Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome

Syed A. Mian,[1]* Alexander E. Smith,[1]* Austin G. Kulasekararaj,[1,2] Aytug Kizilors,[2] Azim M. Mohamedali,[1] Nicholas C. Lea,[2] Konstantinos Mitsopoulos,[3] Kevin Ford,[1] Erick Nasser,[1] Thomas Seidl,[1] and Ghulam J. Mufti[1,2]

[1]King's College London School of Medicine, Department of Haematological Medicine, London, UK; [2]Kings's College Hospital, Department of Haematology, London,UK; and [3]Kings's College Hospital, Department of Histopathology, London; and [3]ICR Breakthrough Breast Cancer

 doi:10.3324/haematol.2012.075325

Online Supplementary Methods

Amplicon sequencing

Amplicon sequencing was used to perform a mutation screen in 154 cases of MDS using a 22-gene panel. We sequenced all coding exons for *ZRSR2*, *TP53*, *DNMT3A*, *EZH2*, *RUNX1*, *CEBPA* and *TET2*. This included mutation data for *TET2* for 142 cases that were available from our previously published study.[1] Mutational hotspots were specifically sequenced for *SF3B1*, *SRSF2*, *U2AF1*, *ASXL1*, *CCBL*, *FLT3*, *NRAS*, *KRAS*, *JAK2* exon12/14, *IDH1*, *IDH2*, *BRAF*, *MPL*, *C-KIT* and *NPM1* in all cases. Gene hotspot regions were selected based on previously published data and frequency of the mutations shown in the COSMIC database (*Online Supplementary Table S2*). In addition, all coding exons of *SF3B1* were sequenced for all 24 RARS/RCMD-RS patients. *SRSF2*, *U2AF1* and *ZRSR2* genes were amplified by using polymerase chain reaction (PCR) primers published previously.[2] PCR primers for all the genes are shown in *Online Supplementary Table S3*. The PCR and sequencing methodology have been described previously.[1] The average sequencing coverage across all genes was 200X and >90% of the coding regions had a coverage of >100X. This coverage enabled us to reliably detect mutant clones down to ≥5-10% mutant allele burden, defined as the proportion of sequence reads containing the mutation. All mutations were confirmed through independent PCR and GS FLX sequencing/Sanger sequencing experiments. For data relating to samples at/prior to transformation to acute myeloid leukemia (AML), for which 454-amplicon sequencing data were not available, relative peak intensity from Sanger sequencing was used to estimate mutant allele burden. The acquired status of novel mutations was also confirmed in 48/54 cases for whom constitutional source of DNA was available: skin biopsy (n=27), $CD3^+$ T cells (n=18) and buccal swab (n=3) (*Online Supplementary Table S6*).

Exome sequencing

Eight patients (5 RARS, 1 RARS-T, 1 RCMD-RS and 1 tMDS) with >50% ringed sideroblasts were selected for whole-exome sequencing, using DNA from $CD34^+$ cells in all cases and paired constitutional source [skin (n=3), $CD3^+$ T cells (n=3) and $CD34^-CD235^+$ (n=1)] (*Online Supplementary Table S3*). The exomic regions of the genome were enriched using Agilent SureSelect Human All Exon Kit covering approximately 50 Mb of the genome. This was followed by paired-end sequencing with a read length of 75 bp using Illumina HiSeq 2000 and version 3 chemistry. The base calls generated by the real time analysis (RTA) or the off-line base-caller (OLB) software were de-multiplexed and converted to fastq format using Casava 1.8. Alignment (NCBI37/Hg19) and variant calling for single nucleotide polymorphisms (SNP) and Indels was performed using both Casava 1.8 (Illumina), and BWA/GATK according to the Broad Institute best practices. Annovar software[3] was used for functional annotation of all variants and also to remove SNP reported in the 1000 Genomes Project, dbSNP132, and genomic super duplications databases. For Casava, remaining variants with a quality score of >QSNP 25 and with a mutant read depth of >1 (total read depth >6) were passed for further analysis. A repeat sequencing run was also used to filter out sequencing artefacts in case of which candidate mutations supported in both runs were passed. For GATK, variants were selected that passed standard filtering, Additionally, variations present in available paired skin and $CD3^+$ T-cell samples at greater than 20% or 50% the level found in $CD34^+$ tissue were designated germline and excluded from subsequent analysis of acquired mutations. Data from both pipelines were largely in agreement for filter-passed variants and variant depth >1. Any discrepancies between the two pipelines at this level were later confirmed to be artefactual calls.

PCR and Sanger sequencing were used to validate selected candidate mutations. An ExoSAP-IT purification kit was used for the purification of PCR products. Sequencing was performed using a Big Dye Terminator V 3.1 kit, according to the manufacturer's protocol. Sanger sequencing was performed using an ABI3010xl (Applied Biosystems) and the sequencing results were analyzed by SeqScape software.

Statistical analysis

The characteristics of the study population were studied with appropriate statistical methods (Mann-Whitney test for continuous variables and Fisher's exact test for categorical

variables) comparing patients with specific mutations *versus* cases without mutations. Clinical characteristics, survival and time to progression to AML were updated to January 2012 and measured from time of sample collection (N=123) or diagnosis (N=31) at King's College Hospital. The median disease duration prior to sample analysis for 123 patients was 11 months (range, 1-119 months): during this period 29 patients showed disease progression (upstaging of WHO category) of whom 10 were treated with either intensive chemotherapy or 5-azacitidine.

The Kaplan-Meier estimate was used to evaluate time to survival and time to progression. The log-rank test was used to assess potential differences in outcome between subgroups. A *P*-value of ≤0.05 was considered statistically significant.

Online Supplementary Results

Exome sequencing reveals common SF3B1 mutations in patients with refractory anemia and ringed sideroblasts

Whole exome sequencing (Illumina) was performed on $CD34^+$ cells from eight patients all of whom had >50% ringed sideroblasts: five with RARS, one with RARS-T, one with RCMD-RS and one with therapy-related MDS (*Online Supplementary Table S3*). Paired constitutional DNA from skin (3 cases) or $CD3^+$ T cells (3 cases) was similarly sequenced. An average of 8 Gbp of sequence data was generated per patient exome and processed using both the Casava 1.8 (Illumina) and Broad Institute 'Best practise' pipelines, supplemented with additional software tools as detailed in the Design and Methods. Functional variants not found in dbSNP132 or 1000 genomes databases were subsequently collated.

Aberrations in one particular gene, splicing factor 3b subunit 1 (*SF3B1*), a component of the major and minor spliceosomes, initially stood out because of its high frequency across all patients (7 of 8 cases). Importantly, a *TP53* mutation was observed in a single case with therapy-related MDS who had wild-type *SF3B1*. Furthermore, mutations in epigenetic factors including *TET2* and *DNMT3A* were found in three cases with *SF3B1* aberrations. Validation of these selected mutations in available paired skin biopsy or $CD3^+$ T-cell samples by Sanger sequencing led to the confirmation of the acquired nature of these mutations. For six patients with paired samples, an additional 61 patient-specific mutation candidates (10 per patient exome on average) were identified and selectively validated via Sanger sequencing (*Online Supplementary Tables S4* and *S5*).

SF3B1 *mutations persist through differentiation*

Initially $CD34^+$ cell samples (n=8) were subjected to exome sequencing and subsequent screening/confirmation of *SF3B1* mutations was performed on total bone marrow cells. We also quantified the mutation load in $CD34^+$, total $CD3^+$ T cells, $CD3^+CD4^+$ T cells, $CD19^+$ B cells, CD235+$CD71^+$ erythroblasts and $CD34^-CD3^-CD235^-$ for three additional patients in whom SF3B1 mutations were detected in total bone marrow cells. $CD235^+CD71^+$ erythroblasts and $CD34^-CD3^-CD235^-$ cells showed approximately the same *SF3B1* mutant allele burden (according to Sanger sequencing) as seen for paired total bone marrow and $CD34^+$ cells. Importantly, no *SF3B1* mutation was detected above background in $CD3^+$ T cells, $CD3^+CD4^+$ T-cells or $CD19^+$ B cells according to Sanger sequencing (3/3 mutant patients), indicating that the mutant clone has a growth advantage only in the myeloid lineage in MDS. Mutant *SF3B1* was not detected in any available paired skin biopsy samples, confirming the acquired nature of these mutations.

References

1. Smith AE, Mohamedali AM, Kulasekararaj A, Lim Z, Gäken J, Lea NC, Przychodzen B, et al. Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low abundance mutant clones with early origins, but indicates no definite prognostic value. Blood. 2010;116(19):3923-32.
2. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478(7367):64-9.
3. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
4. Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med. 2011;365(15):1384-95.
5. Gelsi-Boyer V, Trouplin V, Adélaïde J, Bonansea J, Cervera N, Carbuccia N, Lagarde A, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br J Haematol. 2009;145(6):788-800.
6. Carbuccia N, Trouplin V, Gelsi-Boyer V, Murati A, Rocquain J, Adélaïde J, et al. Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. Leukemia. 2010;24(2):469-73.
7. Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrózek K, et al. ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. Blood. 2011;118(26):6920-9.
8. Nakao M, Yokota S, Iwai T, Kaneko H, Horiike S, Kashima K, et al. Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. Leukemia. 1996;10(12):1911-8.
9. Thiede C, Steudel C, Mohr B, Schaich M, Schäkel U, Platzbecker U, et al. Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. Blood. 2002;99(12):4326-35.
10. Jiang J, Paez JG, Lee JC, Bo R, Stone RM, DeAngelo DJ, et al. Identifying and characterizing a novel activating mutation of the FLT3 tyrosine kinase in AML. Blood. 2004;104(6):1855-8.
11. Loriaux MM, Levine RL, Tyner JW, Fröhling S, Scholl C, Stoffregen EP, et al. High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. Blood. 2008;111(9):4788-96.
12. Hirai H, Kobayashi Y, Mano H, Hagiwara K, Maru Y, Omine M, et al. A point mutation at codon 13 of the N-ras oncogene in myelodysplastic syndrome. Nature. 1987;327 (6121):430-2.
13. Bacher U, Haferlach T, Kern W, Haferlach C, Schnittger S. A comparative study of molecular mutations in 381 patients with myelodysplastic syndrome and in 4130 patients with acute myeloid leukemia. Haematologica. 2007;92(6):744-52.
14. Tyner JW, Erickson H, Deininger MW, Willis SG, Eide CA, Levine RL, et al. High-throughput sequencing screen reveals novel, transforming RAS mutations in myeloid leukemia patients. Blood. 2009;113(8):1749-55.
15. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet. 2005;365 (9464):1054-61.
16. Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med. 2005;352(17):1779-90.
17. Sanada M, Suzuki T, Shih LY, Otsu M, Kato M, Yamazaki S, et al. Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. Nature. 2009;460(7257): 904-8.
18. Makishima H, Cazzolli H, Szpurka H, Dunbar A, Tiu R, Huh J, et al. Mutations of e3 ubiquitin ligase cbl family members constitute a novel common pathogenic lesion in myeloid malignancies. J Clin Oncol. 2009;27(36):6109-16.
19. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009;361(11):

