

**RECURRENT MUTATIONS IN THE *U2AF1* SPLICING FACTOR
IN MYELODYSPLASTIC SYNDROMES**

Timothy A. Graubert^{*1,3,6}, Dong Shen^{*4}, Li Ding^{*2,4}, Theresa Okeyo-Owuor¹, Cara L. Lunn¹, Jin Shao¹, Kilannin Krysiak¹, Christopher C. Harris⁴, Daniel C. Koboldt⁴, David E. Larson⁴, Michael D. McLellan⁴, David J. Dooling⁴, Rachel M. Abbott⁴, Robert S. Fulton⁴, Heather Schmidt⁴, Joelle Kalicki-Veizer⁴, Michelle O’Laughlin⁴, Marcus Grillot¹, Jack Baty⁵, Sharon Heath¹, John L. Frater⁶, Talat Nasim^{7,8}, Daniel C. Link^{1,3}, Michael H. Tomasson^{1,3}, Peter Westervelt^{1,3}, John F. DiPersio^{1,3}, Elaine R. Mardis^{2,3,4}, Timothy J. Ley^{1,2,3,4}, Richard K. Wilson^{2,3,4}, and Matthew J. Walter^{1,2,3}

¹Department of Internal Medicine, Division of Oncology, Washington University, St. Louis, MO, USA

²Department of Genetics, Washington University, St. Louis, MO, USA

³Siteman Cancer Center, Washington University, St. Louis, MO, USA

⁴The Genome Institute, Washington University, St. Louis, MO, USA

⁵Division of Biostatistics, Washington University, St. Louis, MO, USA

⁶Department of Pathology and Immunology, Washington University, St. Louis, MO, USA

⁷Department of Medical and Molecular Genetics, King’s College, Guy’s Hospital, London, UK

⁸National Institute for Health Research (NIHR), Biomedical Research Centre, Guy’s and St. Thomas’ NHS Foundation Trust and King’s College London, UK

*these authors contributed equally.

TABLE OF CONTENTS

A. Supplementary Note

B. Supplementary Figures

- Supplementary Figure 1. Survival analysis in transplanted vs. non-transplanted patients.*
- Supplementary Figure 2. Genomic architecture of UPN 266395 index MDS patient.*
- Supplementary Figure 3. U2AF1 S34F mutation induces splicing alterations in FMR1.*
- Supplementary Figure 4. U2AF1 quantitative RT-PCR and global expression profiling.*

C. Supplementary Tables

- Supplementary Table 1. Sequence metrics.*
- Supplementary Table 2. Tier 1-3 validated SNVs.*
- Supplementary Table 3. Tier 1 SNV annotation.*
- Supplementary Table 4. Annotation of U2AF1 mutations.*
- Supplementary Table 5. Clinical characteristics of MDS patients with U2AF1 mutations.*
- Supplementary Table 6. Significantly dysregulated genes in U2AF1 mutant MDS samples vs. controls.*
- Supplementary Table 7. Clinical characteristics of MDS replication cohort.*
- Supplementary Table 8. Primers used for U2AF1 sequencing.*
- Supplementary Table 9. Microarray-based expression of mutated genes.*

D. Supplementary References

SUPPLEMENTARY NOTE

Clinical history of index case (UPN 266395). The patient was a 65 year old male of European ancestry who presented with anemia and thrombocytopenia (white blood cell count 5,900/mcl, hemoglobin 9.6 g/dl, platelet count 53,000/mcl). A bone marrow biopsy and aspirate showed 60% cellularity with trilineage hematopoiesis, without excess blasts or overt dysplasia. His marrow was felt to be abnormal, but not diagnostic of a myelodysplastic or myeloproliferative disorder. Cytogenetics showed a 46, XY karyotype. He was initially treated with recombinant erythropoietin without improvement, and two months later was referred for evaluation with the following blood counts: white blood cell count 40,300/mcl, hemoglobin 9.4 g/dl, platelet count 33,000/mcl. A bone marrow biopsy and aspirate were repeated (including specimens obtained for tissue banking after obtaining informed consent with specific language authorizing whole genome sequencing which was approved by the Washington University Human Research Protection Office), and showed a 100% cellular marrow, with trilineage dysplasia and 7% blasts, interpreted as MDS/RAEB-1. Cytogenetics again showed a normal male karyotype. His IPSS score was 1.5 (Intermediate-2). At that time, erythropoietin was continued and hydroxyurea was added (1500 mg daily).

One month later, the patient presented with increased fatigue, blurred vision, and dyspnea and was found to have a white blood cell count of 147,000/mcl, hemoglobin 8.9 g/dl, platelet count 36,000/mcl, with 27% circulating blasts. A bone marrow biopsy and aspirate were repeated (including specimens obtained for tissue banking after obtaining informed consent with specific language authorizing whole genome sequencing which was approved by the Washington University Human Research Protection Office). The bone marrow biopsy showed >90% cellularity, with 60% blasts with irregular nuclear contours (NBE positive, largely myeloperoxidase negative). Flow cytometry demonstrated coexpression of CD13 and CD33, with variable expression of CD64 and CD117, most consistent with AML M5b (AML with monocytic differentiation). He received emergency leukopheresis and began 5-azacytidine (160 mg subcutaneously, days 1-7 of 28), for a total of 6 cycles.

Seven months later, the patient developed mental status changes and was found to have leukemic relapse in the central nervous system. He received supportive care until his death 322 days after initial presentation, and 246 days after evolution to sAML.

The patient had a prior history of coronary artery disease, hyperlipidemia, and recurrent pneumonia. There was no significant family history of cancer.

RESULTS

Genome coverage. Individual reads were aligned to the reference genome (NCBI build 36) and only unique reads (deduplicated reads) were used to determine haploid coverage of the genome for the normal and sAML samples (38.23x and 39.12x, respectively, **Supplementary Table 1**). We obtained at least 32x haploid coverage for all autosomal chromosomes in the normal genome and ~half this coverage for chromosomes X and Y, as expected (**Supplementary Fig. 2**). At least 99.42% of the consensus coding DNA sequence (CCDS) was covered by at least 1 read in both genomes (**Supplementary Table 1**). We identified all the heterozygous and homozygous SNPs present in the normal and sAML genomes using the Affymetrix SNP 6.0 arrays and determined diploid coverage for each genome based on the number of these SNPs that were identified using the WGS data, as previously described¹. Using this approach, we achieved 98.91% and 99.31% diploid coverage of the normal and sAML genomes, respectively (**Supplementary Table 1**).

Detection and validation of somatic single nucleotide variants and insertions/deletions. 1281 potential somatic single nucleotide variants (SNVs) and 8801 indels were identified in the

sAML genome after filtering out low quality data and variants present in the matched normal genome and previously identified SNPs. These variant calls were prioritized into non-overlapping tiers for downstream validation, as previously described¹. 1233 SNV predictions were tested using the custom capture approach, of which 507 validated (**Supplementary Table 2**). This represents an overall validation rate of 507/1233 (41%) and 16/38 (42%) for tier 1 high confidence (HC). 16 indel predictions were validated (none in Tier 1). 505 of the 507 SNVs (99.6 %) were present in the MDS genome, and 2 were new in the sAML genome (sAML-specific) (**Supplementary Table 2**).

Copy number alterations and loss of heterozygosity. Using WGS data, we identified 42 putative amplifications. None of these calls were supported by LOH analysis and there were no instances of copy number neutral LOH (**Supplementary Fig. 2**). 4 putative copy number alterations were detected using the Affymetrix SNP 6.0 array data (SNPa). Deep sequencing after solid phase capture of the called sites in the normal, MDS, and sAML samples did not support any of the copy number predictions. Therefore, no confirmed structural variants were detected in this genome.

Luciferase- β -galactosidase double-reporter assay. Transient co-expression of the pTN24 double-reporter plasmid and the wild-type or S34F mutant *U2AF1* cDNA, with or without the Tra2 α splicing enhancer or hnRNPG splicing inhibitor, was performed in 293T cells in 3 independent experiments, with one representative example shown in **Fig. 3a**. Each experiment had 6 replicate data values for each combination of conditions. A significant increase in splicing (as detected by an increase in the luciferase/ β -galactosidase ratio) occurred with expression of the mutant *U2AF1* compared to wild-type *U2AF1* when co-expressed with pTN24 in 3 experiments ($P \leq 0.004$). Co-expression of the Tra2 α splicing enhancer or hnRNPG splicing inhibitor resulted in increased and decreased splicing of the pTN24 reporter plasmid, as expected, that was independent of wild-type or mutant *U2AF1* expression. Experiments were analyzed independently. Transient co-expression of a control plasmid (vector) or the mutant *U2AF1* with or without a siRNA targeting the endogenous *U2AF1* was performed in 3 independent experiments, with one representative example shown in **Fig. 3b**. Each experiment had 6 replicate data values for each combination of conditions. A significant increase in splicing (as detected by an increase in the luciferase/ β -galactosidase ratio) occurred in all 3 experiments when comparing the vector and *U2AF1* mutant cDNAs, independent of knockdown of *U2AF1* ($P \leq 0.04$).

Alternative splicing of *FMR1*. The *FMR1* gene, fragile X mental retardation 1, contains several alternative cryptic 3' splice acceptor sites in exon 15². We performed RT-PCR using primers that flanked the cryptic splice sites and measured the relative abundance of splice acceptor site usage, as previously described². We used RNA extracted from 7 unfractionated bone marrow samples harvested from MDS patients with a *U2AF1* mutation and 5 samples from MDS patients without a *U2AF1* mutation. Two independent experiments were performed using cDNA obtained from these samples and results are shown in **Supplementary Fig. 3a**. The splicing ratio is defined as the ratio of mRNAs resulting from the use of an alternative 3' cryptic splice acceptor site located within exon 15 (resulting in 124 or 85 base pair amplicons) compared to the normal splice acceptor site (160 base pair amplicon). There was an increased utilization of a cryptic 3' splice acceptor site in MDS samples harboring a *U2AF1* mutation compared to MDS samples without *U2AF1* mutations ($p < 0.03$ for both independent experiments, $p = 0.0001$ for pooled data).

Next, we validated this result by creating a *FMR1* minigene splicing reporter construct using the same genomic locus interrogated in the clinical samples. Each experiment was repeated with four biologic replicates. The *FMR1* minigene was transiently transfected into 293T cells with a control plasmid (pcDNA3.1-YFP), wild-type *U2AF1* cDNA, or mutant (S34F)

U2AF1 cDNA in the presence of a control siRNA or siRNA targeting the endogenous *U2AF1*. RNA was harvested 48 hours later and a reverse transcriptase (RT) reaction was performed to create cDNA. Expression of the S34F mutant *U2AF1* resulted in an increase in cryptic 3' splice acceptor usage compared to control or wild-type *U2AF1*, independent of endogenous *U2AF1* levels ($P < 0.04$) (**Supplementary Fig. 3b**).

