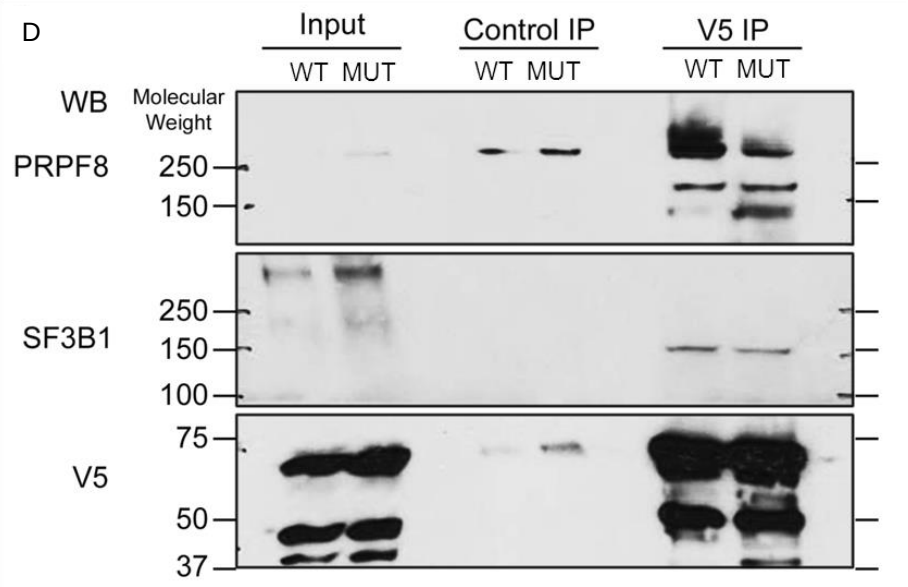


Figure S6: Related to Figure 5.

(A) Mutations of *DDX41*, other DEAD/H box helicases and spliceosomal genes (*U2AF1, SRSF2, ZRSF2, SF3B1, PRPF8* and *LUC7L2)* in our cohort (n=846) and TCGA database (n=197).

(B) Ingenuity based functional classification of the DDX41 proteome. DDX41 co-immunoprecipitated proteins were analyzed with Ingenuity software for functional pathway analysis. The relative frequencies of protein hits are assigned to functional categories. Overall, 96% of identified proteins clustered into five functional categories. Spliceosomal Cycle was the top functional group associated with DDX41.

(C) DDX41 co-immunoprecipitation was analyzed by Western blot (WB) using antibodies again PRPF8 (Top panel), SF3B1 (Middle panel), and DDX41 (Bottom panel). Lane 1-4 represented 5% of input from nuclear fractions of K562 cells obtained from individual cultures. Lane 5-8 represented DDX41 co-immunoprecipitates obtained from corresponding cultures. Control IP (Contr. IP) using normal mouse IgG are shown on the right side of each panel.

(D) V5 Tagged wild-type DDX41 (WT) and mutant (MUT) DDX41 were expressed in HEK293 cells. Nuclear fractions from both wild-type and mutant cells were used for V5 co-immunoprecipitation. The result was analyzed by Western blot (WB) using antibodies against PRPF8 (Top panel), SF3B1 (Middle panel), and DDX41 (Bottom panel). 5% of input of nuclear fractions of HEK293 cells are shown on the left side, control IP (Control IP) using normal mouse IgG are shown in the middle and DDX41 co-immunoprecipitation is shown on the right side of each panel.

Table S5: Related to Figure 5. Provided as an Excel file.

A

Allignment of RNAseq reads (50bp) to a special, splice specific junction genome (hg18) using Bowtie

Removal of low quality sites (exons having less than 10 quality reads in more than 30%)