1058-66.
20. Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. N Engl J Med. 2005;352(3):254-66.
21. Pardanani AD, Levine RL, Lasho T, Pikman Y, Mesa RA, Wadleigh M, et al. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. Blood. 2006;108(10):3472-6.
22. Beer PA, Ortmann CA, Stegelmann F, Guglielmelli P, Reilly JT, Larsen TS, et al. Molecular mechanisms associated with leukemic transformation of MPL-mutant myeloproliferative neoplasms. Haematologica. 2010;95(12): 2153-6.
23. Patnaik MM, Lasho TL, Finke CM, Gangat N, Caramazza D, Siragusa S, et al. MPL mutation effect on JAK2 46/1 haplotype frequency in JAK2V617F-negative myeloproliferative neoplasms. Leukemia. 2010;24(4):859-60.
24. Bowen DT, Padua RA, Burnett AK, deCastro CM, Kaufman RE. Two new polymorphisms but no mutations of the KIT gene in patients with myelodysplasia at positions corresponding to human FMS and murine W locus mutational hot spots. Leukemia. 1993;7(11):1883-5.
25. Lorenzo F, Nishii K, Monma F, Kuwagata S, Usui E, Shiku H. Mutational analysis of the KIT gene in myelodysplastic syndrome (MDS) and MDS-derived leukemia. Leuk Res. 2006;30(10):1235-9.
26. Jankowska AM, Makishima H, Tiu RV, Szpurka H, Huang Y, Traina F, et al. Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. Blood. 2011;118(14):3932-41.
27. Zhang SJ, Rampal R, Manshouri T, Patel J, Mensah N, Kayserian A, et al. Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood. 2012;119(19): 4480-5.
28. Green A, Beer P. Somatic mutations of IDH1 and IDH2 in the leukemic transformation of myeloproliferative neoplasms. N Engl J Med. 2010; 362(4):369-70.
29. Jeziskova I, Razga F, Bajerova M, Racil Z, Mayer J, Dvorakova D. IDH2 mutations in patients with acute myeloid leukemia: missense p.R140 mutations are linked to disease status. Leuk Lymphoma. 2010;51(12):2285-7.
30. Pardanani A, Guglielmelli P, Lasho TL, Pancrazzi A, Finke CM, Vannucchi AM, Tefferi A. Primary myelofibrosis with or without mutant MPL: comparison of survival and clinical features involving 603 patients. Leukemia. 2011;25(12): 1834-9.
31. Thol F, Kade S, Schlarmann C, Löffeld P, Morgan M, Krauter J, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood. 2012;119(15):3578-84.
32. Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. Blood. 2012;119(14):3211-8.
33. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature. 2002;417(6892):949-54.
34. Beer PA, Delhommeau F, LeCouédic JP, Dawson MA, Chen E, Bareford D, et al. Two routes to leukemic transformation after a JAK2 mutation-positive myeloproliferative neoplasm. Blood. 2010;115(14):2891-900.
35. Patnaik MM, Lasho TL, Hodnefield JM, Knudson RA, Ketterling RP, Garcia-Manero G, et al. SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. Blood. 2012; 119(2):569-72.
36. Grand FH, Hidalgo-Curtis CE, Ernst T, Zoi K, Zoi C, McGuire C, et al. Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. Blood. 2009; 113(24):6182-92.
37. Zhang SJ, Rampal R, Manshouri T, Patel J, Mensah N, Kayserian A, et al. Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms reveals recurrent SRSF2 mutations which are associated with adverse outcome. Blood. 2012;119(19):4480-5.
38. Beekman R, Valkhof MG, Sanders MA, van Strien PM, Haanstra JR, Broeders, et al. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. Blood. 2012;119(22):5071-7.
39. Pietra D, Li S, Brisci A, Passamonti F, Rumi E, Theocharides A, et al. Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders. Blood. 2008;111(3):1686-9.
40. Ricci C, Spinelli O, Salmoiraghi S, Finazzi G, Carobbio A, Rambaldi A. ASXL1 mutations in primary and secondary myelofibrosis. Br J Haematol. 2012;156(3):404-7.
41. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet. 2010;42(8):722-6.
42. Stein BL, Williams DM, O'Keefe C, Rogers O, Ingersoll RG, Spivak JL, et al. Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: analysis of molecular genetics and clinical phenotypes. Haematologica. 2011;96(10):1462-9.
43. Tzankov A, Sotlar K, Muhlematter D, Theocharides A, Went P, Jotterand M, et al. Systemic mastocytosis with associated myeloproliferative disease and precursor B lymphoblastic leukaemia with t(13;13)(q12;q22) involving FLT3. J Clin Pathol. 2008;61(8):958-61.
44. Mullighan CG, Zhang J, Kasper LH, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337): 235-9.
45. Jerez A, Sugimoto Y, Makishima H, Verma A, Jankowska AM, Przychodzen B, et al. Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis. Blood. 2012;119(25):6109-17.
46. Metzeler KH, Maharry K, Radmacher MD, Mrózek K, Margeson D, Becker H, et al. TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol. 2011;29(10):1373-81.
47. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med. 2010;363(25):2424-33.
48. Gaidzik VI, Paschka P, Späth D, Habdank M, Köhne CH, Germing U, et al. TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. J Clin Oncol. 2012;30 (12):1350-7.
49. Malcovati L, Papaemmanuil E, Bowen DT, Boultwood J, Della Porta MG, Pascutto C, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood. 2011; 118(24):6239-46.
50. Chen CY, Lin LI, Tang JL, Ko BS, Tsay W, Chou WC, et al. RUNX1 gene mutation in primary myelodysplastic syndrome--the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. Br J Haematol. 2007;139(3):405-14.
51. Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL, et al. Array-based genomic resequencing of human leukemia. Oncogene. 2010;29(25):3723-31.
52. Dicker F, Haferlach C, Sundermann J, Wendland N, Weiss T, Kern W, et al. Mutation analysis for RUNX1, MLL-PTD, FLT3-ITD, NPM1 and NRAS in 269 patients with MDS or secondary AML. Leukemia. 2010;24(8): 1528-32.
53. Gaidzik VI, Bullinger L, Schlenk RF, Zimmermann AS, Röck J, Paschka P, et al. RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. J Clin Oncol. 2011;29(10):1364-72.
54. Chou WC, Lei WC, Ko BS, Hou HA, Chen CY, Tang JL, et al. The prognostic impact and stability of Isocitrate dehydrogenase 2 mutation in adult patients with acute myeloid leukemia. Leukemia. 2011;25(2):246-53.
55. Tatton-Brown K, Hanks S, Ruark E, Zachariou A, Duarte Sdel V, Ramsay E, et al. Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. Oncotarget. 2011;2(12):1127-33.
56. Dicker F, Haferlach C, Kern W, Haferlach T, Schnittger S. Trisomy 13 is strongly associated with AML1/RUNX1 mutations and increased FLT3 expression in acute myeloid leukemia. Blood. 2007;110(4):1308-16.
57. Silva FP, Almeida I, Morolli B, Brouwer-Mandema G, Wessels H, Vossen R, et al. Genome wide molecular analysis of minimally differentiated acute myeloid leukemia. Haematologica. 2009;94(11):1546-54.
58. Visconte V, Makishima H, Jankowska A, Szpurka H, Traina F, Jerez A, et al. SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. Leukemia. 2012;26(3):542-5.
59. Sugimoto Y, Muramatsu H, Makishima H, Prince C, Jankowska AM, Yoshida N, et al. Spectrum of molecular defects in juvenile myelomonocytic leukaemia includes ASXL1 mutations. Br J Haematol. 2010;150(1):83-7.
60. Irahara N, Baba Y, Nosho K, Shima K, Yan L, Dias-Santagata D, et al. NRAS mutations are rare in colorectal cancer. Diagn Mol Pathol. 2010;19(3):157-63.
61. Rossi D, Bruscaggin A, Spina V, Rasi S, Khiabanian H, Messina M, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood. 2011;118(26): 6904-8.
62. Christiansen DH, Andersen MK, Pedersen-Bjergaard J. Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation. Blood. 2004;104(5):1474-81.
63. Dunbar AJ, Gondek LP, O'Keefe CL, Makishima H, Rataul MS, Szpurka H, et al. 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. 2008;68(24):10349-57.
64. Abbas S, Lugthart S, Kavelaars FG, Schelen A, Koenders JE, Zeilemaker A, et al. Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. Blood. 2010;116(12):2122-6.
65. Ewalt M, Galili NG, Mumtaz M, Churchill M, Rivera S, Borot F, et al. DNMT3a mutations in high-risk myelodysplastic syndrome parallel those found in acute myeloid leukemia. Blood Cancer J. 2011;1(3):e9.
66. Olcaydu D, Rumi E, Harutyunyan A, Passamonti F, Pietra D, Pascutto C, et al. The role of the JAK2 GGCC haplotype and the TET2 gene in familial myeloproliferative neoplasms. Haematologica. 2011;96(3):367- 74.

Online Supplementary Table S1. Treatments received by 154 patients. †Denotes other treatments which include: lenalidomide, thalidomide, cyclosporine and antithymocyte globulin (ATG). *Includes 13 patients who had HSCT after receiving 5-azacitidine.

Treatment	No of Cases
5-Azacytidine	68
Other Treatments†	14
HSCT*	35
No treatment	37

Online Supplementary Table S2. Genes screened for hotspot mutations. Mutation frequency within the hotspot regions was obtained from the Cosmic database (Welcome Trust Sanger Institute, *http://www.sanger.ac.uk/genetics/CGP/cosmic/*).