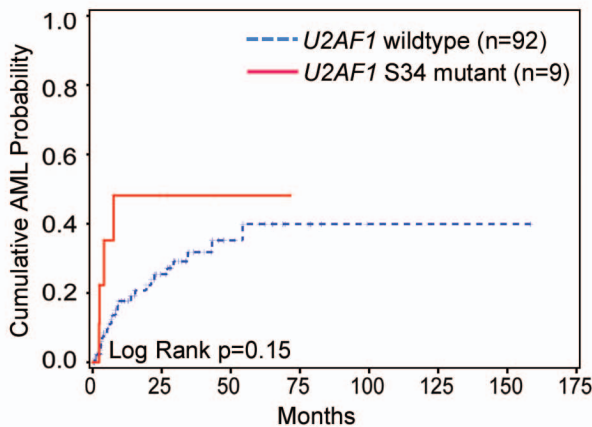
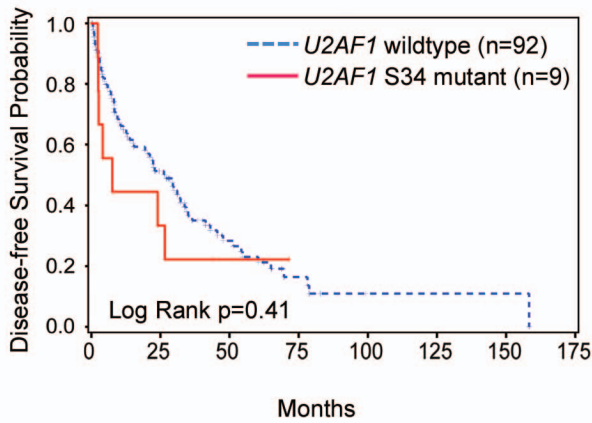
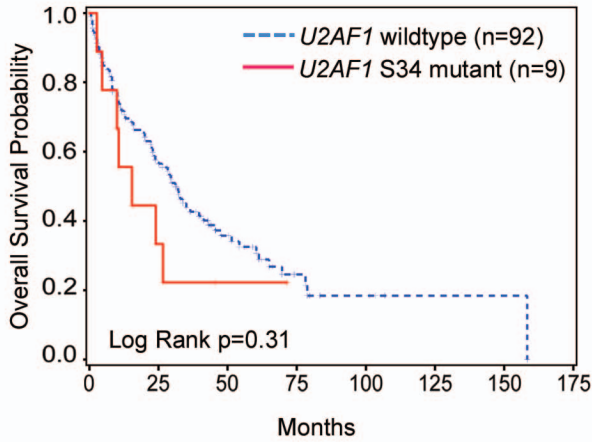
RT-PCR of *U2AF1*. Quantitative RT-PCR of *U2AF1* revealed no difference in the expression levels between mutant and wild-type MDS samples ($p = 0.73$, **Supplementary Fig. 4a, left panel**). A dominant RT-PCR product was identified for both wild-type and mutant *U2AF1*, consistent with isoform “a” expression (**Supplementary Fig. 4a, right panel**). Cloning and sequencing of RT-PCR products obtained from *U2AF1* mutant MDS samples UPN 300813 and UPN 571656 revealed isoform “a” by Sanger sequencing (data not shown) and next-generation sequencing (**Fig. 1c**).

Gene expression profiling. We compared the expression level of the 30 genes containing tier 1 mutations in our index sAML sample (UPN 266395) with the expression in 43 *de novo* AML samples for which exon array data was available (T. Ley, unpublished data). 28 genes had at least 3 probesets on the array. For these 28 genes, at least one probeset was called present in 27 genes in the sAML case (**Supplementary Table 9**). All 27 genes were also expressed in >50% of the *de novo* AML cases and the one gene not expressed in the sAML case (*POPDC3*) had only 1/13 probes expressed in ~50% of AML cases (**Supplementary Table 9**). 27 of the 30 mutated genes had probesets in the mutated exon. 14 of these were expressed in sAML (and in *de novo* AML) and 13 were not (vs. 11 absent in >50% of *de novo* AML cases). Overall, the expression pattern of the 28 genes with tier 1 mutations analyzed by microarray was highly similar in the sAML and *de novo* AML samples.

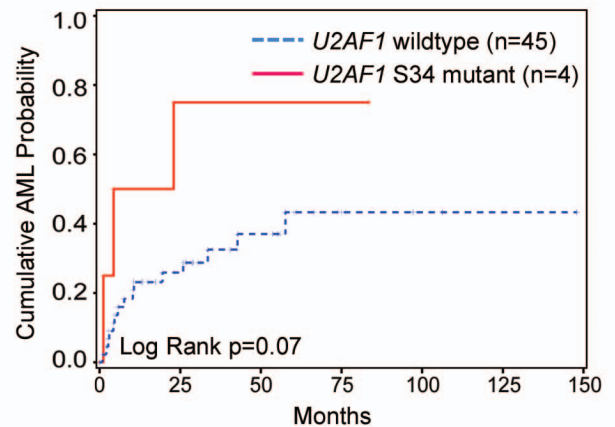
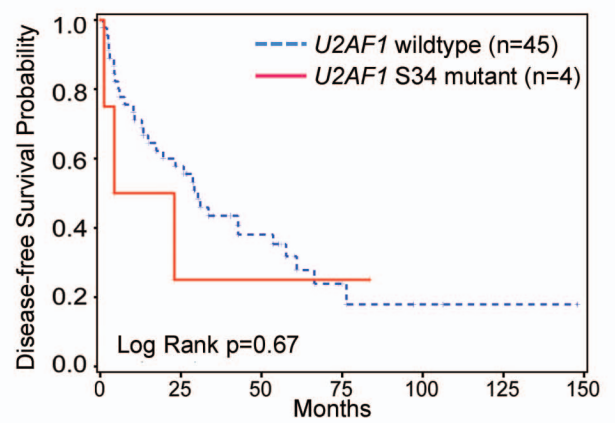
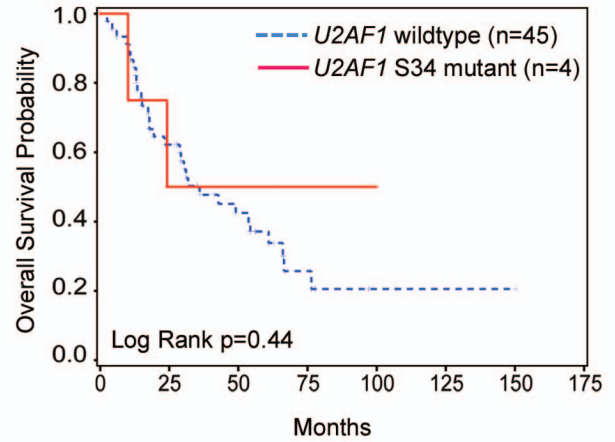
Affymetrix U133plus2 gene expression array data was available for 15 MDS and 4 control samples, as previously described³. RNA was extracted from CD34+ purified bone marrow cells for all samples. Array data was generated from the same batch analysis for the 19 samples and 6 of the 15 MDS samples had *U2AF1* mutations. First, we performed unsupervised hierarchical clustering using 19 arrays (4 control, 6 *U2AF1* mutant MDS samples, and 9 *U2AF1* wild-type MDS samples) and observed that clustering of samples was independent of disease or mutation status. When we restricted our clustering analysis to the 4 control and 6 *U2AF1* mutant samples, using 37,552 probesets that remained after removing probesets absent in all samples and using Ward's, UPGMA, or WPGMA clustering algorithms with Euclidean distance, we observed segregation of the mutant samples (**Supplementary Fig. 4b, left panel**). To identify the most significantly differentially expressed genes between control and *U2AF1* mutant MDS samples from the list of 37,552 probesets, we used the Significance Analysis of Microarray (SAM)⁴ and Gene Set Enrichment Analysis (GSEA)⁵ supervised algorithms. While we found no significant gene sets using GSEA, we did identify 50 probesets that were up-regulated and 351 probesets that were down-regulated in *U2AF1* mutant samples compared to controls with a false discovery rate (FDR) <0.005 (**Supplementary Table 6**). *U2AF1* mutant and control samples segregated by hierarchical clustering when we used 401 SAM defined probesets, as expected (**Supplementary Fig. 4b, right panel**). We analyzed up- and down-regulated genes to identify common pathways based on functional annotation categories (including gene ontology) using DAVID⁶. Nine annotation clusters were identified by DAVID using the 351 down-regulated probesets (300 with unique gene symbols) in *U2AF1* mutant samples that had an enrichment score >2. Three of the 9 clusters contained “mRNA splicing/processing” genes (enrichment score = 2.46) and two clusters with “RRM”(RNA recognition motif) genes (enrichment scores = 3-4.3) involved in splicing (**Supplementary Fig. 4c, Supplementary Table 6**). When genes that segregated control samples from *U2AF1* wild-type MDS samples were analyzed in DAVID, “splicing” and “RRM” categories were not

enriched, suggesting that down-regulation of splicing and RRM genes is not common to all MDS samples, but instead is associated with *U2AF1* mutations. Collectively, these results suggest that a regulatory feedback loop may exist for mRNA splicing genes when *U2AF1* is mutated.

No Transplant



Transplant

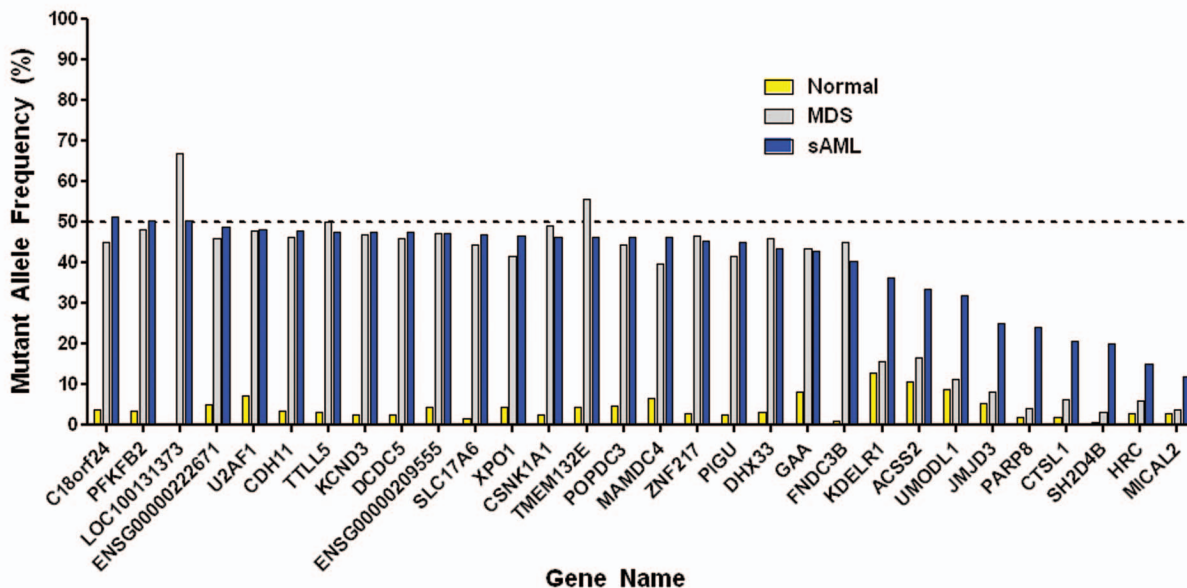


Supplementary Figure 1. Survival analysis in transplanted vs. non-transplanted patients.

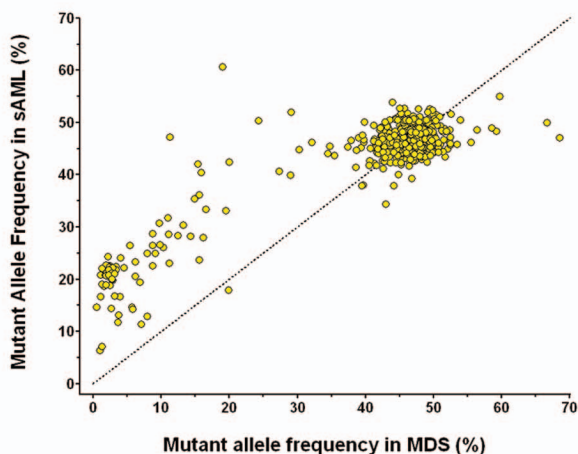
(left column) Overall, event-free survival, and probability of progression to sAML was not significantly different in U2AF1 wildtype vs. mutant MDS patients who did not undergo stem cell transplantation.