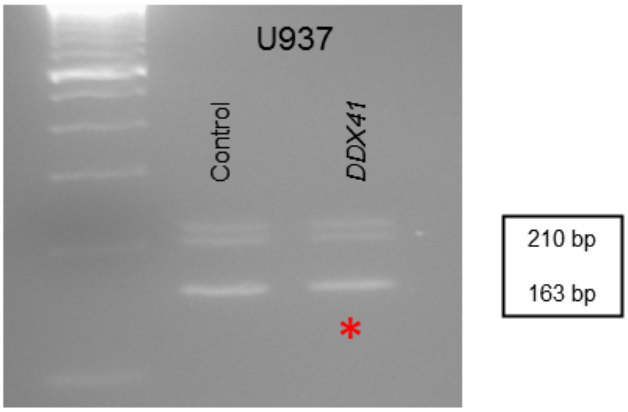
Quantification of valid, high quality reads (splicetrap software) and subsequent frequency estimation per each patient

Statistical evaluation of frequency differences between *DDX41* defects and WT using T-Test ($p<.05$)

Removal of results having small (<10%) frequency difference between *DDX41* defects and WT

B

C



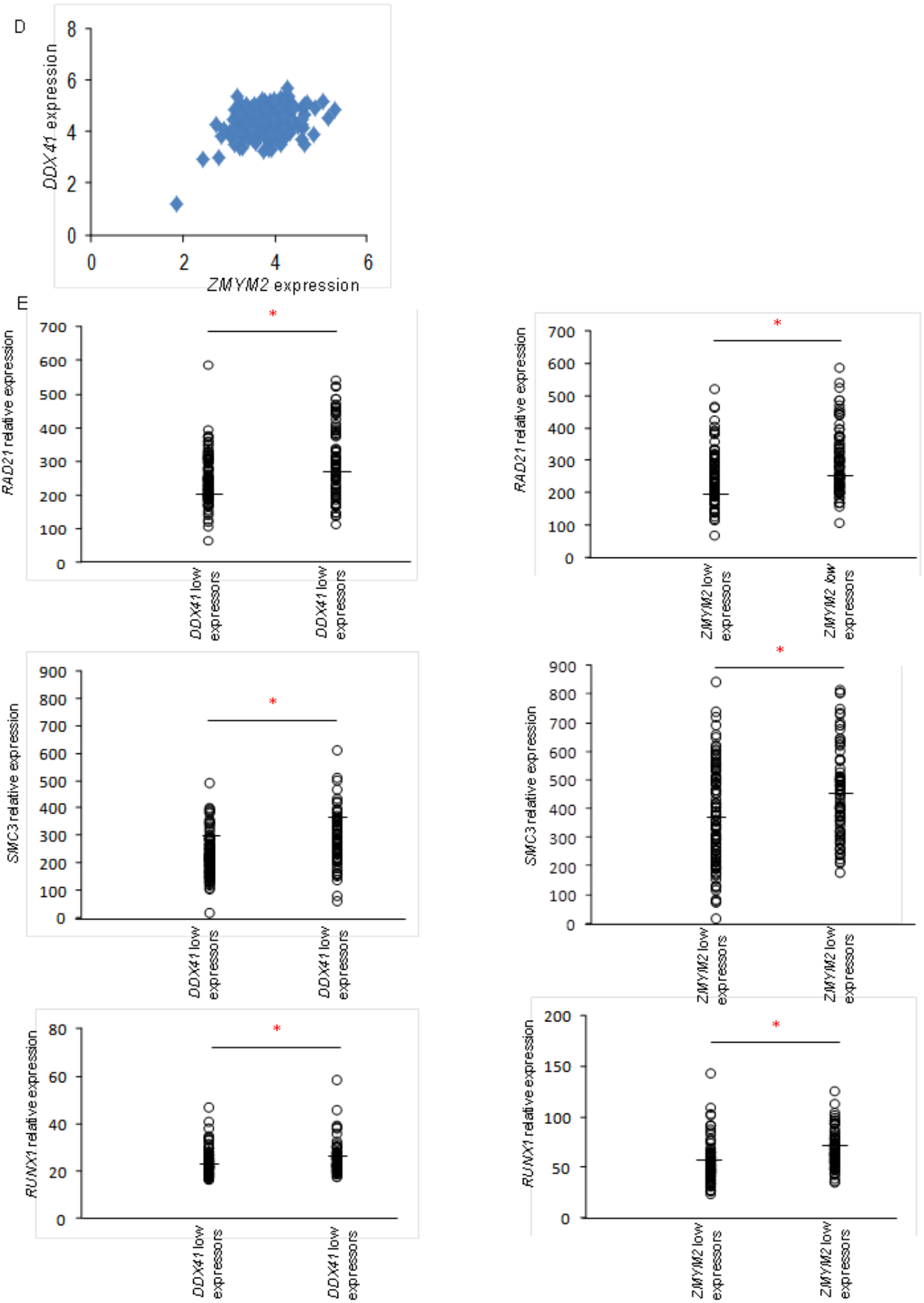

D
DDX41 expression
ZMYM2 expression
E
RAD21 relative expression
DDX41 low expressors
DDX41 low expressors
ZMYM2 low expressors
ZMYM2 low expressors
SMC3 relative expression
RUNX1 relative expression
*

Figure S7: Related to Figure 6.

(A) Work flow for the identification of candidate exons harboring differential exon usage between *DDX41* defects and WT.

(B) PCR demonstrated skipping and retaining of exon 3 in *ZMYM2* in *DDX41* knockdown of K562 and $CD34^+$ progenitor cells compared with control. Asterisks indicate denser bands (more skipped exon) in sh-*DDX41*.

(C) PCR demonstrated skipping and retaining of exon 3 in *ZMYM2* in *DDX41* overexpressing U937 cells. Asterisks indicate fainter bands (less skipped exon) in U937 cells with *DDX41* overexpression.

(D) Correlation of *ZMYM2* and *DDX41* expression levels. ($r=.38$, $p<.001$)

(E) Down-modulation of *ZMYM2* and *DDX41* target genes *SMC3*, *RAD21*, *RUNX1* is depicted by expression levels in both, *DDX41* and *ZMYM2,* low and high expressors (*$p<.05$, high and low expressors were dichotomized by the mean of relative value (shown as bar) of mRNA transcription levels).

Table S6: Related to Figure 6. List of misspliced genes from global differential pattern analysis.

Gene	Wild-type skipping (%)	*DDX41* defect skipping (%)	Differences between Wild-type and *DDX41* defect	p value
LASS4	51	79	28	0.04
ABI2	27	52	25	0.001
LRRC27	26	50	24	0.04
SLC35A1	12	43	31	0.01
CASP1	32	52	20	0.04
SSH1	25	43	18	0.002
IKZF1	32	49	17	0.006
THEM4	71	87	16	0.02