Gene	Total Frequency of Mutations in MDS	Hot Spot Regions Sequenced	Previous Publications	COSMIC Database (Welcome Trust Sanger Institute, http://www.sanger.ac.uk/genetics/CGP/cosmic/)
SF3B1	*30%*	Exon 12 - 16	Yoshida, et al Nature (2011); Papaemmanuil, et al. NEJM (2011)	Shows 99% of SF3B1 somatic mutations in haematopoietic neoplasms within this region (out of 344 reported mutant cases; 2668 sequenced samples)
SRSF2	*15%*	Exon 1	Yoshida, et al. Nature (2011)	Shows 100% of SRSF2 somatic mutations in haematopoietic neoplasms within this region (out of 279 reported mutant cases; 1872 sequenced samples)
ASXL1	*15%*	Exon 12	Gelsi-Boyer, et al. BJH (2009); Carbuccia, et al. LEukemia (2009); Metzeler, et a. Blood (2011)	Shows 99% of ASXL1 somatic mutations in haematopoietic neoplasms within this region (out of 472 reported mutant cases; 3590 sequenced samples)
U2AF1	*12%*	Exon 2, 6	Yoshida, et al. Nature (2011)	Shows 99% of U2AF35 somatic mutations in haematopoietic neoplasms within this region (out of 96 reported mutant cases; 1518 sequenced samples)
FLT3	*6%*	Exon14, 15	Nakao, et al. Leukemia (1996); Thiede, et al. Blood (2002); Jiang et al. Blood (2004); Loriaux et al. Blood (2008)	Shows 100% of FLT3 somatic mutations in haematopoietic neoplasms within this region (out of 11816 reported mutant cases; 51551 sequenced samples)
NRAS	*3.6%*	Exon 2, 3	Hirai, et al. Nature (1987); Bacher, et al. Haematologica (2007); Tyner, et al. Blood (2008)	Shows 99% of NRAS somatic mutations in haematopoietic neoplasms within this region (out of 676 reported mutant cases; 6473 sequenced samples)
JAK2	*3%*	Exon12,14	Baxter et al. Lancet (2005); Kralovics, et al. NEJM (2005)	Shows 100 % of JAK2 somatic mutations in haematopoietic neoplasms within this region (out of 32218 reported mutant cases; 76213 sequenced samples)
CCBL	*2.3%*	Exon 7, 8, 9	Sanada, et al. (2009); Makishima, et al. (2009)	Shows 99% of CCBL somatic mutations in haematopoietic neoplasms within this region (out of 163 reported mutant cases; 3577 sequenced samples)
IDH2	*2.1%*	Exon 4	Mardis, et al. NEJM (2009)	Shows 100 % of IDH2 somatic mutations in haematopoietic neoplasms within this region (out of 905 reported mutant cases; 13777 sequenced samples)
NPM1	*1.8%*	Exon 12	Falini, et al. NEJM (2005)	Shows ≥ 97 % of NPM1 somatic mutations in haematopoietic neoplasms within this region (4181 reported mutant cases; 14167 sequenced samples)
IDH1	*1.4*	Exon 4	Mardis, et al. NEJM (2009)	Shows 100 % of IDH1 somatic mutations in haematopoietic neoplasms within this region (out of 778 reported mutant cases; 16301 sequenced samples)
KRAS	*0.9%*	Exon 2, 3	Tyner, et al. Blood (2008)	Shows >97% of KRAS somatic mutations in haematopoietic neoplasms within this region (out of 152 reported mutant cases ; 3716 sequenced samples)
BRAF	*0.5%*	Exon 11, 15		Shows 100 % of BRAF somatic mutations in haematopoietic neoplasms within this region (6 reported mutant cases; 581 sequenced samples)
MPL	*Rare*	Exon 10	Pardanani et al. Blood (2006); Beer, et al. Haematologica (2010); Patnaik et al. Leukemia (2010)	Shows ≥ 97 % of MPL somatic mutations in haematopoietic neoplasms within this region (534 reported mutant cases; 14202 sequenced samples)
C-KIT	*Rare*	Exon 17	Bowen, et al. Leukemia (1993); Lorenzo, et al. Leukemia Research (2006)	Shows 40 % of C-KIT somatic mutations in haematopoietic neoplasms within this region (1672 reported mutant cases; 7734 sequenced samples)

Online Supplementary Table S3. Primers for 22 genes used for 454-sequencing. Primer names reflect exons being sequenced. Universal sequencing primers for 2nd round PCR and Sanger sequencing are in lower case.

Genes	Primer Sequence
CMPL_F	gtagtgcgatggccagtAGCCTGGATCTCCTTGGTGAC
CMPL_R	cagtgtgcagcgatgacCGAGTCCCAGAGGTGACGT
JAK2_E12F	gtagtgcgatggccagtCAGAACGAATGGTGTTTCTGA
JAK2_E12R	cagtgtgcagcgatgacCCAATGTCACATGAATGTAAATCAA
JAK2_E14F	gtagtgcgatggccagtTCCTCATCTATAGTCATGCTGAAA
JAK2_E14R	cagtgtgcagcgatgacCTGACACCTAGCTGTGATCCTG
IDH1_F	gtagtgcgatggccagtGCGTCAAATGTGCCACTATC
IDH1_R	cagtgtgcagcgatgacTTCATACCTTGCTTAATGGGTGT
IDH2_F	gtagtgcgatggccagtAATTTTAGGACCCCCGTCTG
IDH2_R	cagtgtgcagcgatgacTGTGGCCTTGTACTGCAGAG
FLT3/KTD_F	gtagtgcgatggccagtCCGCCAGGAACGTGCTTG
FLT3/KTD_R	cagtgtgcagcgatgacGCAGCCTCACATTGCCCC
FLT3/ITD_F	gtagtgcgatggccagtCAATTTAGGTATGAAAGCCAGCTA
FLT3/ITD_R	cagtgtgcagcgatgacCTTTCAGCATTTTGACGGCAACC
CKIT_F	gtagtgcgatggccagtGTTTTCTTTTCTCCTCCAACC
CKIT_R	cagtgtgcagcgatgacGGACTGTCAAGCAGAG
NPM1_F	gtagtgcgatggccagtATTTCTTTTTTTTTTTTTCCAGGCTATTCAAG
NPM1_R	cagtgtgcagcgatgacGGTAGGGAAAGTTCTCACTCTGC
NRAS2_F	gtagtgcgatggccagtGATGTGGCTCGCCAATTAA
NRAS2_R	cagtgtgcagcgatgacGATTGTCAGTGCGCTTTTCC
NRAS3_F	gtagtgcgatggccagtATAGGCAGAAATGGGCTTGAATA
NRAS3_R	cagtgtgcagcgatgacGTATTGGTCTCTCATGGCACTGTA
KRAS2_F	gtagtgcgatggccagtGCCTGCTGAAAATGACTGAA
KRAS2_R	cagtgtgcagcgatgacGAATGGTCCTGCACCAGTAA
KRAS3_F	gtagtgcgatggccagtCCAGACTGTGTTTCTCCCTTC
KRAS3_R	cagtgtgcagcgatgacAAAGAAAGCCCTCCCCAGT
BRAF_11_F	gtagtgcgatggccagtTTTTCTGTTTGGCTTGACTTGA
BRAF_11_R	cagtgtgcagcgatgacACTTGTCACAATGTCACCACA
BRAF_15_F	gtagtgcgatggccagtTGCTTGCTCTGATAGGAAAATG
BRAF_15_R	cagtgtgcagcgatgacTGGAAAAATAGCCTCAATTCTTA
CCBL_E7_F	gtagtgcgatggccagtACACCACGTTGCCCTTTTAG
CCBL_E7_R	cagtgtgcagcgatgacAAGCTTGTGTCCAGTGATATGG
CCBL_E8_F	gtagtgcgatggccagtGGAAACAAGTCTTCACTTTTTCTGT
CCBL_E8_R	cagtgtgcagcgatgacAAAAAAGTCGCTGTTTAGATCCGTA
CCBL_E9_F	gtagtgcgatggccagtTGCATCTGTTACTATCTTTTGCTTC
CCBL_E9_R	cagtgtgcagcgatgacCTCACAATGGATTTTGCCA
DNMT3A_E2_F	gtagtgcgatggccagtGCCTCCAAAGACCACGATAA
DNMT3A_E2_R	cagtgtgcagcgatgacCGGCTGTCATCACATAGGG
DNMT3A_E3_F	gtagtgcgatggccagtGGTGGGGGCATATTACACAG
DNMT3A_E3_R	cagtgtgcagcgatgacTCCCTGCAGGACATACATCA
DNMT3A_E4a_F	gtagtgcgatggccagtAGAATTTCAGAGCGGTCAATG
DNMT3A_E4a_R	cagtgtgcagcgatgacCAGCCATTTTCCACTGCTCT
DNMT3A_E4b_F	gtagtgcgatggccagtATTACCCAATGGGGACTTGG
DNMT3A_E4b_R	cagtgtgcagcgatgacAAGCAGACCTTTAGCCACGA
DNMT3A_E5_F	gtagtgcgatggccagtGAACAGCTAAACGGCCAGAG
DNMT3A_E5_R	cagtgtgcagcgatgacCTTCCCACAGAGGGATGTGT
DNMT3A_E6_F	gtagtgcgatggccagtCATTGTGTTTGAGGCGAGTG
DNMT3A_E6_R	cagtgtgcagcgatgacTAGCCTGAAGGGGAAACTGA
DNMT3A_E7_F	gtagtgcgatggccagtTTCCTGGAGAGGTCAAGGTG
DNMT3A_E7_R	cagtgtgcagcgatgacTGGAGAGAGGAGAGCAGGAC
DNMT3A_E8_F	gtagtgcgatggccagtTCTTGCCTCATTCAGATGGA
DNMT3A_E8_R	cagtgtgcagcgatgacCCTGGGATCAAGAACCTTCC
DNMT3A_E9_F	gtagtgcgatggccagtCTCCAGGGCTGAGACTGACT
DNMT3A_E9_R	cagtgtgcagcgatgacACCTGCACTCCAACTTCCAG
DNMT3A_E10_F	gtagtgcgatggccagtCCTGTGCCACCCTCACTACT
DNMT3A_E10_R	cagtgtgcagcgatgacCTCCCTAAGCATGGCTTTCC
DNMT3A_E11_F	gtagtgcgatggccagtAGGTGGGAACAAGTTGGAGA
DNMT3A_E11_R	cagtgtgcagcgatgacAGAGCTGGCGTCAGAGGAG
DNMT3A_E12_F	gtagtgcgatggccagtTAGTTGGCCTGCTTCTGGAG
DNMT3A_E12_R	cagtgtgcagcgatgacGAGTCCCACACCCTGAAGAC
DNMT3A_E13_F	gtagtgcgatggccagtAGATGATGGCGTTCGAGACT
DNMT3A_E13_R	cagtgtgcagcgatgacTGGACACAGTCAGCCAGAAG