(right) Overall, event-free survival, and probability of progression to sAML was not significantly different in U2AF1 wildtype vs. mutant MDS patients who did undergo stem cell transplantation.

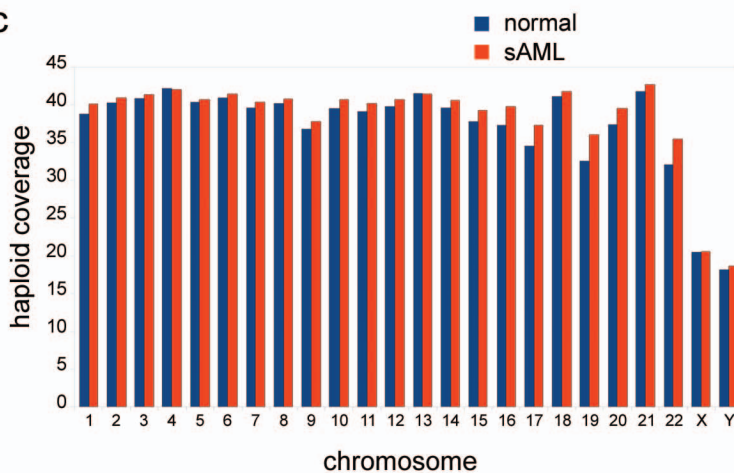
a



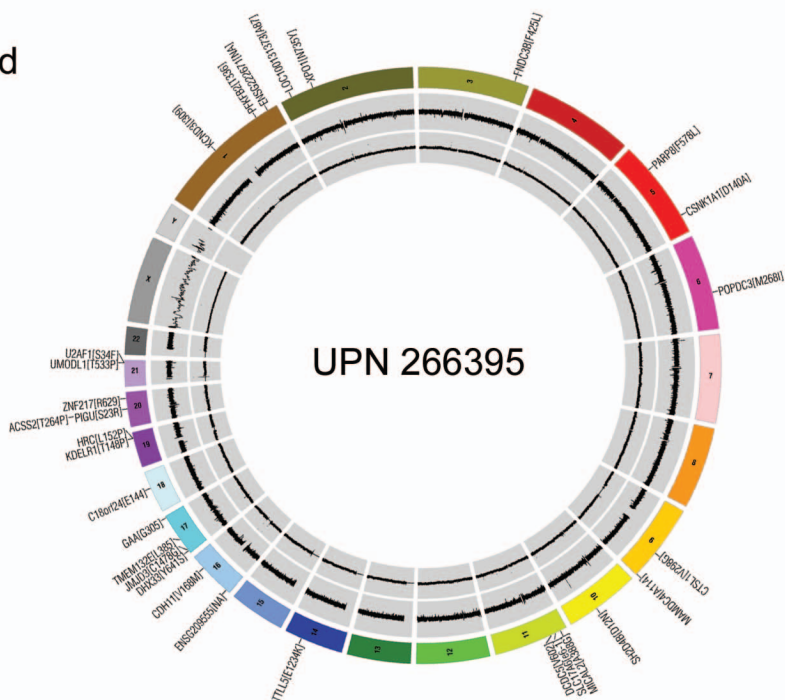
b



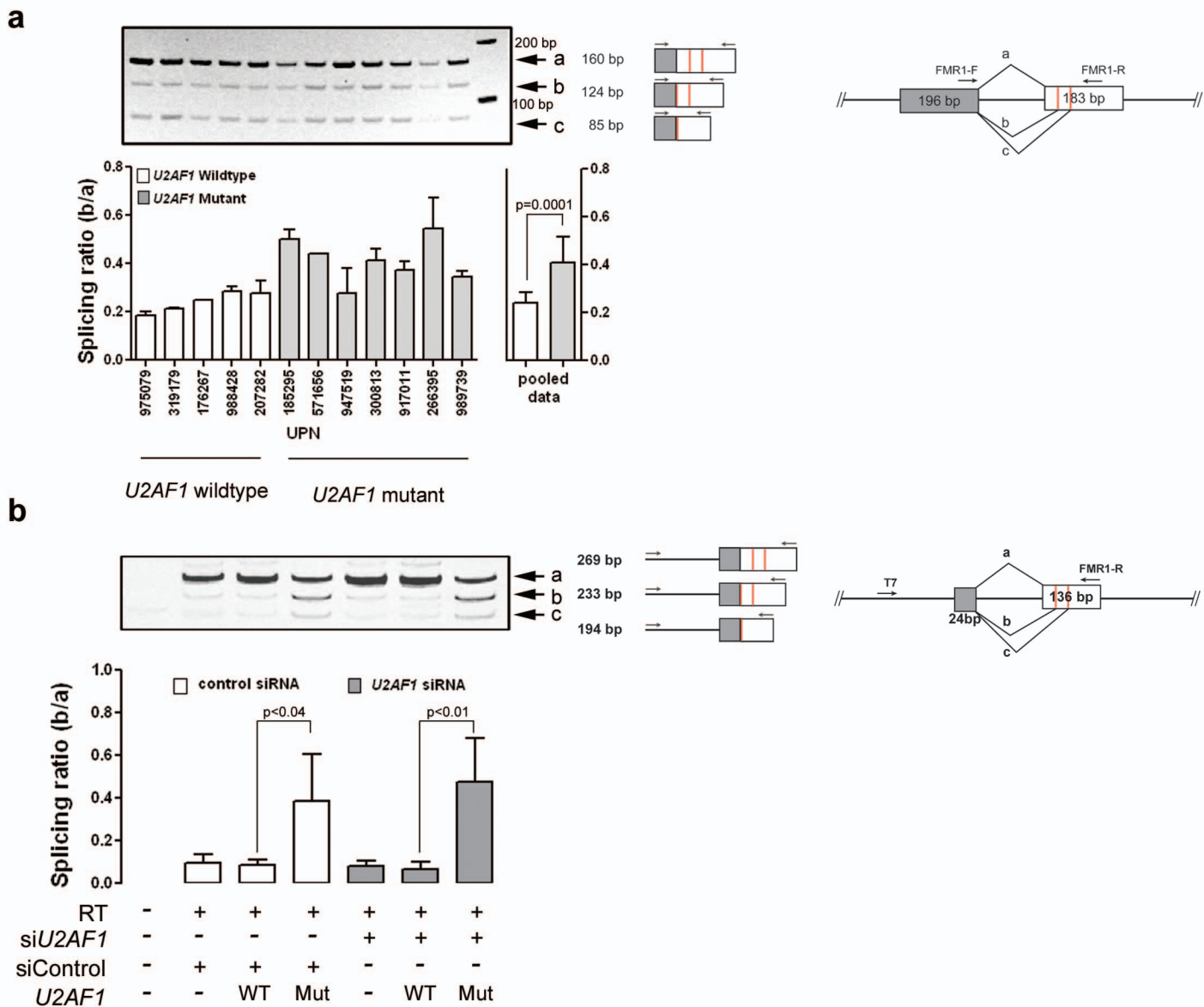
c



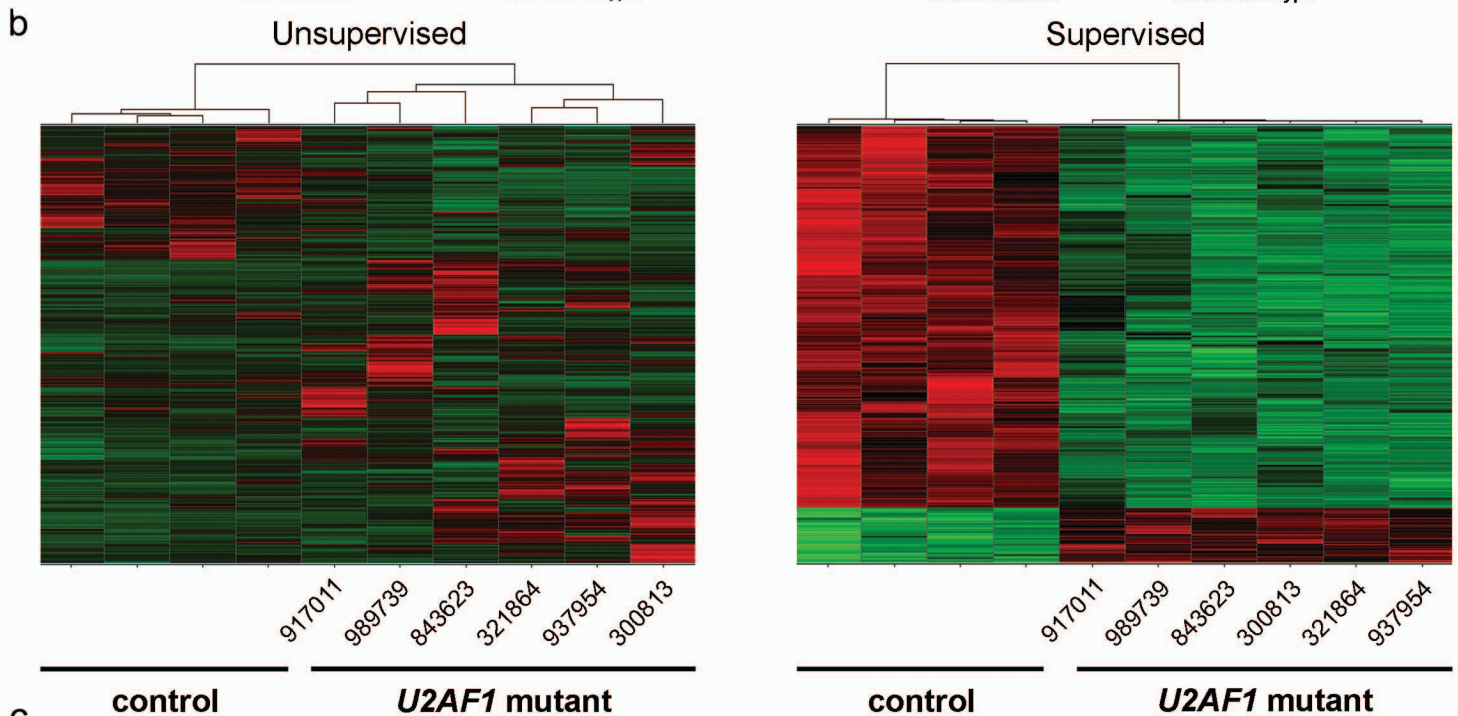
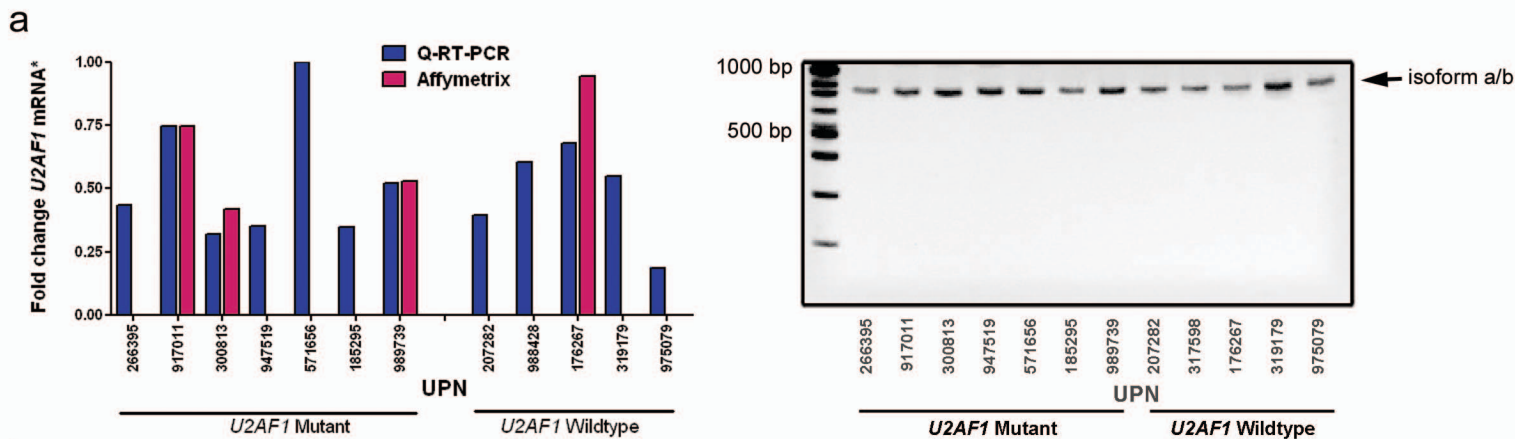
d