CSNK1D	75	91	16	0.004
OSBPL9	37	53	16	0.01
SCYL3	66	82	16	0.04
UBXN11	62	77	15	0.01
FGFR1OP	67	82	15	0.009
SFXN2	7	22	15	0.04
PHACTR4	84	99	15	0.01
AARSD1	18	32	14	0.04
SLC35B3	6	20	14	0.04
ANKMY1	1	15	14	<0.001
C16orf35	59	73	14	0.04
ZMYM2	61	74	13	0.02
ITGB3BP	77	40	37	0.03
PLB1	35	7	28	0.04
PILRB	41	14	27	0.001
RAG1AP1	37	12	25	0.005
PKP4	69	48	21	0.03
BTN3A3	70	49	21	0.005
MKL2	32	11	21	0.04
BRD8	38	18	20	0.01
L3MBTL	21	1	20	0.02
R3HDM1	92	73	19	0.02
DDB2	44	25	19	0.02
MLH3	38	19	19	0.02
METTL6	61	42	19	0.04
PPARA	22	4	18	0.03
SEC31A	67	49	18	0.01
CCNE1	23	5	18	0.03
MTRR	84	66	18	0.04
SRPK2	27	10	17	0.02
SCARF1	36	19	17	0.007
LYPLA1	27	11	16	0.005

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

NGS STUDIES

Whole exome capture was accomplished based on liquid-phase hybridization of sonicated genomic DNA having 150 – 200 bp of mean length to the bait cRNA library synthesized on magnetic beads (SureSelect®, Agilent Technology), according to the manufacturer's protocol. The captured targets were subjected to massive sequencing using Illumina HiSeq 2000 with the pair end 75-108 bp read option, according to the manufacturer's instruction. 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, University of Tokyo, which are summarized in a previous report (Yoshida et al., 2011) The data processing was divided into two steps, i) Generation of a bam file (http://samtools.sourceforge.net/) for paired normal and tumor samples for each case, ii) Detection of somatic point mutations and indels by comparing normal and tumor BAM files. Alignment of sequencing reads on hg19 was visualized using Integrative Genomics Viewer (IGV) software (http://www.broadinstitute.org/igv/) (Robinson et al., 2011). Each potential mutation was compared against databases of known SNPs, including Entrez Gene (http://www.ncbi.nlm.nih.gov/gene) and the Ensembl Genome Browser (http://useast.ensembl.org/index.html).