continued on the next page

continued from the previous page

Genes	Primer Sequence
DNMT3A_E14_F	gtagtgcgatggccagtCTCTGTGAGGCCAGGTGTG
DNMT3A_E14_R	cagtgtgcagcgatgacAGGTGTGCTACCTGGAATGG
DNMT3A_E15F	gtagtgcgatggccagtACCAGGGCTGAGAGTCTCCT
DNMT3A_E15_R	cagtgtgcagcgatgacAGGCTCCTAGACCCACACAC
DNMT3A_E16F	gtagtgcgatggccagtCAGGGTGTGTGGGTCTAGGA
DNMT3A_E16_R	cagtgtgcagcgatgacAAGCTTCCCCTTTGGGATAA
DNMT3A_E17_F	gtagtgcgatggccagtGACTTGGGCCTACAGCTGAC
DNMT3A_E17_R	cagtgtgcagcgatgacCAAAATGAAAGGAGGCAAGG
DNMT3A_E18_F	gtagtgcgatggccagtCAACTTGGTCCCGTTCTTGT
DNMT3A_E18_R	cagtgtgcagcgatgacCAAGGAGGAAGCCTATGTGC
DNMT3A_E19_F	gtagtgcgatggccagtGACAGCTATTCCCGATGACC
DNMT3A_E19_R	cagtgtgcagcgatgacGCTCCACAATGCAGATGAGA
DNMT3A_E20_F	gtagtgcgatggccagtCAGCTTGTGGAATGTGGCTA
DNMT3A_E20_R	cagtgtgcagcgatgacCACTATGGGTCATCCCACCT
DNMT3A_E21_F	gtagtgcgatggccagtATGACGTGTGTGCGTGATTT
DNMT3A_E21_R	cagtgtgcagcgatgacCCACACTAGCTGGAGAAGCA
DNMT3A_E22_F	gtagtgcgatggccagtTTTGGTAGACGCATGACCAG
DNMT3A_E22_R	cagtgtgcagcgatgacCAGGACGTTTGTGGAAAACA
DNMT3A_E23_F	gtagtgcgatggccagtTCCTGCTGTGTGGTTAGACG
DNMT3A_E23_R	cagtgtgcagcgatgacTTTTTCTCTTCTGGGTGCTGA
DNMT3A_E24_F	gtagtgcgatggccagtTGAAGGAGTATTTTGCGTGTG
DNMT3A_E24_R	cagtgtgcagcgatgacCAGAAAACCCCTCTGAAAAGA
ASXL1_E12_1F	gtagtgcgatggccagtGTTCACACAGTCCCACCAGAAA
ASXL1_E12_1R	cagtgtgcagcgatgacACTCTCTATGGCAGTGGTGACCT
ASXL1_E12_2F	gtagtgcgatggccagtGTGGACCCTCGCAGACATTAAA
ASXL1_E12_2R	cagtgtgcagcgatgacACTTGGGAGGCATCTCCTAGC
ASXL1_E12_3F	gtagtgcgatggccagtGTGGACTCACAGATGGGCTAGG
ASXL1_E12_3R	cagtgtgcagcgatgacACCCAAGCCCTAATTCGTCATC
ASXL1_E12_4F	gtagtgcgatggccagtGTAATCCTCACCGACTGATTGC
ASXL1_E12_4R	cagtgtgcagcgatgacACTGCTTCAGAGTCTCCGTTGA
ASXL1_E12_5F	gtagtgcgatggccagtGTTCACTCTGGACTGTGCCATC
ASXL1_E12_5R	cagtgtgcagcgatgacACGCAGCAACTGCATCACAAGT
ASXL1_E12_6F	gtagtgcgatggccagtGTTCCCAGATTCCCTACTGCTG
ASXL1_E12_6R	cagtgtgcagcgatgacACCTGGATGGAGGGAGTCAAAA
ASXL1_E12_7F	gtagtgcgatggccagtGTAAGGCAGTCCCAAGTTTTGA
ASXL1_E12_7R	cagtgtgcagcgatgacACGTTTGTGCTTGGTCCCAACT
ASXL1_E12_8F	gtagtgcgatggccagtGTGATGCCTCTTCCTGCTGAGA
ASXL1_E12_8R	cagtgtgcagcgatgacACAGCTGGTGGAACTCAGTTGG
ASXL1_E12_9F	gtagtgcgatggccagtGTCTTCTCTCCCCTCCCAACTC
ASXL1_E12_9R	cagtgtgcagcgatgacACACAGAGCTTTGAGGGTCCAA
RUNX1_E3_F	gtagtgcgatggccagtAGCTGTTTGCAGGGTCCTAA
RUNX1_E3_R	cagtgtgcagcgatgacCATCCCAAGCTAGGAAGACCGA
RUNX1_E4_F	gtagtgcgatggccagtGTATAACATCCCTGATGTCTGCA
RUNX1_E4_R	cagtgtgcagcgatgacCCATGAAACGTGTTTCAAGC
RUNX1_E5_F	gtagtgcgatggccagtATGTTCAGGCCACCAACCTCAT
RUNX1_E5_R	cagtgtgcagcgatgacGACATGGTCCCTGAGTATACCAG
RUNX1_E6_F	gtagtgcgatggccagtCGAGTCTATGTTGGGGTGAGGGGA
RUNX1_E6_R	cagtgtgcagcgatgacGAAACCCCAGTTGGTCTGGGA
RUNX1_E7B_F	gtagtgcgatggccagtATTTGAACAAGGGCCACTCAT
RUNX1_E7B_R	cagtgtgcagcgatgacCTCAGCTGCAAAGAATGTGT
RUNX1_E8A_F	gtagtgcgatggccagtGGAAGAGCTGTGGCCTCCGCAA
RUNX1_E8A_R	cagtgtgcagcgatgacTACAGGTGGTAGGAGGGCGAGCTG
RUNX1_E8B_F	gtagtgcgatggccagtTCGTCGCAAGCGCAGGGAGG
RUNX1_E8B_R	cagtgtgcagcgatgacGGGCTTGTCGCGAACAGGAGG
CEBPA_1F	gtagtgcgatggccagtGTTCGCCATGCCGGGAGAACTCTAAC
CEBPA_CEBPA AC	cagtgtgcagcgatgacACCCTGCTGCCGGCTGTGCTGGAAC
CEBPA_PP4F	gtagtgcgatggccagtGTCTTCAACGACGAGTTCCTGGCCGA
CEBPA_PP4R	cagtgtgcagcgatgacACAGCTGCTTGGCTTCATCCTCCT
CEBPA_PP5F	gtagtgcgatggccagtGTCCGCTGGTGATCAAGCAGGA
CEBPA_PP5R	cagtgtgcagcgatgacACCCGGTACTCGTTGCTGTTCT
CEBPA_CEBPA C	gtagtgcgatggccagtGTCAAGGCCAAGAAGTCGGTGGACA
CEBPA_PP6RBN	cagtgtgcagcgatgacACGGAGCAGGCCAGGCTTTCAG
P53_E2_F	gtagtgcgatggccagtCTGGATCCCCACTTTTCCTC
P53_E2_R	cagtgtgcagcgatgacTCCCACAGGTCTCTGCTAGG
P53_E3_F	gtagtgcgatggccagtAGCGAAAATTCATGGGACTG
P53_E3 _R	cagtgtgcagcgatgacGGGGACTGTAGATGGGTGAA

continued on the next page

continued from the previous page

Genes	Primer Sequence
P53_E4A_F	gtagtgcgatggccagtCCTGGTCCTCTGACTGCTCT
P53_E4A_R	cagtgtgcagcgatgacTTCTGGGAAGGGACAGAAGA
P53_E4B_F	gtagtgcgatggccagtGTCCAGATGAAGCTCCCAGA
P53_E4B_R	cagtgtgcagcgatgacCAGGCATTGAAGTCTCATGG
P53_E5_F	gtagtgcgatggccagtTTTCAACTCTGTCTCCTTCCTCTT
P53_E5_R	cagtgtgcagcgatgacAGCCCTGTCGTCTCTCCAG
P53_E6_F	gtagtgcgatggccagtCAGGCCTCTGATTCCTCACT
P53_E6_R	cagtgtgcagcgatgacCTTAACCCCTCCTCCCAGAG
P53_E7_F	gtagtgcgatggccagtTCATCTTGGGCCTGTGTTATC
P53_EX7_R	cagtgtgcagcgatgacAGTGTGCAGGGTGGCAAG
P53_EX8_F	gtagtgcgatggccagtGCCTCTTGCTTCTCTTTTCC
P53_EX8_R	cagtgtgcagcgatgacAACTGCACCCTTGGTCTCC
P53_EX9_F	gtagtgcgatggccagtGGAGACCAAGGGTGCAGTTA
P53_EX9_R	cagtgtgcagcgatgacGAAAACGGCATTTTGAGTGTT
P53_EX10_F	gtagtgcgatggccagtACTTCTCCCCCTCCTCTGTT
P53_EX10_R	cagtgtgcagcgatgacGAAGGCAGGATGAGAATGGA
P53_EX11_F	gtagtgcgatggccagtTGTCATCTCTCCTCCCTGCT
P53_EX11_R	cagtgtgcagcgatgacCAAGGGTTCAAAGACCCAAA
EZH2_E2_F	gtagtgcgatggccagtGGTGATCATATTCAGGCTGGT
EZH2_E2_R	cagtgtgcagcgatgacTTGAACTTAGGAGGGGAAAAA
EZH2_E3_F	gtagtgcgatggccagtTTTCTCCTTTCCTCTCCTTCA
EZH2_E3_R	cagtgtgcagcgatgacTCCTCCCAATAACCAAACAAA
EZH2_E4_F	gtagtgcgatggccagtTGGGTAGGCAGCATCTCTTT
EZH2_E4_R	cagtgtgcagcgatgacCTGTCTTGATTCACCTTGACAAT
EZH2_E5_F	gtagtgcgatggccagtTCTGGAGAACTGGGTAAAGACA
EZH2_E5_R	cagtgtgcagcgatgacGCCCCTTTTTCCAAGAGAAG
EZH2_E6_F	gtagtgcgatggccagtGCTATGCCTGTTTTGTCCAAG
EZH2_E6_R	cagtgtgcagcgatgacGCTGTAATGGCTACACAGAATCC
EZH2_E7_F	gtagtgcgatggccagtTGGGTAGAGAAAATGAAAGATCAA
EZH2_E7_R	cagtgtgcagcgatgacGCAAGATTGCCTCAAAGGAA
EZH2_E8_F	gtagtgcgatggccagtCATCAAAAGTAACACATGGAAACC
EZH2_E8_R	cagtgtgcagcgatgacAGCACTCTCCAAGCTGCTTTA
EZH2_E9_F	gtagtgcgatggccagtCCAGTGGAACTGGAAGAGTGA
EZH2_E9_R	cagtgtgcagcgatgacACCTCCACCAAAGTGCAAAG
EZH2_E10_F	gtagtgcgatggccagtTGATGTTGACATTTTTCATTTCG
EZH2_E10_R	cagtgtgcagcgatgacCAGTAAAACCCAGTTATTAGACGTG
EZH2_E11_F	gtagtgcgatggccagtTCCAATCATTTCTTGACCAGTG
EZH2_E11R	cagtgtgcagcgatgacTTTTCTTTGTTTGGACAACGAGT
EZH2_E12F	gtagtgcgatggccagtCTGTCCTCATGGCTCTGTGA
EZH2_E12R	cagtgtgcagcgatgacGCCTTGCCTGCAGTGTCTAT
EZH2_E13F	gtagtgcgatggccagtTTGTAGCTTCCCGCAGAAAT
EZH2_E13R	cagtgtgcagcgatgacCGTCCTCCATTCAAATTGGT
EZH2_E14F	gtagtgcgatggccagtCATCTCCCTGGATTCATTGG
EZH2_E14R	cagtgtgcagcgatgacGCAATTGCATCAAAGCAACA
EZH2_E15F	gtagtgcgatggccagtGTACAGCCCTTGCCACGTAT
EZH2_E15R	cagtgtgcagcgatgacAAGGCAATCCTGACATTTGC
EZH2_E16F	gtagtgcgatggccagtTCCATTTTCACCCTCCTTTTT
EZH2_E16R	cagtgtgcagcgatgacCATTTCCAATCAAACCCACA
EZH2_E17F	gtagtgcgatggccagtTATTCACTCTGTGCGCTTTTG
EZH2_E17R	cagtgtgcagcgatgacCCTCCCAGCTCTGAAACATAC
EZH2_E18F	gtagtgcgatggccagtAGGCAAACCCTGAAGAACTG
EZH2_E18R	cagtgtgcagcgatgacGGACTGAAAAGGGAGTTCCA
EZH2_E19F	gtagtgcgatggccagtCCGTCTTCATGCTCACTGAC
EZH2_E19R	cagtgtgcagcgatgacAAAAACCCTCCTTTGTCCAGA
EZH2_E20F	gtagtgcgatggccagtGCCATCCAGCGGACATCTCC
EZH2_E20R	cagtgtgcagcgatgacCTAAATTGCCCACAGTACTCGAGG
U2AF1_ex2f	gtagtgcgatggccagtTCATGCTGCTGACATATTCC
U2AF1_ex2r	cagtgtgcagcgatgacCACACTTATGAACACAAATGGAAA
U2AF1_ex6f	gtagtgcgatggccagtCCTGAGTGTGTATATCTCTCTCTGAT
U2AF1_ex6r	cagtgtgcagcgatgacTGGGTTGGAAGGAGACATTT
SRSF2_ex1f	gtagtgcgatggccagtGTCGGCGACGTGTACATCC
SRSF2_ex1r	cagtgtgcagcgatgacCGCGGACCTTTGTGAGGT
ZRSR2_ex1 f	gtagtgcgatggccagtCGTTTCAAGTCCCACGCT
ZRSR2_ex1r	cagtgtgcagcgatgacTTCCTGCCACCACATCCT
ZRSR2_ex2f	gtagtgcgatggccagtCTGCATTGTAGCCGCTGA
ZRSR2_ex2r	cagtgtgcagcgatgacGCTGGAGTGGACAGAGCAA