Supplementary Figure 2. Genomic architecture of UPN266395 index MDS patient. (a) Mutant allele frequencies of tier 1 SNVs from capture/deep sequencing of normal (skin), MDS, and sAML samples are shown. (b) Mutant allele frequencies of 507 validated somatic tier 1-3 SNVs from capture/deep sequencing are shown for the MDS vs. sAML samples. A single dominant clone was present in MDS and persisted at the sAML stage. Only 2 SNVs were acquired or enriched between MDS and sAML. (c) Sequence coverage from WGS was uniform across chromosomes and between normal (skin) and sAML samples. (d) Circos plot summarizing somatic changes identified by WGS in index case. Outer ring: validated tier 1 SNVs; middle ring: heterozygosity of sAML sample (no loss of heterozygosity detected); inner ring: copy number of sAML sample (no copy number alterations detected). SNV, single nucleotide variant. UPN, unique patient number.



Supplementary Figure 3. U2AF1 S34F mutation induces splicing alterations in FMR1. (a) The relative ratios of alternatively spliced products was measured following RT-PCR of RNA extracted from unpurified bone marrow cells from patients with (n=7) and without (n=5) U2AF1 mutations. A cartoon of FMR1 exons 14 and 15 that are interrogated by the assay and the possible splice products are indicated (labeled a, b, or c). A representative PCR gel image is shown and the ratio of the middle band (amplicon b, utilizing the 3' cryptic splice acceptor site) relative to the fully spliced upper band (amplicon a) is shown for each condition. The ratio of splice products was measured by densitometry and plotted in the bar graph with the p-value listed. Patient samples with U2AF1 mutations had increased usage of a cryptic 3' splice acceptor site in exon 15 of FMR1 relative to nonmutant patient samples (left panel, error bars represent two independent experiments) (wildtype vs. mutant splicing ratio, $p < 0.03$ for both independent experiments in the left panel, $p = 0.0001$ for pooled data in the right panel). (b) The FMR1 minigene (generated using primers shown in panel a) was transiently transfected into 293T cells with a control plasmid (pcDNA3.1-YFP), wild-type U2AF1 cDNA, or mutant (S34F) U2AF1 cDNA in the presence of a control siRNA or siRNA targeting endogenous U2AF1. RNA was harvested 48 hours later and a reverse transcriptase reaction was performed to create cDNA. PCR using the indicated primers resulted in a fully spliced 269 base pair amplicon or smaller amplicons (233 or 194 basepairs) that skip the first splice acceptor site. A representative PCR gel image is shown and the ratio of the middle band (cryptic 3' splice acceptor site = amplicon b) relative to the fully spliced upper band (amplicon a) is shown for each condition. Expression of S34F mutant U2AF1 resulted in an increase in cryptic 3' splice acceptor usage compared to control or wild-type U2AF1, independent of endogenous U2AF1 levels (n=4, $P < 0.04$).



c

Annotation Cluster 1		Enrichment Score: 7.715			
Category	Term	# genes	% total genes	Fold Enrichment	Benjamini p-value
GOTERM_CC_FAT	GO:0031974:membrane-enclosed lumen	67	22.33	2.09	8.89E-07
GOTERM_CC_FAT	GO:0070013:intracellular organelle lumen	64	21.33	2.08	1.41E-06
GOTERM_CC_FAT	GO:0043233:organelle lumen	64	21.33	2.03	2.29E-06
GOTERM_CC_FAT	GO:0031981:nuclear lumen	52	17.33	2.07	3.73E-05
Annotation Cluster 2		Enrichment Score: 4.311			
Category	Term	# genes	% total genes	Fold Enrichment	Benjamini p-value
SMART	SM00360:RRM	14	4.67	4.18	3.33E-03
INTERPRO	IPR000504:RNA recognition motif, RNP-1	14	4.67	3.92	3.70E-02
INTERPRO	IPR012677:Nucleotide-binding, alpha-beta plait	14	4.67	3.88	2.05E-02

Supplementary Figure 4. U2AF1 quantitative RT-PCR and global expression profiling. (a, left panel) qRT-PCR of *U2AF1* (exons 2-3 of isoform A) mRNA was performed using RNA extracted from unpurified bone marrow cells from patients with (n=7) and without (n=5) *U2AF1* mutations. Affymetrix U133plus2 array data were available for 4 of the patients and are plotted for comparison. There was no significant difference in the expression of *U2AF1* in mutant vs. wild-type MDS patient samples (p=0.73). (a, right panel) RT-PCR was performed using RNA extracted from unpurified bone marrow cells from patients with (n=7) and without (n=5) *U2AF1* mutations. Primers were designed to detect all 3 isoforms of *U2AF1*. Full-length isoform a was detected in all samples, independent of mutation status. (b, left panel) Unsupervised hierarchical clustering results using Ward's method and 37,552 probesets from Affymetrix U133plus2 arrays (probesets were excluded only if absent in all samples being compared). RNA was obtained from CD34+ purified bone marrow cells from 4 normal healthy controls and 6 MDS samples with a *U2AF1* mutation. (b, right panel) Supervised hierarchical clustering using Ward's method and 401 probesets identified by Significance Analysis of Microarrays (SAM), comparing control and *U2AF1* mutant MDS samples with a false discovery rate (FDR) < 0.005. (c) Nine gene ontology annotation clusters were identified by DAVID using 351 probesets (300 gene symbols) that were down-regulated in *U2AF1* mutant compared to control samples (identified by SAM) and had an enrichment score > 2 in DAVID. The top 2 clusters are shown. RNA recognition motif (RRM) gene categories were down-regulated in the *U2AF1* samples compared to control samples. UPN, unique patient number. * relative to GAPDH mRNA expression.

Supplementary Table 1. Sequence metrics.

Parameter	Normal Genome	MDS Tumor Genome	sAML Tumor Genome
Whole Genome Sequence			
lanes	8	ND	12
sequence production (Gbp)	134.91	ND	132.24
aligned (Gbp)	122.80	ND	124.37
deduplicated (Gbp)	114.50	ND	116.32
haploid coverage	38.23	ND	39.12
bases with $\geq 1x$ coverage (%)	92.97	ND	93.00
bases with $>10x$ coverage (%)	91.87	ND	92.49
bases with $>20x$ coverage (%)	86.23	ND	89.12
CCDS bases with $\geq 1x$ coverage (%)	99.42	ND	99.67
CCDS bases with $\geq 10x$ coverage (%)	92.24	ND	97.21
CCDS bases with $\geq 20x$ coverage (%)	73.61	ND	88.52
Het SNPs (Affy)	226,603	ND	224,808
Hom SNPs (Affy)	166,483	ND	167,609
Diploid coverage Het SNPs (%)	98.91	ND	99.31
Diploid coverage Hom SNPs (%)	99.55	ND	99.17
Capture Validation			
target bases with $\geq 1x$ coverage (%)	93.06	93.78	94.86
target bases with $>10x$ coverage (%)	87.16	88.53	91.34
target bases with $>20x$ coverage (%)	84.22	86.15	89.84

CCDS, consensus coding DNA sequence; ND, not done.

Het SNPs, single nucleotide polymorphic sites genotyped as heterozygous using Affymetric 6.0 SNP array and both wildtype and variant alleles detected in whole genome sequence data

Hom SNPs, single nucleotide polymorphic sites genotyped as homozygous using Affymetric 6.0 SNP array and only variant allele detected in whole genome sequence data