TARGETED DEEP SEQUENCING

To detect allelic frequencies for mutations or SNPs, we applied deep sequencing to targeted exons as previously described (Yoshida et al., 2011). Briefly, we screened for possible mutations of *DDX41* and other genes that were concomitantly mutated in the cases with *DDX41* mutation (*DNMT3A, PHF6, JAK2, RUNX1, TP53, CDH26, CUX1, SUZ12, BCOR, SMC3* and *LUC7L2*). Each targeted exon was amplified with NotI linker attached to each primer as previous described (Yoshida et al., 2011). After digestion with NotI, amplicons were ligated with T4 DNA ligase and sonicated into fragments that were on average up to 200 bp in size using Covaris. Sequencing libraries were generated according to an Illumina paired-end library protocol and were subjected to deep sequencing on the Illumina Genome Analyzer IIx or Hiseq 2000 sequencers according to the standard protocol.

List of 62 genes for targeted deep sequencing.

Genes	ID	Genes	ID
APC	NM_001127511	*KDM6A*	NM_021140
ASXL1	NM_015338	*KIT*	NM_000222
BCOR	NM_001123384	*KRAS*	NM_004985
BCOR1L1	NM_021946	*MECOM*	NM_001105078
BTRC	NM_001256856	*MED12*	NM_005120
LUC7L2	NM_016019	*MLL*	NM_001197104
CALR	NM_004343.3	*NF1*	NM_000267
CBL	NM_005188	*NPM1*	NM_199185
CDH23	NM_001171930	*NRAS*	NM_002524
CEBPA	NM_004364	*OGT*	NM_181672
CFTR	NM_000492	*PHF6*	NM_001015877
CSF1R	NM_005211	*PRPF8*	NM_006445
CSF3R	NM_156039.2	*PTCH1*	NM_001083603
CUX1	NM_001202543	*PTPN11*	NM_002834
DDX41	NM_016222	*RAD21*	NM_006265.2
DDX54	NM_001111322	*RNF25*	NM_022453.2
DHX29	NM_019030	*RUNX1*	NM_001754.2
DNMT3A	NM_153759	*SETBP1*	NM_015559
EED	NM_003797	*SF3B1*	NM_012433
ERBB4	NM_001042599	*SIMC1*	NM_198567
ETV6	NM_001987	*SMC3*	NM_005445
EZH2	NM_001203249	*SRSF2*	NM_001195427
FLT3	NM_004119	*STAG2*	NM_006603
GATA2	NM_001145662	*STAT3*	NM_003150
GLI1	NM_005269.2	*SUZ12*	NM_015355
GLI2	NM_005270	*TET2*	NM_001127208
GNB1	NM_002074	*TP53*	NM_001126115
IDH1	NM_005896.2	*U2AF1*	NM_001025203
IDH2	NM_002168.2	*U2AF2*	NM_001012478
IRF4	NM_001195286	*WT1*	NM_000378
JAK2	NM_004972	*ZRSR2*	NM_005089

SANGER SEQUENCING AND ALLELE-SPECIFIC PCR

Exons of selected genes were amplified and underwent direct genomic sequencing by standard techniques on the ABI 3730Xl DNA analyzer (Applied Biosystems) as previously described (Dunbar et al., 2008; Jankowska et al., 2009; Makishima 2011). When a mutant allele with small burden was not confirmed by Sanger sequencing, cloning and sequencing of individual colonies (TOPO TA cloning, Invitrogen) was performed for validation. All mutations were

detected by bidirectional sequencing and were scored as pathogenic if not present in non-clonal paired DNA from CD3-selected cells (or buccal swab DNA). The allelic presence of p.Arg525His alteration was determined by allele-specific PCR.