continued on the next page

continued from the previous page

Genes	Primer Sequence
ZRSR2_ex3f	gtagtgcgatggccagtTTGACCAAGGATTTGCAGC
ZRSR2_ex3r	cagtgtgcagcgatgacGACTGGTACTGGTTAGTAAAGGTTGA
ZRSR2_ex4f	gtagtgcgatggccagtTGTGGATTAATGCCCATTTC
ZRSR2_ex4r	cagtgtgcagcgatgacCCAACCTCCCAAGATAGGC
ZRSR2_ex5f	gtagtgcgatggccagtTGTGCGCTGTATGTGAAATG
ZRSR2_ex5r	cagtgtgcagcgatgacCCCAAACTCTGACATGCCT
ZRSR2_ex6f	gtagtgcgatggccagtTGTGTGCGTGTGTGTGTTTT
ZRSR2_ex6r	cagtgtgcagcgatgacCCACGAAACTAACATTACTGGAAC
ZRSR2_ex7f	gtagtgcgatggccagtTCATGGGTTTTTACTCCACCA
ZRSR2_ex7r	cagtgtgcagcgatgacCCTCTCCCAAAAGGGGAA
ZRSR2_ex8f	gtagtgcgatggccagtCCACCATGCCTGGTCTAAAG
ZRSR2_ex8r	cagtgtgcagcgatgacTGTGTCCCAGCTCTCTTGTG
ZRSR2_ex9f	gtagtgcgatggccagtGGGGAATGTTAGCCTGGA
ZRSR2_ex9r	cagtgtgcagcgatgacCAGGAAGACATCCACAAGCA
ZRSR2_ex10f	gtagtgcgatggccagtCAGTGAACTTGGTGGTCCTACA
ZRSR2_ex10r	cagtgtgcagcgatgacACTGGGTTTCCCCCAAAG
ZRSR2_ex11Af	gtagtgcgatggccagtTTCGGAAAAGGATAAAGTAGCA
ZRSR2_ex11Ar	cagtgtgcagcgatgacCTTCCCCCTGTGACGACTAC
ZRSR2_ex11Bf	gtagtgcgatggccagtAACCCTAGTCCAGACCACTCC
ZRSR2_ex11Br	cagtgtgcagcgatgacGCCTTCTATCCGAGTATGTAGCA
SF3B1 1f	gtagtgcgatggccagtAGCCCCCAGCTATTTTTCTC
SF3B1 1r	cagtgtgcagcgatgacTGTAAGAGGAGGACGCCATT
SF3B1 2f	gtagtgcgatggccagtGAAATGCATTGTGTTGGGAGT
SF3B1 2r	cagtgtgcagcgatgacTAAACCAGATGGCTGCAACA
SF3B1 3f	gtagtgcgatggccagtTGAAGGAGGGCTTAGACATCA
SF3B1 3r	cagtgtgcagcgatgacTGGGAACTCAGACATTCACTTTT
SF3B1 4f	gtagtgcgatggccagtGGCATGTATTAAACATTTGTGCTT
SF3B1 4r	cagtgtgcagcgatgacAATTCAGAAGCATGCCAAAAA
SF3B1 5f	gtagtgcgatggccagtGCAGGGCAGATAAATCAGTTG
SF3B1 5r	cagtgtgcagcgatgacTGGGGGTAAGATTCTTTCTCAG
SF3B1 6f	gtagtgcgatggccagtGGAAGTGATTGCGCTAATGG
SF3B1 6 r	cagtgtgcagcgatgacCTATGGCAACCCAAGCAGAC
SF3B1 7f	gtagtgcgatggccagtTTCTGTGTGGGTGTGTGAAA
SF3B1 7r	cagtgtgcagcgatgacATACGTGTCCACCCAGGAAT
SF3B1 8 f	gtagtgcgatggccagtAATTGTGGTTTTACTCACTCTTTCTTT
SF3B1 8 r	cagtgtgcagcgatgacAACAATTATGTCCAATGAGACAGTTC
SF3B1 9 f	gtagtgcgatggccagtAAATTTAAGTCTTGGTTTGCGTTT
SF3B1 9 r	cagtgtgcagcgatgacTCCTAAATACCACCTCATTCAAA
SF3B1 10 f	gtagtgcgatggccagtTGCAAATATTGTTCATTATGCTGTT
SF3B1 10 r	cagtgtgcagcgatgacAAAAATGTTAAGGGAAGTTGAAATG
SF3B1 11 f	gtagtgcgatggccagtAATGACCAGCCATCTGGAAA
SF3B1 11 r	cagtgtgcagcgatgacTGACAATTTAACAAACTGGGACTT
SF3B1 12 f	gtagtgcgatggccagtGAAACCACACCTATTACTCTGCTC
SF3B1 12 r	cagtgtgcagcgatgacAAGGAAAAGGTCTAGGAGAATATGT
SF3B1 13 f	gtagtgcgatggccagtCATGAGCATTTCATCAGTAATTG
SF3B1 13 r	cagtgtgcagcgatgacGTAGCCAGACCAGCAGCCTA
SF3B1 14 f	gtagtgcgatggccagtCCAACTCATGACTGTCCTTTCTT
SF3B1 14 r	cagtgtgcagcgatgacGGGCAACATAGTAAGACCCTGT
SF3B1 15 f	gtagtgcgatggccagtTTGGGGCATAGTTAAAACCTG
SF3B1 15 r	cagtgtgcagcgatgacTTCAAGAAAGCAGCCAAACC
SF3B1 16 f	gtagtgcgatggccagtGTATCCGCCAACACAGAGGA
SF3B1 16 r	cagtgtgcagcgatgacTGTTAGAACCATGAAACATATCCA
SF3B1 17 f	gtagtgcgatggccagtTTCTCTTCATTTCAGGTCAGTTG
SF3B1 17 r	cagtgtgcagcgatgacTCCATCTCCTTTCATAATCAAGC
SF3B1 18 f	gtagtgcgatggccagtTGCTTTATTTCCTTGGAAAAGC
SF3B1 18 r	cagtgtgcagcgatgacGCAATGTGCCATAATAGTTTTCAT
SF3B1 19 f	gtagtgcgatggccagtGCAACAGATGTTTGGGTGGT
SF3B1 19 r	cagtgtgcagcgatgacTTTGGGGAAGAAGTAAGAATTTG
SF3B1 20 f	gtagtgcgatggccagtTGTCATGAAGACTTGTCAAGAGG
SF3B1 20 r	cagtgtgcagcgatgacAACAAAAACCTCCCAACTCC
SF3B1 21 f	gtagtgcgatggccagtATCTGGGGCTTTCTCTTTCC
SF3B1 21 r	cagtgtgcagcgatgacATTGAATACAAAGTGGCCAAA
SF3B1 22 f	gtagtgcgatggccagtTCATGTTTTTAGAACTGAATTTGC
SF3B1 22 r	cagtgtgcagcgatgacTCAGACCATGCCTCAAAAGA
SF3B1 23 f	gtagtgcgatggccagtCAGCTTGTTGACCCATTTGTT
SF3B1 23 r	cagtgtgcagcgatgacTTCACGATGTTCTAAAATGAAGGA
SF3B1 24 f	gtagtgcgatggccagtCCGCATCTTAAAGGACTTTTT
SF3B1 24 r	cagtgtgcagcgatgacATGCATGCAGGGCTTAAAAC
SF3B1 25 f	gtagtgcgatggccagtCCCCAAGGGAAATAATTTTGA
SF3B1 25 r	cagtgtgcagcgatgacAGGTGTGAAGTAGCTGTGCATT

Online Supplementary Table S4. Exome data processed through the CASAVA pipeline showing a complete list of somatic mutations in six MDS exomes. †SNV present filtered out during the analysis as a result of low quality read or low read numbers.