21	26,646,726	G	A		0	3		0	215	0	0.00	385	9	2.28	2.28	Somatic	0.018	522	137	20.79	20.79	Somatic	0.000	2
21	30,459,290	C	T	CLDN8	3	3_prime_flanking_region		370	16	4.15	921	766	45.41	45.41	Somatic	0.000	354	330	48.25	48.25	Somatic	0.000	1	
21	41,416,902	G	A	BACE2	3	5_prime_flanking_region		1038	32	2.99	664	560	45.75	45.75	Somatic	0.000	399	339	45.93	45.93	Somatic	0.000	1	
21	42,402,818	A	C	UMODL1	1	missense		84	8	8.70	65	8	10.96	10.96	Somatic	0.577	161	75	31.78	31.78	Somatic	0.000	0	
21	43,397,525	G	A	U2AF1	1	missense		404	30	6.91	91	83	47.70	47.70	Somatic	0.000	343	317	48.03	48.03	Somatic	0.000	1	
21	46,252,211	C	T	COL6A1	2	3_prime_flanking_region		92	1	1.08	89	67	42.95	42.95	Somatic	0.000	213	170	44.39	44.39	Somatic	0.000	1	
21	46,301,845	G	T	COL6A2	2	5_prime_flanking_region		46	0	0.00	11	2	15.38	15.38	Reference	0.041	251	182	42.03	42.03	Somatic	0.000	0	
22	18,187,764	C	T	GNB1L	3	intronic		387	7	1.78	104	92	46.94	46.94	Somatic	0.000	223	211	48.62	48.62	Somatic	0.000	1	
22	19,294,334	C	G	ENSG00000220150	3	5_prime_flanking_region		265	25	8.62	128	119	48.18	48.18	Somatic	0.000	360	298	45.29	45.29	Somatic	0.000	1	
22	20,866,781	G	A	IGLV11-55	3	5_prime_flanking_region		328	21	6.02	185	239	56.37	56.37	Somatic	0.000	165	156	48.60	48.60	Somatic	0.000	0	
22	36,194,879	A	C	ELFN2	3	5_prime_flanking_region		735	74	9.15	291	56	16.14	16.14	Somatic	0.000	502	195	27.98	27.98	Somatic	0.000	0	
22	38,089,713	C	T	SYNGR1	3	intronic		954	46	4.60	97	89	47.85	47.85	Somatic	0.000	145	139	48.94	48.94	Somatic	0.000	1	
22	46,395,195	C	T	FLJ46257	3	3_prime_flanking_region		378	10	2.58	137	111	44.76	44.76	Somatic	0.000	302	308	50.49	50.49	Somatic	0.000	1	
22	46,512,620	A	G		0	3		471	11	2.28	526	17	3.13	3.13	Somatic	0.421	431	115	21.06	21.06	Somatic	0.000	2	
X	3,387,861	G	A		0	3		375	0	0.00	362	21	5.48	2.74	Somatic	0.000	167	112	40.14	20.07	Somatic	0.000	2	
X	29,779,410	G	A	IL1RAPL1	2	intronic		121	6	4.72	21	384	94.81	47.41	Somatic	0.000	19	292	93.89	46.95	Somatic	0.000	1	
X	32,777,396	C	T	DMD	3	intronic		414	38	8.41	31	496	94.12	47.06	Somatic	0.000	21	292	93.29	46.65	Somatic	0.000	1	
X	34,705,765	A	G		0	3		217	29	11.79	7	287	97.62	48.81	Somatic	0.000	12	311	96.28	48.14	Somatic	0.000	1	
X	39,157,340	A	G		0	3		283	22	7.21	22	391	94.67	47.34	Somatic	0.000	22	413	94.94	47.47	Somatic	0.000	1	
X	52,897,186	C	T	FAM156B	3	5_prime_flanking_region		226	11	4.64	18	368	95.34	47.67	Somatic	0.000	14	282	95.27	47.64	Somatic	0.000	1	
X	104,353,541	G	C	IL1RAPL2	3	intronic		141	21	12.96	24	327	93.16	46.58	Somatic	0.000	19	424	95.71	47.86	Somatic	0.000	1	
X	111,088,560	G	A	TRPC5	3	intronic		253	11	4.17	39	584	93.74	46.87	Somatic	0.000	15	321	95.54	47.77	Somatic	0.000	1	
X	116,699,743	C	A		0	3		219	25	10.25	34	461	93.13	46.57	Somatic	0.000	11	396	97.30	48.65	Somatic	0.000	1	
X	118,033,017	G	A	LONRF3	3	intronic		473	37	7.25	18	437	96.04	48.02	Somatic	0.000	16	351	95.64	47.82	Somatic	0.000	1	
X	132,324,635	C	T	GPC4	3	intronic		642	64	9.07	61	841	93.24	46.62	Somatic	0.000	21	412	95.15	47.58	Somatic	0.000	1	
X	142,016,065	C	T	ENSG00000202473	3	3_prime_flanking_region		59	6	9.23	10	139	93.29	46.64	Somatic	0.000	8	164	95.35	47.67	Somatic	0.000	1	
X	151,265,229	C	T	GABRA3	3	intronic		621	28	4.31	50	711	93.43	46.71	Somatic	0.000	14	376	96.41	48.21	Somatic	0.000	1	
Y	9,231,939	G	A	TTTY20	2	intronic		115	8	6.50	6	96	94.12	47.06	Somatic	0.000	17	226	93.00	46.50	Somatic	0.000	1	
Y	15,681,536	T	C		0	3		359	9	2.45	512	41	7.41	3.71	Somatic	0.015	308	110	26.32	13.16	Somatic	0.000	0	
Y	22,378,497	C	T	ENSG00000217635	3	5_prime_flanking_region		217	24	9.96	9	209	95.87	47.94	Somatic	0.000	17	196	92.02	46.01	Somatic	0.000	1	
Y	22,830,969	C	T	RBMV1J	3	5_prime_flanking_region		158	11	6.51	12	215	94.71	47.36	Somatic	0.000	13	321	96.11	48.05	Somatic	0.000	1	

Supplementary Table 3. Annotated Tier 1 Mutations in the MDS and sAML genomes.

Chromosome	Position*	Reference allele	Variation allele	Gene symbol	Transcript ID	Strand	Mutation type	Mutated in MDS	Mutated in sAML	Zygoty	Coding position	Amino acid change
1	112,325,945	G	T	KCND3	NM_004980	-1	synonymous	Y	Y	Het	c.927	I309
1	205,309,412	C	G	PFKFB2	NM_006212	1	synonymous	Y	Y	Het	c.1008	T336
1	229,096,537	T	C	ENSG00000222671	ENST00000410739	1	non-coding RNA	Y	Y	Het	NA	NA
2	20,289,990	C	T	LOC100131373	XM_001720891	1	synonymous	Y	Y	Het	c.261	A87
2	61,569,230	T	A	XPO1	NM_003400	-1	missense	Y	Y	Het	c.2203	N735Y
3	173,529,456	C	A	FNDC3B	NM_022763	1	missense	Y	Y	Het	c.1275	F425L
5	50,152,869	C	G	PARP8	NM_024615	1	missense	Y	Y	Het	c.1734	F578L
5	148,880,083	T	G	CSNK1A1	NM_001025105	-1	missense	Y	Y	Het	c.419	D140A
6	105,713,110	C	A	POPDC3	NM_022361	-1	missense	Y	Y	Het	c.804	M268I
9	89,535,224	T	G	CTSL1	NM_001912	1	missense	Y	Y	Het	c.893	V298G
9	138,867,831	C	T	MAMDC4	NM_206920	1	synonymous	Y	Y	Het	c.342	A114
10	82,288,101	G	A	SH2D4B	NM_207372	1	missense	Y	Y	Het	c.34	D12N
11	12,198,538	C	G	MICAL2	NM_014632	1	missense	Y	Y	Het	c.1163	A388G
11	22,340,860	G	T	SLC17A6	NM_020346	1	splice_site	Y	Y	Het	c.662-1	e6-1
11	30,881,639	A	G	DCDC5	ENST00000406071	-1	synonymous	Y	Y	Het	c.1806	V602
14	75,418,958	G	A	TLL5	NM_015072	1	missense	Y	Y	Het	c.3700	E1234K
16	7,716,860	C	T	ENSG00000209555	ENST00000386820	-1	non-coding RNA	Y	Y	Het	NA	NA
16	63,589,987	C	T	CDH11	NM_001797	-1	missense	Y	Y	Het	c.502	V168M
17	5,288,451	T	G	DHX33	NM_020162	-1	missense	Y	Y	Het	c.1922	Y641S
17	7,696,343	T	G	JMJD3	NM_001080424	1	missense	Y	Y	Het	c.4432	C1478G
17	29,981,226	C	T	TMEM132E	NM_207313	1	synonymous	Y	Y	Het	c.1155	L385
17	75,696,250	G	A	GAA	NM_000152	1	synonymous	Y	Y	Het	c.915	G305
18	46,165,704	G	A	C18orf24	NM_001039535	1	synonymous	Y	Y	Het	c.432	E144
19	53,579,461	T	G	KDELR1	NM_006801	-1	missense	Y	Y	Het	c.442	T148P
19	54,349,852	A	G	HRC	NM_002152	-1	missense	Y	Y	Het	c.455	L152P
20	32,728,482	A	C	PIGU	NM_080476	-1	missense	Y	Y	Het	c.69	S23R
20	32,965,857	A	C	ACSS2	NM_018677	1	missense	Y	Y	Het	c.790	T264P
20	51,626,823	T	C	ZNF217	NM_006526	-1	synonymous	Y	Y	Het	c.1887	R629
21	42,402,818	A	C	UMODL1	NM_173568	1	missense	Y	Y	Het	c.1597	T533P
21	43,397,525	G	A	U2AF1	NM_001025203	-1	missense	Y	Y	Het	c.101	S34F

*build 36

Supplementary Table 4. U2AF1 Somatic Mutations in MDS.

UPN	gene name	chromosome	start	stop	reference allele	variant allele	transcript	strand	mutation type	coding position	amino acid change	SIFT	SIFT score	PolyPhen2	PolyPhen2 score
02-86	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
185295	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
266395	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
300813	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
321864	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
463763	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
571656	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
843623	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
917011	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
947519*	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
989739	U2AF1	21	43397525	43397525	G	A	NM_001025203	-1	missense	c.101	p.S34F	damaging	0	probably damaging	1
137250	U2AF1	21	43397525	43397525	G	T	NM_001025203	-1	missense	c.101	p.S34Y	damaging	0	probably damaging	1
937954	U2AF1	21	43397525	43397525	G	T	NM_001025203	-1	missense	c.101	p.S34Y	damaging	0	probably damaging	1
947519*	U2AF1	21	43387846	43387846	T	C	NM_001025203	-1	missense	c.470	p.Q157R	damaging	0	probably damaging	0.999

UPN, unique patient number

* same UPN

Supplementary Table 5. Characteristics of MDS patients with *U2AF1* Mutations

UPN	Gender	Age at diagnosis	FAB	Karyotype	%BM Blast	IPSS	Therapy after MDS banking (months)	AML Progression?	OS (months)	EFS (months)
266395 ^a	M	65	RAEB	46,XY[20]	7	1	azacytidine (6)	yes	11	2
137250	F	54	RAEB	46,XX[20]	19	2	standard induction, allogeneic SCT	yes	100	1
185295	M	50	RA	46,XY,del(20)(q11.2q13.3)[7]/46,XY,[13]	0	0.5	PTK/ZK (<1), azacytidine (2), allogeneic SCT	no	97	83
300813	F	42	RA	46,XX[20]	2	0	none	no	71	71
321864	F	63	RAEB	46,XX+8[20]	12	2.5	standard induction	no	3	3
571656	M	71	RAEB	46,XY,del(20)(q11.2q13.3)[12]/46XY[8]	4	0.5	standard induction	yes	10	8
843623	M	60	RAEB	46,XY,del(20)(q11.3)[2]/46,XY[18]	6	1	none	yes	5	2
917011	M	79	RA	46,XY,del(20)(q11.2)[12]/ 46,XY[8]	0	0.5	lenalidomide (3)	no	27	25
937954	M	51	RAEB	47,XY,+8[1]/47, idem, del(16)(q12.1)[13]/48, idem,+Y, del(16)(q11.1)[5]/46,XY[1]	12	2.5	azacytidine (3)	no	24	24
947519	M	59	RAEB	46,XY,1-5dms[10]/46,XY[12]	7	1.5	decitabine (3), allogeneic SCT, standard induction, salvage chemotherapy	yes	24	23
02-86	M	53	RAEB-T	92<4n>,XXYY,del(4)(q33)x2,del(5)(q13),add(11) (q23)x2[cp8]/46,XY[12]	21	3	decitabine (2)	yes	10	4
463763	M	54	RA	47,XY,+Y[6]/46,XY[14]	0	0.5	none	no	46	44
989739	M	71	RAEB	46,XY,del(7)(q21)[11]/47,XY,+8[5]/46,XY[4]	18	3	decitabine (8)	yes	15	4

UPN, unique patient number; OS, overall survival; EFS, event-free survival; PTK/ZK (Vatalanib), VEGF inhibitor.

^aindex case

Supplementary Table 7. Clinical Characteristics of MDS replication cohort.