Primers for *DDX41* sequencing and PCR specific for mutant alleles

DDX41 sequencing	
Exon3_Forward	GACCGACGGCTTGATCTG
Exon3_Reverse	CCTTCTCTTTAAGGTGCTGGT
Exon5_Forward	CAGGCATGTTATATCTCAAGGGA
Exon5_Reverse	GATGGGTAACAGGGATCAAGAG
Exon6_Forward	GTATGTCTGTACAGTCTGCAGTT
Exon6_Reverse	CCTTGAAGCTCTTGATGGGT
Exon7_Forward	GGCTCTGAGAAAGTACCTGTG
Exon7_Reverse	CAGAAGATGAAGGACACCTAGC
Exon8_Forward	TGAGGAGGGCTGGAACA
Exon8_Reverse	CGCGCTTTGAGAAGGGTAA
Exon10_Forward	TGCCTGCTTGCCTCTAGATA
Exon10_Reverse	TACGGATGTCACCCTCGAA
Exon11_Forward	TGTCTCAGTTGCTCAGCTTC
Exon11_Reverse	AATCAGCTTCAGGGAGACTTG
Exon15_Forward	AGAGACTCTGTCCTTCTCTCTG
Exon15_Reverse	CACCTTCTGCTTGGCTTCTA
Exon11-15_Forward	CCAACCCACTGCTCATACTT
Exon11-15_Reverse	CACCTTCTGCTTGGCTTCTA
Allele specific PCR (p.Arg525His)	
*DDX41*_R525H_Forward_out	GGGAANCATCAGGGCCCATCCTGGGCTC
*DDX41*_R525H_Reverse_out	GCCTGGACTTCCCTGCCATCCAGCACGT
*DDX41*_R525H_Forward_in_C	GTAGTGGCGATGCCTGTGTTTCCCGCGC
*DDX41*_R525H_Reverse_in_T	AGTACACCGGATTGGCCGCACCGGTCA

List of genes and their annotation screened by Sanger sequencing

Genes	Ensembl cDNA sequence	Amino acid	Sequenced exons
TP53	ENSG00000141510	302	4
DNMT3A	ENSG00000119772	723	19
CDH26	ENSG00000124215	832	8
PHF6	ENSG00000156531	312	2
RUNX1	ENSG00000159216	453	1
JAK2	ENSG00000096968	1132	9

PROLIFERATION ASSAYS

Cell growth was determined by counting cell numbers in culture. Briefly, 1 mL of $5x10^4$ cells were cultured in IMDM containing 10% fetal bovine serum in the six-well plates at day 0, and the cell numbers were scored by Trypan blue exclusion at 24, 48, 72, 96 and 120 hours.

APIGETRIN INDUCED ERYTHROID DIFFERENTIATION (Tsolmon et al., 2011)

K562 cell line was transfected with lentiviruses carrying control shRNA, or *DDX41*-targeting shRNA, and cultured in IMDM medium supplemented with 10%v/v heat-inactivated fetal bovine serum at 37°C in a humidified 5% CO_2 incubator. Erythroid differentiation was induced by the addition of 75 μM apigetrin (Sigma) to both the control and sh-*DDX41* K562 cells ($2x10^4$ cells/mL). Medium and apigetrin (75 μM) were renewed every 3 days. DMSO-dissolved apigetrin at 100mM was stored at -20˚C and was freshly diluted in culture medium immediately before use.

BENZIDINE STAINING (Murphy, M.J. Jr. 1978)

Benzidine (Sigma) was prepared at 0.2% in 0.5 M acetic acid. The staining reagent consisted of 100 mL of the benzidine solution, to which 0.4 mL of 30% hydrogen peroxide was added just prior to use. One mL of this reagent was added to each petri dish and 5 minutes later, the dish was scored for the number of cells which were uniformly benzidine-unreactive (colorless), and uniformly benzidine-reactive (blue).

APOPTOSIS ASSAYS

For analysis of apoptotic cells, 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).