PATIENT ID	WHO DIAGNOSIS	GENE	NCBI REFERENCE SEQUENCE ID	CHROMOSOME	CHROMOSOME LOCATION	REF BASE	MUTANT BASE	BASES USED	MUTANT READS	QSNP	CODON CHANGE	CONFIRMED	SOMATIC/ GERMLINE	GSK PIPELINE	CONSTITUTIONAL DNA SOURCE
RC-06-0256	MDS RARS	*AGXT2L1*	NM_001146590	4	109680914	G	A	30	11	104	T103I	Yes	Somatic	No	Skin
RC-06-0256	MDS RARS	*IKBKE*	NM_001193321	1	206658341	A	G	38	11	102	S394G	Yes	Somatic	Yes	Skin
RC-06-0256	MDS RARS	*GAB4*	NM_001037814	22	17472858	C	T	294	117	599	R128H	Yes	Somatic	No	Skin
RC-06-0256	MDS RARS	*OC90*	NM_001080399	8	133036862	C	T	41	15	113	A434T	Yes	Somatic	Yes	Skin
RC-06-0256	MDS RARS	*SF3B1*	NM_012433	2	198266494	T	C	76	27	185	D781G	Yes	Somatic	Yes	Skin
RC-06-0256	MDS RARS	*ARHGAP39*	NM_02525	8	145806353	A	C	12	5	92	T389G	Yes	Germline	No	Skin
RC-06-0256	MDS RARS	*VWA5B2*	NM_138345	3	183954168	T	C	10	3	23	L417P			No	Skin
RC-06-0256	MDS RARS	*C1orf85*	NM_144580	1	156265317	C	T	10	7	104	Q40Q			No	
RC-06-0256	MDS RARS	*COG5*	NM_001161520	7	107194778	T	C	26	9	112	R113G			Yes	
RC-06-0275	MDS RCMD-RS	*FGD1*	NM_004463	X	54492200	G	A	22	5	47	R476W	Yes	Somatic	No	Skin
RC-06-0275	MDS RCMD-RS	*TESC*	NM_001168325	12	117479796	G	A	8	2	23	P148S	Yes	Somatic	No	Skin
RC-06-0275	MDS RCMD-RS	*SF3B1*	NM_012433	2	198267491	C	G	68	30	237	E622D	Yes	Somatic	Yes	Skin
RC-06-0275	MDS RCMD-RS	*NAGS*	NM_153006	17	42082405	C	A	7	3	51	C374A	yes	Germline	Yes	Skin
RC-06-0275	MDS RCMD-RS	*TRIOBP*	NM_138632	22	38155502	A	C	84	40	23	T422P			No	
RC-06-0275	MDS RCMD-RS	*LTBP3*	NM_021070	11	65325357	G	A	14	3	24	A25V			No	
RC-06-0275	MDS RCMD-RS	*LTBP3*	NM_021070	11	65325358	C	G	14	3	20	A25P			No	
RC-06-0278	MDS RARS	*PKHD1*	NM_170724	6	51947297	C	A	150	56	314	L58F	Yes	Somatic	Yes	Skin
RC-06-0278	MDS RARS	*MC2R*	NM_000529	18	13885474	G	A	43	27	253	A15V	Yes	Somatic	Yes	Skin
RC-06-0278	MDS RARS	*PCNT*	NM_006031	21	47836271	A	G	87	35	211	I2147V	Yes	Somatic	Yes	Skin
RC-06-0278	MDS RARS	*SF3B1*	NM_012433	2	198267371	G	C	89	33	205	H662Q	Yes	Somatic	Yes	Skin
RC-06-0278	MDS RARS	*ATG9B*	NM_173681	7	150721444	C	T	9	4	58	G23R	Yes	Somatic	No	Skin
RC-06-0278	MDS RARS	*FITM2*	NM_001080472	20	42935567	C	T	50	24	168	A163T	Yes	Somatic	Yes	Skin
RC-06-0278	MDS RARS	*AFF3*	NM_001025108	2	100209883	G	T	62	12	28	P747H	Yes	Somatic	No	Skin
RC-06-0278	MDS RARS	*TET2*	NM_001127208	4	106164862	CT	-	18	8	354	1244_1244del	Yes	Somatic	No	Skin
RC-06-0278	MDS RARS	*CEP110*	NM_007018	9	123860787	A	G	67	17	73	S249G	Yes	Somatic	Yes	Skin
RC-06-0278	MDS RARS	*ZC3H18*	NM_144604	16	88694342	G	A	78	34	241	G2284A	Yes	Germline	Yes	Skin
RC-06-0278	MDS RARS	*FRMD4A*	NM_018027	10	13699338	T	G	8	5	20	T751P	Yes		No	Skin
RC-08-0010	MDS RARS	*KIAA1109*	NM_015312	4	123109050	G	A	92	37	151	G210R	Yes	Somatic	No†	CD3+ T-cells
RC-08-0010	MDS RARS	*SF3B1*	NM_012433	2	198266834	T	C	94	40	275	K700E	Yes	Somatic	Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*TET2*	NM_001127208	4	106157287	AC	-	110	46	2232	730_730del	Yes	Somatic	No	CD3+ T-cells
RC-08-0010	MDS RARS	*PPM1D*	NM_003620	17	58740624	A	-	182	76	1812	Q510fs	Yes	Somatic	No	CD3+ T-cells
RC-08-0010	MDS RARS	*AR*	NM_000044	X	66765212	A	C	28	9	97	A224C	Yes	Germline	No†	CD3+ T-cells
RC-08-0010	MDS RARS	*RP1L1*	NM_178857	8	10466010	C	A	164	47	82	G5598T	Yes	Germline	No	CD3+ T-cells
RC-08-0010	MDS RARS	*CREM*	NM_001881	10	35468147	A	G	67	30	244	A304G	Yes	Germline	No	CD3+ T-cells
RC-08-0010	MDS RARS	*SLC12A6*	NM_00104249	15	34543180	C	G	69	38	336	G1367C	Yes	Germline	No	CD3+ T-cells
RC-08-0010	MDS RARS	*FBLIM1*	NM_001024215	1	16101435	T	C	25	13	84	T1034C	Yes	Germline	No	CD3+ T-cells
RC-08-0010	MDS RARS	*FIZ1*	NM_032836	19	56104423	A	C	7	4	69	L295R			No†	CD3+ T-cells
RC-08-0010	MDS RARS	*COL6A6*	NM_001102608	3	130287289	G	A	205	99	658	E748K			Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*VAV3*	NM_006113	1	108507419	A	G	47	26	237	W25R			Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*DONSON*	NM_017613	21	34960634	G	C	17	5	75	P105R			Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*LOC401097*	NM_001168214	3	159943822	C	G	24	11	117	P134A			Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*SYCP2*	NM_014258	20	58461000	C	G	39	24	258	V838L			Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*B4GALNT4*	NM_178537	11	376345	C	T	15	8	121	P431S			Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*PKHD1L1*	NM_177531	8	110465039	C	T	39	63	182	I2200M			Yes	CD3+ T-cells
RC-08-0010	MDS RARS	*SPTB*	-	14	65249007	C	T	86	20	51	Splicesite 5' C>T			No	CD3+ T-cells
RC-08-0182	MDS RARS	*ATP1A4*	NM_144699	1	160143980	C	A	59	14	59	H691N	Yes	Somatic	No†	CD3+ T-cells
RC-08-0182	MDS RARS	*MCM3AP*	NM_003906	21	47666752	G	A	111	25	49	Q1447X	Yes	Somatic	No	CD3+ T-cells
RC-08-0182	MDS RARS	*SF3B1*	NM_012433	2	198267371	G	C	99	40	271	H662Q	Yes	Somatic	Yes	CD3+ T-cells
RC-08-0182	MDS RARS	*GPR112*	NM_153834	X	135455180	T	C	107	27	67	M2578T	Yes	Somatic	Yes	CD3+ T-cells
RC-08-0182	MDS RARS	*CAMLG*	NM_001745	5	134076908	C	A	69	32	240	C328A	Yes	Germline	No	CD3+ T-cells
RC-08-0182	MDS RARS	*PLEC*	NM_201381	8	145005769	C	T	9	4	63	G2141A	Yes	Germline	No	CD3+ T-cells
RC-08-0182	MDS RARS	*ACAN*	NM_013227	15	89399738	G	A	66	20	63	A1308T			No	CD3+ T-cells
RC-08-0182	MDS RARS	*SAMD11*	NM_152486	1	878744	G	C	12	8	125	G559A			No†	CD3+ T-cells
RC-10-0089	MDS RARS	ANXA7	NM_004034	10	75143371	C	-	39	10	410	T259fs	Yes	Somatic	No	CD3+ T-cells
RC-10-0089	MDS RARS	SF3B1	NM_012433	2	198266834	T	C	57	22	164	K700E	Yes	Somatic	Yes	CD3+ T-cells
RC-10-0089	MDS RARS	DNMT3A	NM_022552	2	25457242	C	T	41	20	175	R882H	Yes	Somatic	Yes	CD3+ T-cells
RC-10-0089	MDS RARS	SLC7A2	NM_001164771	8	17417899	C	T	80	31	212	S494L	Yes	Somatic	Yes	CD3+ T-cells
RC-10-0089	MDS RARS	IGF1R		15	99192762	T	C	10	8	83		Yes		No	CD3+ T-cells
RC-10-0089	MDS RARS	RAPGEF3	NM_001098531	12	48151789	G	A	8	4	52	79T	Yes	Germline	Yes	CD3+ T-cells
RC-10-0089	MDS RARS	ST3GAL3	NM_174964	1	44360114	G	A	82	39	289	.G407A	Yes	Germline	No	CD3+ T-cells
RC-10-0089	MDS RARS	C20orf118	-	20	35507598	T	G	21	7	24	Splicesite 3' T>G			No	CD3+ T-cells

Online Supplementary Table S5. Exome data processed through BWA/GATK (Broad Institute best practices pipeline) showing a complete list of somatic mutations in six MDS exomes. † SNV filtered out during the analysis as a result of low quality read or low read numbers.

PATIENT ID	WHO DIAGNOSIS	GENE NAME	ENS_GENE	CHROMOSOME	CHROMOSOME LOCATION	READ DEPTH	MUTANT READS	QUALITY	FILTER	CODON CHANGE	CODON CHANGE	CASAVA PIPELINE	CONSTITUTIONAL DNA SOURCE
RC-06-0256	MDS RARS	COG5	ENSG00000164597	7	107194778	28	11	311.59	PASS	Aga/Gga	R113G	Yes	Skin
RC-06-0256	MDS RARS	IKBKE	ENSG00000143466	1	206658341	38	11	238.34	PASS	Agc/Ggc	S479G	Yes	Skin
RC-06-0256	MDS RARS	OC90	ENSG00000253117	8	133036862	41	15	409.99	PASS	Gca/Aca	A434T	Yes	Skin
RC-06-0256	MDS RARS	SF3B1	ENSG00000115524	2	198266494	74	28	848.12	PASS	gAt/gGt	D781G	Yes	Skin
RC-06-0275	MDS RCMD-RS	NAGS	ENSG00000161653	17	42082405	6	3	63.08	PASS	aCg/aAg	T125K	Yes	Skin
RC-06-0275	MDS RCMD-RS	IL2RA	ENSG00000134460	10	6060053	17	8	225.67	PASS	Gtc/Atc	V253I	NO†	Skin
RC-06-0275	MDS RCMD-RS	SF3B1	ENSG00000115524	2	198267491	65	28	1621.88	PASS	gaG/gaT	E622D	Yes	Skin
RC-06-0278	MDS RARS	ENSG00000248602	ENST00000397266	7	150721444	9	4	74.58	PASS	-	G/R	NO	Skin
RC-06-0278	MDS RARS	MC2R	ENSG00000185231	18	13885474	43	27	789.9	PASS	gCa/gTa	A15V	Yes	Skin
RC-06-0278	MDS RARS	FITM2	ENSG00000197296	20	42935567	48	24	664.06	PASS	Gcc/Acc	A163T	Yes	Skin
RC-06-0278	MDS RARS	CEP110	ENSG00000119397	9	123860787	63	16	388.87	PASS	Agt/Ggt	S249G	Yes	Skin
RC-06-0278	MDS RARS	ZC3H18	ENSG00000158545	16	88694342	77	34	1053.25	PASS	Gaa/Aaa	E762K	Yes	Skin
RC-06-0278	MDS RARS	PCNT	ENSG00000160299	21	47836271	81	32	722.23	PASS	Ata/Gta	I2147V	Yes	Skin
RC-06-0278	MDS RARS	SF3B1	ENSG00000115524	2	198267371	90	33	2937.78	PASS	caC/caG	H662Q	Yes	Skin
RC-06-0278	MDS RARS	PKHD1	ENSG00000170927	6	51947297	144	53	1678.62	PASS	ttG/ttT	L58F	Yes	Skin
RC-10-0089	MDS RARS	RAPGEF3	ENSG00000079337	12	48151789	8	4	81.45	PASS	Cgg/Tgg	R27W	Yes	CD3+ T-cells
RC-10-0089	MDS RARS	DNMT3A	ENSG00000119772	2	25457242	39	21	1678.95	PASS	cGc/cAc	R882H	Yes	CD3+ T-cells
RC-10-0089	MDS RARS	SF3B1	ENSG00000115524	2	198266834	56	22	2355.32	PASS	Aaa/Gaa	K700E	Yes	CD3+ T-cells
RC-10-0089	MDS RARS	SLC7A2	ENSG00000003989	8	17417899	75	31	957.57	PASS	tCg/tTg	S493L	Yes	CD3+ T-cells
RC-08-0010	MDS RARS	MICALL2	ENSG00000164877	7	1484556	7	6	170.48	PASS	Gtg/Atg	V384M	NO†	CD3+ T-cells
RC-08-0010	MDS RARS	B4GALNT4	ENSG00000182272	11	376345	13	6	302.17	PASS	Ccg/Tcg	P431S	Yes	CD3+ T-cells
RC-08-0010	MDS RARS	DONSON	ENSG00000159147	21	34960634	15	5	404.27	PASS	cCg/cGg	P105R	Yes	CD3+ T-cells
RC-08-0010	MDS RARS	ENSG00000180044	ENST00000326474	3	159943822	22	10	236.76	PASS	Ccg/Gcg	P134A	Yes	CD3+ T-cells
RC-08-0010	MDS RARS	SYCP2	ENSG00000196074	20	58461000	40	24	1202.94	PASS	Gtt/Ctt	V838L	Yes	CD3+ T-cells
RC-08-0010	MDS RARS	VAV3	ENSG00000134215	1	108507419	46	26	1020.97	PASS	Tgg/Cgg	W25R	yes	CD3+ T-cells
RC-08-0010	MDS RARS	PKHD1L1	ENSG00000205038	8	110465039	116	57	2381.04	PASS	atT/atG	I2200M	Yes	CD3+ T-cells
RC-08-0010	MDS RARS	COL6A6	ENSG00000206384	3	130287289	193	92	5156.23	PASS	Gaa/Aaa	E748K	Yes	CD3+ T-cells
RC-08-0182	MDS RARS	GPR112	ENSG00000156920	X	135455180	107	27	585.01	PASS	aTg/aCg	M2578T	Yes	CD3+ T-cells
RC-08-0182	MDS RARS	SF3B1	ENSG00000115524	2	198267371	97	40	2937.78	PASS	caC/caG	H662Q	Yes	CD3+ T-cells
RC-08-0182	MDS RARS	WDR62	ENSG00000075702	19	36545988	13	9	284.2	PASS	Cca/Aca	P39T	No†	CD3+ T-cells