UPN	Gender	Age @ Diagnosis	%BM Blast	FAB	Cytogenetics	[-5, del(5q)] (0=no/1=yes)	[-20, del(20q)] (0=no/1=yes)	[-7, del(7q)] (0=no/1=yes)	[+8] (0=no/1=yes)	IPSS	Disease Status (0=alive/1=dead)	Days from Dx to Death or Last Follow-up	Evolution to AML (0=no/1=yes)	Days from Dx to AML or last Follow-up	DNMT3A status (0=wt/1=mut)	U2AF1 status (0=wt/1=mut)
100901	M	70	4	RA	46,XY,del(3)(q21)[16]/46,XY[4]	0	0	0	0	0.5	1	1857	0	1654	0	0
105327	F	54	1	RA	46,XX[20]	0	0	0	0	0.0	1	1647	0	1646	0	0
112848	M	62	5	RAEB	46,XY,del(5)(q13q33)[18]/46,XY[1]/45,XY,-5[1]	1	0	0	0	0.5	1	2116	0	2033	0	0
116960	F	55	2	RA	46,XX,t(11;13)(q23;q14)[14]/47,idem,+2[2]/46,XX[3]/46,XX,inv(14)(q11,q32)[1]	0	0	0	0	1.0	0	2362	0	2246	0	0
123363	M	64	0	RA	46,XY[19]/46,XY,del(20)(q11,q13.1)[1]	0	0	0	0	0.0	0	2421	0	2402	0	0
124124	F	66	18	RAEB	46,XX[20]	0	0	0	0	2.0	1	998	1	891	0	0
127804	M	63	2	RA	46,XY[20]	0	0	0	0	0.5	1	2375	0	2368	0	0
131153	M	78	8	RAEB	44,XY,dup(3)(p21p25),-5,-6,del(12)(q15q22),psu dic(15;14)(p13;q32),add(17)(p11.2),add(19)(q13.3),7i(22;22)(q12.13),+mar[20]	1	0	0	0	2.0	1	621	0	619	0	0
137250	F	54	19	RAEB	46,XX[20]	0	0	0	0	2.0	0	3037	1	38	0	1
137313	F	77	0	RA	47,XX,+8[6]/46,XX[14]	0	0	0	0	1.0	1	320	0	144	0	0
137404	F	70	0	RAR	46,XX,add(1)(p11.3),del(1)(q23),add(3)(p13),del(5)(q13q33),add(8)(p11.2),add(9)(q22),add(10)(q22),add(11)(p14),-12,del(17)(q21),+mar[17]/46,XX[3]	1	0	0	0	1.0	1	674	0	663	0	0
173005	M	77	10	RAEB	45,XY,der(3)(5)(p10;p10),-18,-20,+2mar[12]/46,idem,+21[5]/46,XY[3]	0	0	0	0	2.0	1	800	0	71	0	0
176267	M	56	7	RAEB	43-47,XY,del(5)(q13q33),der(7)(7;11)(q11.2;q21),-9,-12,der(16)(12;16)(q10;q22),add(16)(q23),+mar1,+mar2[cp12]/44-46,XY,del(5)(q13q33),der(7)(7;11)(q11.2;q21),t(12;16)(q10;q22)[cp8]	1	0	0	0	2.0	1	137	1	93	1	0
177356	F	76	2	RA	46,XX[20]	0	0	0	0	0.0	0	1304	0	747	0	0
182896	M	75	0	RA	46,XY,add(4)(p16),t(4;20)(q25;p13),del(5)(q13q33)[13]/46,idem,del(6)(q25)[1]/47,XY,+8[6]	0	0	0	1	1.0	1	1205	1	1047	0	0
184368	F	49	0	RA	46,XX[20]	0	0	0	0	0.0	0	2013	0	81	0	0
184874	F	56	2	RAR	47,XX,+8[10]/46,XX[10]	0	0	0	1	0.5	1	2020	0	2020	0	0
185295	M	50	0	RA	46,XY,del(20)(q11,q13.3)[7]/46,XY[13]	0	1	0	0	0.5	0	2539	0	2538	0	1
189012	M	64	3	RA	46,XY,del(5)(q15q33)[16]/46,XY,del(5)(q12q33),del(7)(q21q32)[1]/46,XY[3]	1	0	0	0	0.5	0	1250	0	1230	0	0
189474	M	61	4	RAEB	46,XY[3]	0	0	0	0	0.5	0	1336	0	907	0	0
194328	M	52	1	RA	46,XY[20]	0	0	0	0	0.0	0	1702	0	809	0	0
197795	M	50	17	RAEB	46,XY[20]	0	0	0	0	1.5	1	1064	0	457	0	0
199019	F	60	0	RA	45-49,XX,+1,add(3)(p11),del(5)(q13q33),del(10)(q24q26),add(18)(q21),+19,-20,+21,+1-2mar[cp20]	1	1	0	0	1.5	1	1491	1	1020	0	0
201641	M	62	14	RAEB	44-45,XY,del(3)(q12),add(5)(q35),del(5)(q12q33),del(7)(q22),-12,der(17)(7;17)(q22p13)[15]/46,XY[5]	1	0	0	0	3.0	1	220	0	199	0	0
207282	M	76	11	RAEB	46,XY[20]	0	0	0	0	2.0	1	251	1	122	1	0
222841	F	46	17	RAEB	60<3ns>,XX,-X,add(1)(q44),ins(1)(p36)add(1)(q12),-5,del(5)(q11,q23q3),+6,-7,add(7)(p22),-10,-11,del(12)(q13),add(12)(p13),-13,-14,hsr(14)(p11.1),-16,-16,-18,der(19;21)(q10;q10),+21,-22,+mar5/46,XX[15]	1	0	1	0	2.5	1	528	0	528	0	0
266395	M	65	7	RAEB	46,XY[20]	0	0	0	0	1.0	1	321	1	75	0	1
289033	F	31	13	RAEB	46,XX[20]	0	0	0	0	2.0	1	367	1	28	0	0
298273	M	27	18	RAEB-T	46,XY[20]	0	0	0	0	2.0	1	540	1	131	0	0
300813	F	42	2	RA	46,XX[20]	0	0	0	0	0.0	0	2172	0	2172	0	1
312519	F	71	0	RA	46,XX[20]	0	0	0	0	0.5	1	1026	0	436	0	0
313175	M	74	5	RAEB	46,XY[20]	0	0	0	0	0.0	0	961	0	702	0	0
315529	M	56	12	RAEB	46,XY,add(7)(q32),r(8)(p21q24.3),add(11)(p15),t(12;13)(q24.2;q13),-17,r(19)(p13.3q13.4),-20,add(20)(q13.1),-21,+3mar[18]/46,idem,del(5)(q22q34)[1]/46,XY[1]	0	1	0	0	3.0	1	147	0	147	0	0
317598	M	86	11	RAEB	46,XY[20]	0	0	0	0	2.0	1	30	0	30	1	0
319179	F	74	6	RAEB	46,XX[20]	0	0	0	0	0.5	1	483	1	482	1	0
321864	F	63	12	RAEB	47,XX,+8[20]	0	0	0	0	2.5	1	80	0	80	0	1
329877	M	65	0	RA	46,XY[20]	0	0	0	0	1.5	1	249	0	186	0	0
335224	M	65	3	RA	45,XY,del(5)(q22q33),hsr(11)(p15),der(17;22)(q10;q10)[17]/45,XY,idem,hsr(11)(p15)(q23)[3]	1	0	0	0	1.5	1	202	1	104	0	0
335875	F	46	2	RA	46,XX[20]	0	0	0	0	0.0	0	1387	0	1229	0	0
337081	M	67	10	RAEB	46,XY[20]	0	0	0	0	0.5	1	1108	0	31	0	0
342760	F	77	12	RAEB	46,XX[20]	0	0	0	0	1.5	1	1447	0	1447	0	0
343306	F	62	15	RAEB	40-45,XX,-4,-5,add(10)(p15),add(15)(p11.2),-17,+add(19)(p13.3),add(21)(p11.2),-22,+mar[cp3]/46,XX[10]-only 13 analyzed	1	0	0	0	3.0	1	686	0	683	0	0
346313	M	62	6	RAEB	46,XY[20]	0	0	0	0	1.0	1	965	1	787	0	0
350460	M	30	4	RA	46,XY[17]	0	0	0	0	0.5	1	905	0	904	0	0
362699	F	73	2	RA	46,XX[18]/47,XX,+4[1]/47,XX,+9[1]	0	0	0	0	0.5	1	307	0	127	0	0
369682	M	59	ND	RA	45,XY,der(5)t(3;5)(q21;q13)[9]/46,idem,+8[6]/46,idem,+mar[1]/46,XY[4]	0	0	0	1	UK	1	34	1	91	0	0
372440	M	59	2	RA	46,XY[20]	0	0	0	0	0.0	0	1891	0	1894	0	0
379929	F	39	3	RA	46,XX[20]	0	0	0	0	0.5	1	290	1	224	1	0
391828	M	50	3	RA	46,XY,inv(3)(q21q26.2)[6]/46,XY[14]	0	0	0	0	1.0	1	1852	0	1852	0	0
392543	M	74	13	RAEB	55-57,XY,+1,+2,+6,der(7)(7;11)(p22;q13),+9,-10,+11,+13,+14,add(14)(p11.2),add(17)(p11.2),-18,+21,add(21)(p11.2),+22,+1-2r,+1-3mar[cp7]/46,XY[13]	0	0	0	0	2.5	1	599	1	589	0	0
395621	F	35	6	RAEB	46,XX,der(15)t(1;15)(q10;q10)[15]/46,XX[5]	0	0	0	0	1.0	1	541	1	140	0	0
411330	M	83	2	RA	45,X,-Y[3]/46,XY[17]	0	0	0	0	0.0	1	18	0	18	0	0
421452	F	33	10	RAEB	46,XX[20]	0	0	0	0	1.0	0	2016	1	90	0	0
429531	F	38	0	RA	46,XX,del(5)(q12q33)[18]/46,XX[2]	1	0	0	0	0.5	0	2951	0	2577	0	0
430998	M	56	0	RA	47,XY,+14[5],46,XY,t(4;6)(q10;5)/46,XY[10]	0	0	0	0	0.5	0	1067	0	706	0	0
431137	M	55	7	RAEB	46,XY,add(2)(q26),del(3)(p13),del(6)(q15q31),-6,+8,add(12)(p11.2),add(15)(p11.1),-18,+22[cp4]/44,sl,-8,-20,-22,+mar[cp5]/44,sl,-del(3),-7,+2mar[cp5]/43,sl,add(X)(q28),+add(12p),+12,-mar[cp6]	1	1	1	1	2.0	1	181	0	181	0	0
431586	M	68	6	RAEB	87,XY,+X,+1,+1,+2,+2,+3,+3,+4,+4,+5,+5,+6,+6,+7,+7,+8,+8,+9,+9,+10,+10,+11,+11,+12,+12,+13,+13,+14,+14,+15,+15,+16,+16,+17,+17,+18,+18,+19,+19,+20,+20,+21,+21,+22,+22[3]/46,XY[22]	0	0	0	1	1.5	1	337	1	275	0	0
447189	M	76	0	RA	46,XY[20]	0	0	0	0	0.5	0	1795	0	1436	0	0
455113	F	68	2	RA	46,XY[20]	0	0	0	0	0.0	0	2244	0	1982	0	0
455194	M	62	2	RA	47,XY,+8[19]/46,XY[1]	0	0	0	1	0.5	0	1529	0	1302	0	0
457721	M	63	22	RAEB-T	46,XY[20]	0	0	0	0	2.5	1	398	1	173	1	0
461282	M	65	6	RAEB	46,XY,-17,del(20)(q11.2),f(5)/46,idem,del(5)(q31q33)[10]/46,idem,del(5)(q13q33)[3]/46,idem,-X,add(X)(q22),der(12)t(12;7)(p11.2;7)del(12)(q22)[2]	1	1	0	0	2.0	1	2009	1	1751	0	0
461321	F	78	6	RAEB	46,XX[18]	0	0	0	0	0.5	0	981	1	178	0	0
462596	M	78	0	RA	46,XY[20]	0	0	0	0	0.5	1	395	0	177	0	0
463763	M	54	0	RA	47,XY,+Y[6]/46,XY[14]	0	0	0	0	0.5	0	1385	0	1332	0	1
485523	M	62	11	RAEB	46,XY[20]	0	0	0	0	2.0	0	814	0	804	0	0
497942	M	38	0	RA	46,XY,der(1;22)(q10;q10)[4]/46,XY[16]	0	0	0	0	1.0	1	4626	0	4553	0	0
511734	M	69	4	RAEB-T	46,XY[20]	0	0	0	0	0.0	1	353	0	353	0	0
518747	F	47	6	RAEB	46,XX[11]	0	0	0	0	1.0	0	844	0	816	0	0
530722	M	69	13	RAEB	46,XY[20]	0	0	0	0	2.0	1	311	1	24	0	0
530724	F	67	11	RAEB	Failed	0	0	0	0	UK	1	32	0	12	0	0
537426	F	57	2	RA	45,XX,-7[20]	0	0	1	0	1.5	0	1908	0	1833	0	0
538267	F	49	12	RAEB	42-88,XX,-X,(3;17)(q21;q21),del(5)(q15q33),-10,t(2;18)(q11.2;q11.2),-12,-13,t(17;20)(q21;11.2)add(17)(p11.2),dic(19;21)(p13q22),+r,+mar[cp18]/46,XX[2]	1	0	0	0	3.0	1	407	0	407	0	0
543465	M	75	11	RAEB	46,XY[20]	0	0	0	0	2.0	0	3080	1	160	0	0
547403	M	61	16	RAEB	46,X,-Y,+14[2]/45,X,-Y[2]/45,XY,-20[1]	0	0	0	0	2.5	1	132	0	132	0	0
545462	M	80	0	RA	46,XY,del(20)(q11.2q13.3)[8]/46,XY,+8[3]/46,XY[9]	0	1	0	0	0.5	1	1256	1	1314	0	0
557468	M	76	18	RAEB	46,XY[20]	0	0	0	0	2.0	1	712	1	217	0	0
558896	M	73	15	RAEB	47-50,XY,t(2;19)(p16;p13),del(5)(q12q33),-11,-18,-20,+r,+2-6mar[4]/41-50,sl,+5,del(5)(q12q33)[8]/47-50,sl,t,+r[8]	1	1	0	0	3.0	1	57	0	57	0	0
571656	M	71	4	RAEB	46,del(20)(q11.2q13.3)[12]/46,XY[8]	0	0	0	0	0.5	1	303	1	231	0	1
574490	M	57	1	RA	46,XY[20]	1	0	0	0	0.0	1	874				