CELL SORTING

Following lysis of red blood cells, immunomagnetic selection of cells was performed using anti-CD34-FITC (Life Technologies) followed by anti-FITC microbeads (Miltenyi Biotec). Samples were separated using LS Columns (Miltenyi Biotec), and purity was verified in each fraction by flow cytometry on a Beckman Coulter FC500.

GLOBAL DIFFERENTIAL SPLICING PATTERN ANALYSIS

We quantified exon inclusion ratios based on paired-end RNAseq data. SpliceTrap software (http://rulai.cshl.edu/splicetrap/) was used to quantify the frequency of inclusion of each exon (Wu et al., 2011) and extract counts of paired end reads that span each exon junction in the genome. For this purpose, each exon was tested for inclusion or exclusion with respect to adjacent exons. SpliceTrap considers individual exons in whole genomes and is not limited by analysis of known repository of transcripts. This unbiased method is a suitable approach for possible novel discovery of unknown/unexpected splicing variants. Each exon was tested with respect to adjacent exons. Within each triplet, each exon was labeled as A, B, and C; exon B being the one screened for every triplet in the transcriptome. According to this method, we counted reads spanning between exon A/B, B/C, and A/C, where reads spanning the A/C junction reflect the proportion of mRNA missing exon B. The sum of reads between exons A/B and B/C divided by 2 reflects the proportion of mRNA that contains exon B. In order to estimate the frequency of exon B skipping, we divided the number of reads spanning A/C by half of the sum of the reads spanning A/B and B/C. By following these guidelines, we extracted alternative splicing patterns for 16 patients (5 *DDX41* defect patients and 11 spliceosomal WT patients). Using the frequency of skipped reads to represent the skipping ratio is independent from variation in coverage between

different RNAseq samples and is a normalization step itself. The unpaired *t* test was used to assess the difference of exon usage between these 2 groups. For each exon tested we compared average exon usage between *DDX41* defects and the WT group, with associated p values generated. Statistical difference of $p<.05$ and average difference of ± 10% in frequency of exon usage was considered valid for an exon tested. Using this approach, we detected changes in exon skipping (excess of shorter mRNA missing an exon) as well as in exon retention (excess of longer mRNA incorporation an exon).

GENE KNOCKDOWN IN PURIFIED MOUSE LIN$^-$SCA-1$^+$C-KIT$^+$ (LSK) CELLS

LSK cells are purified from C57BL/6 mice as previously described (Oakley et al. 2012). After culturing in DMEM with addition of 15% fetal bovine, murine SCF (100 ng/ml), IL-6 (6 ng/ml), and IL-3 (3 ng/ml) for 24 hours, LSK cells were infected twice (once daily) with pLKO.1 lentivirus containing shRNAs targeting *Ddx41* (Targeting sequences*:* sh1*,* 5'-GCCAAGATGGTGTACTTGCTT-3'; sh2, 5'- GCATCACCTATGACGATCCAA-3') or negative control shRNA (Sigma, SHC002). To generate infectious lentivirus, the corresponding constructs were co-transfected using Fugene 6 into 293T cells along with packaging plasmid Δ8.9 and a plasmid expressing VSV-G, and virus particles were harvested at 72 hours after transfection. Viral titers were calculated by infecting NIH-3T3 cells with serial dilutions of viral stocks and isolating puromycin resistant colonies. Lentiviral infections were performed twice by spinoculation in which a mixture of lentivirus and target cells at 4:1 ratio in 48-well or 24-well plates were centrifuged at 2000 x g for 90 minutes at 37°C. Puromycin (2 µg/ml) was added to the infected cells 24 hours after the second infection. Colony formation assays were performed after another 24 hours using 2 x10^4 puromycin resistant cells on IMDM methylcellulose medium supplemented with 15% horse serum, murine SCF (100 ng/ml), IL-6 (6 ng/ml), IL-3 (3 ng/ml) and puromycin (2 µg/ml). Colony numbers were counted after 7 days.