Online Supplementary Table S6. A complete list of somatic missense and nonsense mutations in 117 MDS patients detected by 454-Roche NGS sequencing. †Denotes confirmed novel somatic mutations found in our cohort of patients.*Denotes novel nonsense/splicesite mutations found in samples in which a paired constitutional source of DNA was not available. ‡Denotes previously confirmed somatic mutations in the *TP53* gene according to the International Agency for Research on Cancer (IARC) TP53 database.

Patient ID	Diagnosis	IPSS	Constitutional Source of DNA	Splicing Factor Mutations	Epigenetic Modifier Mutations	Cell signalling pathway/Transcription regulator and Other Gene Mutations
040015	MDS RARS	Low		**SF3B1** K700E[†] 51%	**DNMT3A** Q816X[†] 36%	
050024	MDS RCMD	Int-I	CD3+ T-Cells		**ASXL1** E635fsX15[†] 47% **EZH2** C548X† 73%	
050145	MDS RAEB-II	Int-II			**ASXL1** G635R fsX15[†] 17% **IDH2** R140Q[†]	
050187	MDS RCMD-RS	Int-I			**IDH2** R140W[‡] 24%	
050196	MDS RAEB-II	Int-II		**ZRSR2** E133Gfs11X† 54%		**MPL** W515K[†] 41%
050218	AML/MDS	N/A		**U2AF1** Q157P[#] 41%	**ASXL1** E635fsX15[†] 36% **IDH2** R140Q[†] 33%	**TP53** T253fsX57[‡] 80%
050231	MDS RAEB-I	Int-I		**U2AF1** S34F[‡] 10%	**TET2** G1275E[†] 68% **TET2** G1282D[†] 68% **TET2** D1844G[†] 68%	
050237	MDS RCMD	Int-II				**TP53** K132Q[‡] 42% **TP53** G262V[‡] 35%
050276	MDS RAEB-II	Int-II		**SRSF2** P95L[20] 39%		**NRAS** G13V[11] 11%
050278	MDS/MPD	N/A	CD3+ T-Cells	**SRSF2** Y93fsX121† 16%		**NRAS** G12S[12] 42%
060007	MDS RCMD-RS	Low		**SF3B1** K666R[13] 49%		**CCBL** R420Q [14]13%
060009	MDS RAEB-II	Int-II			**IDH2** R140Q[†] 33%	**TP53** K164E[†] 7% **RUNX1** S322X* 20%
060017	MDS RARS	Low		**SF3B1** K700E[†] 51%	**TET2** S1494X[15] 40%	
060018	MDS RCMD-RS	Low		**SF3B1** K666Q[‡] 43%	**DNMT3A** SA 3' Exon 19* 33%	**CEBPA** G100_G101del* 6%
06-0031	MDS RARS	Low		**SF3B1** E622D[‡] 60%		
060039	MDS RCMD	Int-I	Skin		**DNMT3A** W313X† 32%	
060070	AML/MDS	N/A			**ASXL1** R774fsX* 10 %	**RUNX1** R162K[16] 72%
060075	MDS RCMD	Int-I	CD3+ T-Cells	**SRSF2** Y93fsX121† 24%		
060084	MDS RCMD	Int-I	Skin	**SRSF2** Y93fsX121† 35%	**IDH2** R140Q[†] 34%	
060103	AML/MDS	N/A	CD3+ T-Cells		**IDH2** R140Q[†] 12%	**RUNX1** S369X† 11%
060114	AML/MDS	N/A	CD3+ T-Cells			**FLT3** I.601_K602Ins9† 40% **CCBL** Y368S† 8%
060129	MDS RARS	Low		**SF3B1** K700E[†] 41%		**JAK2** V617F[17] 24%
060145	MDS RCMD	Int-I		**SRSF2** P95L[21] 49%	**TET2** L371X[†] 47.5% **TET2** Y1569fsX1[†] 37%	
060152	MDS RAEB 1	Int-I		**SRSF2** P95L[22] 50%	**ASXL1** R693X[18] 26%	
060158	MDS RAEB-II	High				**TP53** R273H[‡] 18%
060160	MDS RAEB-II	High	CD3+ T-Cells		**DNMT3A** F751V† 25% **DNMT3A** P904S† 20%	
060187	MDS RCMD	Low			**TET2** C1193W[†] 49% **EZH2** R690H[19] 70%	
060188	MDS RARS	Low	CD3+ T-Cells		**ASXL1** G643V† 32%	
060199	tMDS	Int-II	CD3+ T-Cells			**TP53** Y220C[‡] 42% **TP53** Q331H[‡] 41% **CEBPA** P183_P185del† 7%
060219	MDS RAEB 1	Int-I	CD3+ T-Cells	**SRSF2** P95L[23] 41%	**ASXL1** Q760X[23] 37%	**RUNX1** SA 3' Exon4† 32%
060237	CMML	N/A		**SRSF2** P95H[†] 34%	**TET2** G1288V[†] 43% **TET2** Q1699X[†] 49%	**CKIT** D816V[21] 35%
060247	MDS RAEB-II	High		**SRSF2** P95H[†] 45%	**ASXL1** G646Wfs X12[23] 22%	**FLT3** R595_612dup† 5% **NRAS** Q61R[12] 11%
060256	MDS RARS	Low	Skin	**SF3B1** D781G[19] 45%		
060261	MDS RAEB-II	High				**JAK2** V617F[17] 58%
060262	MDS/MPD	Low	Skin	**U2AF1** Q157P[#] 45%	**DNMT3A** N551S† 32% **ASXL1** W583X† 31% **TET2** R1216X[†] 41%	
060264	MDS RAEB 1	Int-II			**TET2** R1878H[†] 6%	**TP53** L43X[‡] 40% **TP53** C238Y[‡] 41 %
060265	MDS RAEB 1	Int-II	Skin		**DNMT3A** R326H† 22% **ASXL1** G646Wfs X12[23] 18% **EZH2** R583X[19] 38%	
060266	MDS RCMD	Low		**U2AF1** Q157P[#] 27%		
060270	AML/MDS	N/A	CD3+ T-Cells	**SF3B1** K666N[19] 34%		
060272	MDS RAEB 1	Int-I		**SRSF2** P95L[22] 56%	**TET2** Q888X[†] 45% **TET2** R1465X[†] 35%	
060273	MDS RCMD	Int-I		**SRSF2** P95L[24] 63%	**TET2** S716X[†] 45%	
060275	MDS RCMD-RS	Int-I	Skin	**SF3B1** E622D[19] 40%		
060278	MDS RARS	Int-I	Skin	**SF3B1** H662Q[19] 37%	**TET2** L1245delCT fsX22[25] 45%	
060282	MDS RCMD	Int-II			**DNMT3A** A741V[26] 44% **ASXL1** G646Wfs X12[23] 17%	
060285	MDS RCMD	Int-I			**TET2** I1195V[†] 47%	
060286	CMML	N/A			**ASXL1** Q1234X * 10% **TET2** N1743fsX[†] 17 46%	**NRAS** G12S[12] 41%
060290	MDS RAEB-II	N/A		**U2AF1** Q157P[#] 28%		
060297	MDS RCMD	Low			**TET2** Q1529X[†] 38%	
060309	MDS RAEB 1	Int-II	CD3+ T-Cells		**TET2** L1322fsX† 22%	
060329	CMML	N/A		**SF3B1** K666R[13] 50%	**ASXL1** E635fsX15[†] 27%	
060331	AML/MDS	Int-II		**U2AF35** Q157P[#] 34%	**IDH2** R140Q[†] 37%	
060335	MDS RAEB 1	Int-II				**TP53** V157F[‡] 24%
060336	MDS RAEB 1	Int-II			**TET2** Y1148C[27] 25%	**TP53** G266E[‡] 28% **TP53** G325fsX12[‡] 28%
060337*[‡]	MDS RCMD-RS	Low	Skin	**SF3B1** R625C[28] 39%	**TET2** S835X† 60%	
060337*[‡]	MDS RCMD-RS		Skin	**SF3B1** R625C[28] 42%	**TET2** S835X†	**FLT3** F590_W603dupInsP† **RUNX1** F163Y[29] 30%
060337*[‡]	MDS RCMD-RS		Skin	**SF3B1** R625C[28] 46%	**TET2** S5835X† 45%	**FLT 3** F590_W603dupInsP† **RUNX1** F163Y[29] 40%
060338	MDS RCMD-RS	Int-I	Skin		**IDH2** R140Q[†]	**CEBPA** P183_P185del† 8%
060339	MDS RCMD	Int-I	Skin	**U2AF1** R156H[#] 36%	**DNMT3A** R882P[1,20,30] 35%	