643607	F	79	3	RA	46_XX[20]						0.0	1	897	0	14	0	0
657314	F	28	4	RAEB	45_XX-7[5]				1	0	1.5	0	1704	0	1681	0	0
667720	F	66	10	RAEB	46_XX[20]				0	0	1.0	1	724	1	644	0	0
672525	F	73	5	RAEB	46_XX[20]				0	0	0	1	2395	0	2247	0	0
684211	F	62	1	RAR	46_XX.del(5)(q13q33).del(20)(q11.2)[cp5]				1	1	0	0	0.5	1	325	1	231
690100	M	66	15	RAEB	43-45.XY.del(5)(q22q35).der(16)(16.17)(q22;q21).-17.add(17)(p11.2).-18.dic(18:21)(p11.1;q22).der(22)(1:22)(p13;p11.2).+0-2mar[cp19]47.XY.+12[1]				1	0	0	0	3.0	1	331	1	324
693881	F	78	11	RAEB	44_XX-3.del(5)(q13q33).-7.add(9)(q22).add(13)(p11.2).-18.add(18)(p11.2).-20.+2mar[18]46_XX[2]				1	1	1	0	3.0	1	383	0	369
696026	M	48	0	RA	47_XY.+8[2]47.idem.add(3)(q21)[6]47.add(3)(q21).del(12)(p11.2p13)[12]				0	0	0	1	1.5	0	4566	0	4498
697296	M	71	3	RA	46_XY[20]				0	0	0	0	0.5	1	1385	0	87
698840	F	77	13	RAEB	46_XX.del(5)(q13q23)[1]42-47.si.dic(15:20)(p11.2;q13.3).add(22)(p11.1)[cp3]44-47.si.dic(3)(3;15)(p11.2;q11.1).der(14;22)(q10;q10).+0-5mar[cp6]82-87.si2.-3.del(4)(q727)x2.-7.add(12)(p11.2)x2.der(15:21)(q10;q10).der(15;22)(q10;q10).-17.-17.+19.-21.add(21)(p11.2).-22.-22.+0-5mar[cp9]46_XX[1]				0	0	1	0	3.0	1	253	0	7
703653	F	20	0	RA	46_XY[20]				0	0	0	0	0.5	0	2257	0	2092
705782	M	34	5	RAEB	46_XX.[1;3;5](q21;q31)[13]46.idem.2add(12)(q24.3)[5]46_XY.2add(12)(q24.3)[2]				0	0	0	0	2.0	0	1711	0	1703
717045	F	48	2	RA	46_XX.del(5)(q13q33)[20]				1	0	0	0	0.5	1	393	0	393
732776	F	47	8	RAEB	46_XX[20]				0	0	0	0	1.0	0	2255	0	11
736895	M	54	2	RA	43-49.XY.-4.-5.-7.add(12)(p11.2).-20.-21.+1-6mar[cp7]				1	1	1	0	1.5	1	38	0	26
747967	M	41	2	RA	46_XY.+1.der(1:15)(q10;q10)[8]46_XY.+1.der(1:7)(q10;p10)[4]46_XY[8]				0	0	0	0	1.0	0	1940	0	1899
753222	M	79	0	RA	60-61.XXY.add(3)(p21).del(5)(q13q33).add(6)(q25).-7.-8.-9.add(14)(p11.2).-16.-17.+19.add(19)(q13.4)x2.-22[cp2]46_XY[68]				1	0	1	0	1.5	1	123	0	123
783437	M	63	8	RAEB	46_XY[20]				0	0	0	0	1.0	1	1095	1	594
785104	F	44	2	RA	46_XX[20]				0	0	0	0	0.0	0	1317	0	1172
791334	F	66	3	RA	46_XX[20]				0	0	0	0	0.5	1	923	0	923
804059	M	69	9	RAEB	47_XY.+8[20]				0	0	0	1	2.5	1	878	1	267
809605	M	75	14	RAEB	46_XY[12]				0	0	0	0	2.0	0	849	0	849
817019	F	65	3	RA	46_XX.idel(20)(q11.2)[9]46_XX[11]				0	1	0	0	0.0	0	1172	0	861
827584	M	62	8	RAEB	47_XY.+8[15]46_XY[5]				0	0	1	1.0	0	1710	0	1703	
843623	M	60	6	RAEB	46_XY.del(20)(q11.2q13.3)[2]46_XY[18]				0	1	0	1.0	1	139	1	71	
845446	M	73	13	RAEB	47_XY.+8.del(12)(p11.2p13)[18]47.idem.del(3)(q21)[2]				0	0	0	1	3.0	1	251	0	251
846155	M	64	0	RA	46_XY[20]				0	0	0	0	0.5	0	1023	0	1026
854400	M	63	0	RA	46_XY.yqh+20]				0	0	0	0	0.0	1	945	0	923
855934	M	62	26	RAEB-T	46_XY[20]				0	0	0	0	2.5	0	453	0	453
857645	M	60	ND	RA	46_XY[20]				0	0	0	0	UK	1	860	1	821
859640	F	64	3	RA	46_XX[20]				0	0	0	0	1.5	1	609	1	252
861691	M	62	13	RAEB	45_XY-7[15]46_XY[5]				0	0	1	0	3.0	0	3144	0	118
875301	F	62	14	RAEB	46_XX[20]				0	0	0	0	1.5	1	463	1	81
879047	F	74	17	RAEB	46_XX.del(11)(q21q24)[13]46_XX[7]				0	0	0	0	2.5	1	980	0	980
882525	M	57	7	RAEB	41-49.XY.+X.XX.del(5)(q22q35).del(7)(q22).+8.-12.add(12)(p12).-13.-18.+1-4mar[cp20]				1	0	1	1	2.0	1	148	1	88
889352	M	72	14	RAEB	39-44.XY.del(4)(q31)[14].der(5)(5:?)q21.?)[17].del(6)(q21)[19].-7[19].add(12)(p13)[18].-18[18]cp19]				0	0	1	0	3.0	1	42	0	42
911999	M	36	4	RA	46_XY[20]				0	0	0	0	0.0	0	3250	0	3232
912290	M	59	1	RA	46_XY[20]				0	0	0	0	0.5	1	696	0	696
917011	M	79	0	RA	46_XY.del(20)(q11.2)[12]46_XY[8]				0	1	0	0	0.5	1	809	0	769
936943	M	43	ND	RAEB	failed.				0	0	0	0	UK	1	1300	0	1289
937954	M	51	12	RAEB	47_XY.+8[1]47.idem.del(16)(q12.1)[13]48.idem.+Y.del(16)(q11.1)[5]46_XY[1]				0	0	0	1	2.5	1	729	0	131
943264	M	65	11	RAEB	46_XY[20]				0	0	0	0	1.5	0	891	0	848
944541	M	54	20	RAEB	45_XY-7[19]46_XY[1]				0	0	1	0	3.0	1	1630	0	1630
947519	M	59	7	RAEB	46_XY.1-5dms[10]46_XY[12]				0	0	0	0	1.5	1	734	1	699
949197	F	60	2	RA	failed				0	0	0	0	UK	1	108	0	89
958595	F	82	4	RA	46_XX.del(11)(q23)[20]				0	0	0	0	1.0	1	975	0	975
973424	M	56	3	RA	46_XY[20]				0	0	0	0	0.0	0	1228	0	1027
975079	M	62	0	RA	46_XY[20]				0	0	0	0	0.0	1	1563	0	1563
983677	F	60	8	RAEB	46_XX.del(5)(q21q35)[4]69_XX.+X.+1.+1.+2.+del(5)(q21q35).+5.+6.+8.+9.+9.+10.+11.+11.+13.+13.+14.+14.+16.+16.+19.+19.+20.+21.-22.+mar1.+mar2]7]46_XX[9]				1	0	0	1	2.0	1	85	1	44
987089	F	64	3	RAEB	56-58.XX.+X.+del(3)(p12).del(5)(q13q33).+6.+8.+der(8)add(8q22).+10.+11.+13.+14.-17.+19.+21.+mar[cp20]				1	0	0	1	2.0	1	762	1	678
988428	M	70	2	RAR	47_XY-9.+2mar[1]46_XY[19]				0	0	0	0	0.5	1	1975	0	1891
989739	M	71	18	RAEB	46_XY.del(7)(q21)[11]47_XY.+8[5]46_XY[4]				0	0	1	1	3.0	1	468	1	125
994096	M	72	3	RA	46_XY.del(5)(q14q34).+8.del(11)(q21q23)[19]46.idem.del(17)(q24)[1]				1	0	0	1	1.0	1	1307	0	1178
999103	F	66	6	RAEB	46_XX.del(5)(q15q13)[20]				1	0	0	0	1.0	1	4813	0	2839
01-13	F	41	12	RAEB	46_XX.inv(3)(q21q26)[13]46.idem.t(14:17)(q13;p13)[1]				0	0	0	0	2.0	1	1649	1	1301
02-21	M	57	3	RA	46_XY.t(3:12)(q27;p13)[20]				0	0	0	0	1.0	1	709	0	709
02-28	M	60	1	RA	46_XY.t(5:15)(p10;q10)[2]46_XY[18]				0	0	0	0	1.0	1	76	0	76
02-30	M	72	2	RA	46-48.XY.add(5)(q11.2).-7oradd(7)(q11.2).add(8)(p22).del(13)(q21).add(16)(q722).del(22)(q12).+1-3mar[cp19]46_XY[1]				0	0	1	0	1.5	1	239	1	205
02-75	F	46	10	RAEB	46_XX[20]				0	0	0	0	0.5	0	3247	0	3014
02-77	M	67	28	RAEB-T	46_XY[20]				0	0	0	0	2.0	1	1863	1	1662
02-86	M	53	21	RAEB-T	92-<4n>-XXYY.del(4)(q33)x2.del(5)(q13).add(11)(q23)x2[cp8]46_XY[12]				1	0	0	0	3.0	1	308	1	133
02-99	M	66	16	RAEB	46_XY.del(2)(q33).del(5)(q15q33).-7.+del(11)(q273)[cp6]				1	0	1	0	3.0	1	945	0	933