LENTIVIRAL-MEDIATED DDX41 WT AND MUTANT OVEREXPRESSION

QuikChange site directed mutagenesis kit (Agilent Technologies) was used to generate R525H mutant *DDX41* gene construct. In brief, the primer pair for generating the R525H mutation was designed using the web-based primer design program (www.agilent.com/genomics/qcpd) and the primers obtained from IDT technologies (***DDX41*-R525H-F:** CGCACCGGGCACTCGGGAAAC, ***DDX41*-R525H-R:** GCCAATCCGGTGTACATAGTTCTC). The mutagenesis PCR and the subsequent steps to generate R525H *DDX41* mutant vector were performed according to the kit instructions. Lentiviral expression vector expressing a C-terminal V5 epitope tagged wild-type or R525H mutant *DDX41* gene in pLX304 vector (Clone ID: HsCD00442077; DNASU Plasmid Repository) was used to generate the lentiviral supernatants in 293T cells as described above. The viral supernatant was used for infecting HEK293 cells in the presence of 8 µg/mL polybrene for 24 hours. Subsequently, cells were selected with 10 µg/ml of blasticidin for a week to obtain stably transfected cells overexpressing *DDX41*.

CELL FRACTIONATION AND NUCLEAR PROTEIN EXTRACTION

V5 tagged *DDX41* was expressed in HEK293 cells. Approximately 5x10^7 of HEK293 cells were transferred to 15 ml conical tubes and washed twice with 10 ml of ice-cold 1× PBS that contained protease inhibitors (Sigma, A8340). Cells were resuspended in 500 µl of 1x hypotonic buffer containing 10 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol, 10 mM PMSF, and protease inhibitors (Sigma, A8340). A total of 20 µl of 10% Nonidet P-40 was added to cell suspensions to break the cell membrane. After 5 minutes incubation on ice, cell suspensions were centrifuged at 344 *x g* for 10 minutes. The supernatant was transferred to clean tubes. Nuclear pellets were washed

twice with ice-cold PBS, and resuspended in 100 µl of 50 mM Tris-HCl, pH 8.0, 1 mM $MgCl_2$, 10 mM PMSF, protease inhibitor mixture (Sigma, A8340). Benzonase (Sigma, D5915, 250 units) was added to fragment DNA and RNA. The nuclear suspensions were incubated on ice for 90 minutes with vigorous vortexing every 5 minutes. At the end of the incubation, 500 µl of protein extraction buffer containing 1.5% Nonidet P-40, 500 mM NaCl, 5 mM dithiothreitol, 10 mM PMSF, and 5 µl of protease inhibitor mixture (Sigma, A8340) in 50mM phosphate buffer, pH 7.4, were added. After 30 minutes incubation on ice with vortexing every 5 minutes, the mixture was centrifuged at 12,396 × *g* for 15 minutes. The same extraction process was repeated twice with 300 and 200 µl of extraction buffer, respectively. The supernatant containing nuclear proteins was combined and transferred to clean tubes, and the protein concentration was determined by BCA assay.

COVALENT BOUND ANTIBODY TO PROTEIN G BEADS

Rabbit anti-V5 and control rabbit IgG were covalently coupled to Sepharose-protein G beads using dimethylpimelimidate. Briefly, 25 mg of protein G-Sepharose CL-4B (GE Healthcare, 17-0780-01) was swelled in 1 ml of PBS overnight and incubated with 200 µl of antibody (50 µg) solution (1× PBS) for 1 hour at room temperature. Antibody bound protein G beads were then incubated with 1% chicken egg ovalbumin in PBS for another hour to block nonspecific binding sites. After 3 washes with PBS, 25 mg of dimethylpimelimidate in 1 ml of 200 mM triethanyl amine was added, and the coupling reaction proceeded at room temperature for 30 minutes. The reaction was repeated 2 more times with fresh addition of dimethylpimelimidate and quenched with 50mM ethanol amine. The reacted protein G beads were washed extensively with PBS before immunoprecipitation.