Supplementary Table 8. Primers used for U2AF1 sequencing.							
Ensembl TranscriptID	Chromosome	Exon#	Exon Start	Exon Stop	Primer 1 Sequence*	Primer 2 Sequence	Ensembl ExonIDs
gDNA primers: targets U2AF1 hotspot only							
ENST00000398137	21	2	43397494	43397581	tgtaactcAAACAAACCTGGTAAACGTCG	acatgatgcaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	acgactacagAAACAAACCTGGTAAACGTCG	tgctgatgcAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	cgtgactagAAACAAACCTGGTAAACGTCG	gcatctgacAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	taagatgatAAACAAACCTGGTAAACGTCG	atgtcctacAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	tactctgtgAAACAAACCTGGTAAACGTCG	atgagagcacAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	tagagagagAAACAAACCTGGTAAACGTCG	atctctgtcAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	tgctgctcgAAACAAACCTGGTAAACGTCG	agcagcagcAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	acatcagcgtAAACAAACCTGGTAAACGTCG	tgtagcgcaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	acgcgactatAAACAAACCTGGTAAACGTCG	tgctcgataAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	actactatgAAACAAACCTGGTAAACGTCG	tgtagatacAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	actgacagtAAACAAACCTGGTAAACGTCG	tgcatgtcaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	agcactataAAACAAACCTGGTAAACGTCG	tcgatatgaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	agrcctgctAAACAAACCTGGTAAACGTCG	tgccagcagaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	agtagcgtAAACAAACCTGGTAAACGTCG	tcactgataAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	caagctagAAACAAACCTGGTAAACGTCG	ggtgagtgaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	tgtagagcgtAAACAAACCTGGTAAACGTCG	gtcatctgcaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	cgacgtgacAAACAAACCTGGTAAACGTCG	gctgcaactgaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	taacacactAAACAAACCTGGTAAACGTCG	atgtgtggaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	taacagtgatAAACAAACCTGGTAAACGTCG	atgtgcaactAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	taacagatgAAACAAACCTGGTAAACGTCG	atgtctagcaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	taacgtgtcAAACAAACCTGGTAAACGTCG	atgcagcagaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	taagttagaAAACAAACCTGGTAAACGTCG	atcacatcaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	tgatcacgtAAACAAACCTGGTAAACGTCG	agctagtgaAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	tgccactagAAACAAACCTGGTAAACGTCG	agcctgacAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
cDNA primers: targets U2AF1 hotspot only							
ENST00000398137	21	1-4	43393632	43400757	acagtgctgGGAGTATCTGGCCTCCATCTTC	tgctcacgcaGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	acgtctgacaGGAGTATCTGGCCTCCATCTTC	tgccagctgtGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	agaccctacGGAGTATCTGGCCTCCATCTTC	tcctgctgtagGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	agactgttagGGAGTATCTGGCCTCCATCTTC	tcgtgacacGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	atcacagcagGGAGTATCTGGCCTCCATCTTC	tagctgtgtcGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	atatcgcagGGAGTATCTGGCCTCCATCTTC	tatagcctcGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	cgltgctcaGGAGTATCTGGCCTCCATCTTC	gcacagagatGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	ctcgcgtgtGGAGTATCTGGCCTCCATCTTC	gagcgcacagGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
ENST00000398137	21	1-4	43393632	43400757	taqtatcagGGAGTATCTGGCCTCCATCTTC	atcatatcgGTCAGCAGACTGGGAAGAGTTT	ENSE00001582950, ENSE00001531800, ENSE00001531797, ENSE00001531794
3730 recurrent screening primers: targets all U2AF1 coding exons							
ENST00000398137	21	9	43386136	43386428	tgtaaaacgcgcagcITGAGAGAGTGGGTGGTTCTG	caggaaacagctatgaccCCCTTATGAACTGGTTGGTCAAT	ENSE00001138177
ENST00000398137	21	9	43386136	43386428	tgtaaaacgcgcagcGAGAGTGGGTGGTTCGTGC	caggaaacagctatgaccTCTTACGATCTCTGACCCG	ENSE00001138177
ENST00000398137	21	8	43387650	43387742	tgtaaaacgcgcagcITAACCGTGGTTAATGGACGCC	caggaaacagctatgaccGAGGGTGGTAGGGGAGAAAGAA	ENSE00001268168
ENST00000398137	21	7	43387834	43387967	tgtaaaacgcgcagcITAAACCGTGGTTAATGGACGCC	caggaaacagctatgaccGAGGGTGGTAGGGGAGAAAGAA	ENSE00001050518
ENST00000398137	21	7	43387834	43387967	tgtaaaacgcgcagcITAAAGTCTTATAAAGCGTGGATGGC	caggaaacagctatgaccCACTACTGACGGCAGCAGCG	ENSE00001050518
ENST00000398137	21	6	43388617	43388715	tgtaaaacgcgcagcITGATCTTCTGTACTGTTGGAAAG	caggaaacagctatgaccCACTGTCTGCTTTGCGCTG	ENSE00001050529
ENST00000398137	21	5	43388873	43388922	tgtaaaacgcgcagcITGTTGTGTACTGTAGCAGTTGTAG	caggaaacagctatgaccACTCTGAAGGGAGACCACAG	ENSE00001531791
ENST00000398137	21	4	43393632	43393698	tgtaaaacgcgcagcITCAACTCACTGAAAGCCTTGCC	caggaaacagctatgaccCCTTGCCACTCTAGTGAAACG	ENSE00001531794
ENST00000398137	21	3	43394545	43394611	tgtaaaacgcgcagcITGCCATTTGCAACAATAATGTCTCT	caggaaacagctatgaccCAGCAGCTCAGAACGCCAACAT	ENSE00001531797
ENST00000398137	21	2	43397494	43397581	tgtaaaacgcgcagcITGAACCAAAATGGAAAATCAACTACG	caggaaacagctatgaccTTCAAATTTGAGCATGTCTGCT	ENSE00001531800
ENST00000398137	21	2	43397494	43397581	tgtaaaacgcgcagcITAAACAAACCTGGTAAACGTCG	caggaaacagctatgaccAATTTGATAGTGTATGTCATGCTGCTG	ENSE00001531800
ENST00000398137	21	1	43400630	43400757	tgtaaaacgcgcagcITAGAGCGTCTGCCTGCTCCG	caggaaacagctatgaccCCACCCTCAACCACCC	ENSE00001582950

*lower case, bar code; upper case,U2AF1 sequence

Supplementary Table 9. Expression of Tier 1 mutated genes in sAML sample.

Chromosome	Position*	Gene name	Transcript ID	Exons	Mutant Exon	Total Probes	Present Probes				Probes in Mutant Exon	Present Probes in Mutant Exon			
							in UPN	in > 50% of AML	in > 75% of	in > 90% of AML		in UPN	in > 50% of AML	in > 75% of AML	in > 90% of AML
							266395	samples	AML samples	samples		266395	samples	samples	samples
1	112325945	KCND3	NM_004980	8	2	20	3	2	2	1	3	0	0	0	0
1	205309412	PKFEB2	NM_006212	15	11	29	2	15	9	4	1	0	1	0	0
1	229096537	ENSG00000222671	ENST00000410739	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	20289990	LOC100131373	XM_001720891	1	1	3	1	3	1	1	1	1	0	0	0
2	61569230	XPO1	NM_003400	25	18	38	31	33	33	33	1	1	1	1	1
3	173529456	FNDC3B	NM_022763	25	11	43	41	43	43	42	2	2	2	2	2
5	50152869	PARP8	NM_024615	26	16	31	28	30	29	27	1	1	1	1	1
5	148880083	CSNK1A1	NM_001025105	11	4	26	24	25	24	24	1	1	1	1	1
6	105713110	POPDC3	NM_022361	4	4	13	0	1	0	0	3	0	0	0	0
9	89535224	CTSL1	NM_001912	8	7	20	9	11	5	5	1	1	1	1	1
9	138867831	MAMDC4	NM_206920	27	4	49	12	13	11	9	1	0	0	0	0
9	139117274	MAN1B1	ENST00000371587	9	9	45	17	21	17	15	8	5	6	5	5
10	82288101	SH2D4B	NM_207372	7	1	17	4	6	4	1	1	0	0	0	0
11	22340860	SLC17A6	NM_020346	12	splice_site	15	2	2	0	0	NA	NA	NA	NA	NA
11	30881639	DCDC5	ENST00000406071	15	12	54	3	3	3	2	2	0	0	0	0
14	75418958	TTLL5	NM_015072	32	30	50	41	45	44	40	2	2	2	2	2
16	7716860	ENSG00000209555	ENST00000386820	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16	63589987	CDH11	NM_001797	13	4	28	5	8	4	4	2	0	0	0	0
17	5288451	DHX33	NM_020162	12	12	17	6	13	11	7	1	0	0	0	0
17	7696343	JMJD3	NM_001080424	22	19	44	19	27	22	21	1	0	0	0	0
17	29981226	TMEM132E	NM_207313	10	6	13	4	3	2	2	1	0	0	0	0
17	75696250	GAA	NM_000152	20	5	27	5	6	4	2	1	0	0	0	0
18	46165704	C18orf24	NM_001039535	7	5	14	3	6	2	2	1	0	1	0	0
19	53579461	KDELR1	NM_006801	5	4	12	7	8	8	6	2	1	2	2	1
19	54349852	HRC	NM_002152	6	1	16	1	3	3	2	9	1	2	2	2
20	32728482	PIGU	NM_080476	12	1	18	16	16	15	12	1	1	1	1	0
20	32965857	ACSS2	NM_018677	18	7	28	18	16	11	7	2	1	1	1	0
20	51626823	ZNF217	NM_006526	5	3	15	15	15	15	12	5	5	5	5	5
21	42402818	UMODL1	NM_173568	22	10	28	4	5	4	4	1	0	0	0	0
21	43397525	U2AF1	NM_001025203	8	2	22	18	22	22	20	1	1	1	1	1

*using NCBI Build 36. NA, not available.

Supplementary References

1. Mardis, E.R. et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-66 (2009).
2. Kralovicova, J. & Vorechovsky, I. Allele-specific recognition of the 3' splice site of INS intron 1. *Hum Genet* **128**, 383-400 (2010).
3. Graubert, T.A. et al. Integrated genomic analysis implicates haploinsufficiency of multiple chromosome 5q31.2 genes in de novo myelodysplastic syndromes pathogenesis. *PLoS ONE* **4**, e4583 (2009).
4. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-21 (2001).
5. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
6. Dennis, G., Jr. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).