PROTEIN IDENTIFICATION BY LC-MS/MS

Anti-V5 and isotype antibody immunoprecipitation products were subjected to SDS-polyacrylamide gel electrophoresis and stained with colloidal Coomassie Blue (Gel Code Blue, Pierce Chemical). Gel slices were excised from the top to the bottom of the lane; proteins were reduced with dithiothreitol (10 mM), alkylated with iodoacetamide (55 mM), and digested *in situ* with trypsin. Peptides were extracted from gel pieces 3 times using 60% acetonitrile and 0.1% formic acid/water. The dried tryptic peptide mixture was re-dissolved in 15 µl of 0.1% formic acid and 5% acetonitrile for mass spectrometric analysis. Tryptic peptide mixtures were analyzed by on-line LC-coupled tandem mass spectrometry (LC-MS/MS) on a Orbitrap mass spectrometer (Theomo), a 0.3 × 5-mm trapping column (C18 PepMap 100, LC Packings), a reverse phase separating column (75 µm × 5 cm, Vydac C18), and a flow rate of 250 nl/min. Gradient LC separation was achieved with aqueous formic acid/acetonitrile solvents. The Orbitrap mass spectrometer was operated in standard MS/MS switching mode with the 4 most intense ions in each survey scan subjected to MS/MS analysis. Instrument operation and data acquisition used Theomo Xcalibur software. Initial protein identifications from MS/MS data used the Mascot search engine and the Swiss Protein Human sequence database. The Swiss Protein database search parameters included the allowed 2 missed tryptic cleavage sites, precursor ion mass tolerance of 10ppm, fragment ion mass tolerance of 0.6 Da, protein modifications for Met oxidation, and Cys carbamidomethylation. A minimum Mascot ion score of 25 was used for automatically accepting all peptide MS/MS spectra. A Mascot-integrated decoy database search calculated a false discovery of ≤3.55% when searching was performed on the concatenated mgf files with an ion score cut-off of 25 and a significance threshold of $p \leq .01$. Only peptides with ion scores of ≥ 25 and only proteins with at least two unique peptide ranked as a top candidate (bold red in Mascot) were accepted.

PROTEIN NETWORK ANALYSIS

The “core analysis” function included in IPA (Ingenuity System Inc.) was used to interpret the data in the context of biological processes, pathways, and networks. All identifier types were selected because more than one type of identifier exists in our dataset (working file). Only direct interaction was chosen for analysis.

SUPPLEMENTAL REFERENCES

Cheson, B.D., Bennett, J.M., Kantarijian, H., Pinto, A., Schiffer, C.A., Nimer, S.D., Löwenberg, B., Beran, M., Witte, T.M., Stone, R.M., et al. (2000). Report of an international working group to standardize response criteria for myelodysplastic syndromes. Blood. *96*, 3671-4.

Dunbar, A.J., Gondek, L.P., O'Keefe, C.L., Makishima, H., Rataul, M.S., Szpurka, H., Sekeres, M.A., Wang, X.F., McDevitt, M.A., Maciejewski, J.P. (2008). 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of *c-Cbl,* in myeloid malignancies. Cancer Res. *68*, 10349-10357.

Jankowska, A.M., Szpurka, H., Tiu, R.V., Makishima, H., Afable, M., Huh, J., O'Keefe, C.L., Ganetzky, R., McDevitt, M.A., Maciejewski, J.P. (2009). Loss of heterozygosity 4q24 and *TET2* mutations associated with myelodysplastic/myeloproliferative neoplasms. Blood 113, 6403-6410.

Makishima, H., Jankowska, A.M., McDevitt, M.A., O'Keefe, C., Dujardin, S., Cazzolli, H., Przychodzen, B., Prince, C., Nicoll, J., Siddaiah, H., et al. (2011). *CBL, CBLB, TET2, ASXL1,* and *IDH1/2* mutations and additional chromosomal abberations constitute molecular events in chronic myelogenous leukemia. Blood *117*, e198-e206.

Murphy, M.J. Jr. (1978). In Vitro Aspects of Erythropoiesis (New York: Springer-Verlag).

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24-26.

Tsolmon, S., Nakaziki, E., Han, J. & Isoda, H. (2011). Apigetrin induces erythroid differentiation of human leukemia cells K562: Proteomics approach. Mol. Nutr. Food Res. *55*, S93-S102.

Wu, J., Akerman, M., Sun, S., McCombie, W.R., Krainer, A.R., Zhang, M.Q. (2011). SpliceTrap: a method to quantify alternative splicing under single cellular conditions. Bioinformatics *27*, 3010-3016.