continued on the next page

continued from the previous page

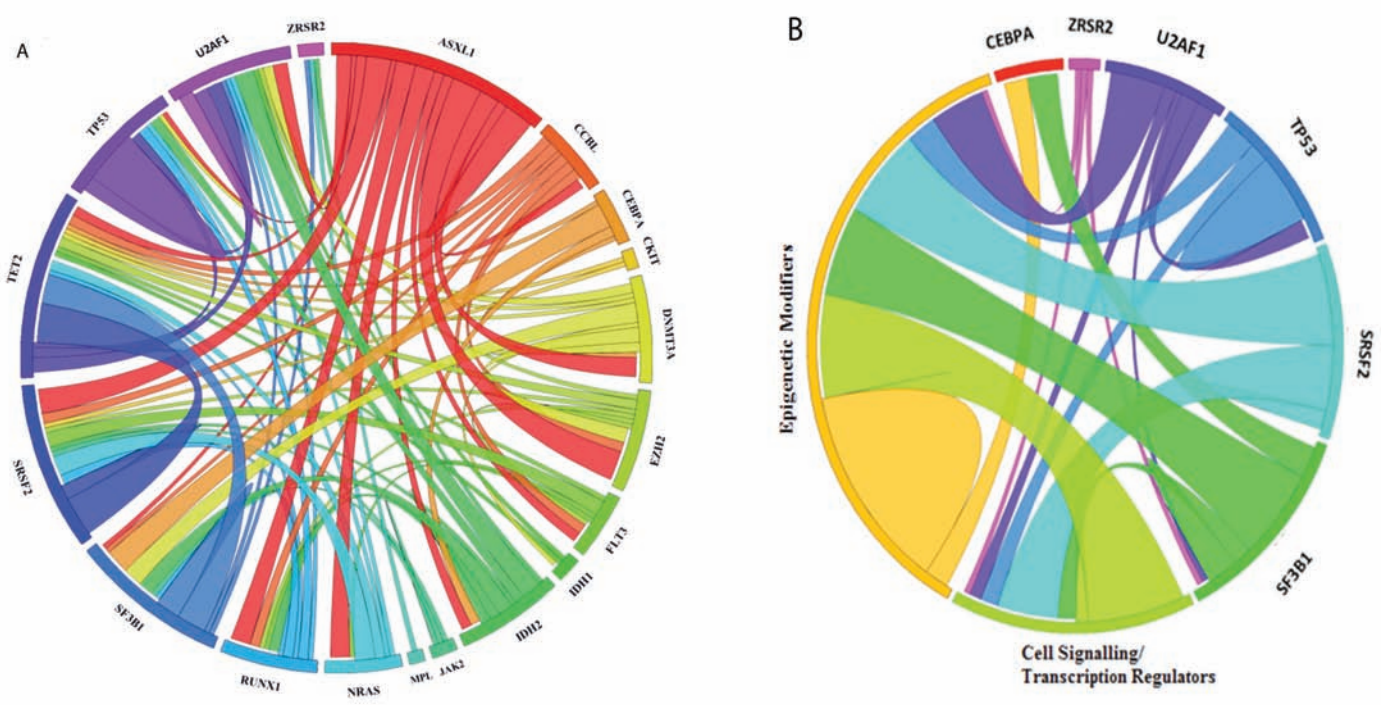
Patient ID	Diagnosis	IPSS	Constitutional Source of DNA	Splicing Factor Mutations	Epigenetic Modifier Mutations	Cell signalling pathway/Transcription regulator and Other Gene Mutations
060342	MDS RA	Low			**TET2** V1862fsX24[4] 98%	
060361	MDS RCMD	Low			**TET2** Q644X[5] 36% **TET2** I.1360fs[5] 36% **TET2** P1962L[4] 40%	
060365	MDS RCMD	Int-II	CD3+ T-Cells		**ASXL1** R860X† 34%	**RUNX1** R201Q[21] 31%
070066	MDS RAEB-II	Int-II			**TET2** A1355P[6] 26% **TET2** V1606fsX22[5] 19%	
070071	CMML	N/A			**TET2** SD 3 exon 10[6] 42% **TET2** Y1631fsX28[6] 51%	**NRAS** G12D[12] 35%
070076	MDS RAEB-I	Int-II		**U2AF1** S34F[4] 44%		**NRAS** G12D[12] 44%
070077	MDS RARS	Low	Skin	**SF3B1** K700E[2] 38%		**CEBPA** P183_P185del† 12%
070088	MDS RCMD	Int-I			**TET2** T1393P[6] 39%	
070093	MDS/MPD	N/A	Skin		**EZH2** G686D† 22%	
070095	MDS RCMD-RS	Int-II				**TP53** V173M[2] 24% **TP53** H214R[2] 13%
070104	MDS RCMD	Low			**TET2** P1962L[4] 50%	
070141	MDS RAEB-I	Int-I		**SF3B1** K666R[2] 28% **U2AF1** Q157R[4] 28%	**IDH2** R140Q[4] 16%	
070148	MDS/MPD	N/A		**SRSF2** P95R[8] 30%		
070180	AML/MDS	High	Buccal	**U2AF1** Q157P[4] 27%		**TP53** R248Q[2] 57% **RUNX1** G165R†[21] 28%
070186	MDS/MPD	N/A	Skin		**ASXL1** E635fsX15[4] 23% **EZH2** N693H† 50%	**CCBL** Y368S† 37%
070192	MDS/MPD	N/A	Buccal		**TET2** C1273F[6] 71% **TET2** R1452X[6] 20% **EZH2** SD 5' Exon 19† 20%	
070203	MDS RAEB-II	High				**TP53** SD Exon 3'[2] 35% **TP53** H178D[2] 40%
070207	MDS/MPD	N/A		**SRSF2** P95H[8] 44%		**JAK2** V617F[15] 32%
070217	MDS RAEB-II	Int-II			**ASXL1** G646WfsX12[22] 22 %	
070221	AML/MDS	N/A		**U2AF1** Q157P[4] 27%	**IDH2** R172K[23] 11%	
070222	MDS RCMD	Int-I	Skin	**U2AF1** Q157P[4] 16%	**ASXL1** W898X† 13% **EZH2** G743V† 16%	
070226	MDS RCMD	Int-I		**SRSF2** P95L[10] 48%	**TET2** Q1687X[6] 41% **TET2** I1873N[6] 39%	
070259	CMML	N/A		**SRSF2** P95R[8] 42%	**TET2** H1380Y[6] 50%	**FLT3** F593_D600dupInsF† 44%
070280	MDS RCMD	Int-I	Skin		**DNMT3A** A376T† 34% **EZH2** V626M[24] 40%	
070297	MDS RCMD	Int-I		**U2AF1** S34F[4] 10%		
070301	MDS RCMD	Int-II	Skin		**TET2** C1135S† 12%	
070311	MDS RCMD	Int-I	Skin	**SRSF2** Y936X121† 13%		
070317	MDS RAEB-I	Int-II				**TP53** L257Q[2] 60%
070319	MDS RAEB-II	Int-II			**IDH2** R140Q[4]	
070329	MDS RAEB-II	High	Skin	**SRSF2** P95L[10] 38%	**TET2** Q255R† 35% **TET2** R1452X[6] 35%	
080010	MDS RARS	Int-I	CD3+ T-Cells	**SF3B1** K700E[2] 43%		
080027	CMML	High	CD3+ T-Cells	**SRSF2** P95H[8] 32%	**ASXL1** S666Sfs3X22† **EZH2** G564fsX107† 40% **EZH2** V679L† 43%	**FLT3** E596_Y597Ins14† 15% **CCBL** R348X† 10% **CCBL** K396Y† 40% **RUNX1** R166X[20] 38%
080040	tMDS	Int-II				**TP53** V173L[2] 93%
080046	MDS RAEB-I	Int-II	Skin		**DNMT3A** T835A 22%	**TP53** Y163C[2] 2.5% **TP53** C275Y[2] 47%
080077	AML/MDS	High				**TP53** P152fsX28[2] 70%
080117	MDS RAEB-II	Int-II	CD3+ T-Cells		**ASXL1** G646WfsX12[22] 10% **EZH2** N130I 9%	
080129	MDS RAEB-I	Int-II				**TP53** S149fsX20[2] 90%
080161	MDS RCMD-RS	Int-I	Skin		**TET2** Q634X† 34% **TET2** Q916X† 42%	
080169	MDS RAEB-II	Int-II	Skin		**TET2** S1369P† 23%	**CCBL** SD del Exon-8† 5% **RUNX1** R166Q [20] 32%
080171	MDS RARS	Low	Skin	**SF3B1** K700E[2] 42%		**CEBPA** G100del†13% **CEBPA** P183_P183del† 8%
080182	MDS RARS	Low	CD3+ T-Cells	**SF3B1** H662Q[25] 45%		
080186	MDS RAEB-II	High	CD3+ T-Cells		**DNMT3A** R366H† 22% **ASXL1** R693X[26]	
080194	MDS RAEB-II	High			**DNMT3A** R882C[24] 34% **EZH2** V626M[24] 78%	
080215	MDS RAEB-I	Int-II			**ASXL1** G646WfsX12[22] 15%	**NRAS** G12V[20] 23%
080223	MDS RAEB-II	High				**TP53** SD Exon 5'[2] 28% **TP53** R348W[2] 38% **NRAS** G12D[12] 35%
080236	MDS RAEB-II	High	Skin		**ASXL1** Q708X†	
080260	MDS RCMD-RS	Int-I	Skin	**SF3B1** K700E[2] 40%	**IDH2** R140Q[4] 35%	**CEBPA** G100del† 6%
090006[A1]	MDS RARS	Low		**SF3B1** H662Q[25] 42%		**RUNX1** Y281X[A1] 50%
090006[A3]	MDS RARS			**SF3B1** H662Q[25] 42%		**RUNX1** Y281X [A3] 50%
090035	MDS RCMD	Int-II			**ASXL1** G646WfsX12[22] 9%	
090045	MDS RAEB-II	High		**U2AF1** S34F[4] 29%		**TP53** F134L[2] 14%
090083	tMDS	Int-II				**TP53** Y163C[2] 45% **TP53** Y234C[2] 35%
090152	CMML	Int-II		**SRSF2** P95R[8] 42%	**TET2** R1465X[6] 50%	**CCBL** C404Y[A2] 80%
090182	MDS RCMD-RS	Int-I	CD3+ T-Cells	**SF3B1** H662Q[25] 30%	**DNMT3A** W581S† 22%	
090214	MDS RAEB-II	High	Skin	**U2AF1** Q157P[4] 23%	**TET2** R1878H† 27% **IDH1** R132H[23]	
090251	MDS RAEB-II	High			**ASXL1** E635fsX15[4] 40%	
090296	MDS RAEB-II	High		**SRSF2** P95R[8] 25%	**ASXL1** G647X* 30%	**NRAS** G12D[12] 34%
090387	MDS RAEB-I	Int-II			**DNMT3A** SA 3' Exon 18* 43% **IDH1** R132L[24]	
100089	MDS RARS	Low		**SF3B1** K700E[2] 40%	**DNMT3A** R882H[24] 50%	
100101	AML/MDS	N/A	Skin	**SF3B1** K700E[2] 35% **ZRSR2** S439_R440del† 64%	**IDH2** R140Q[4] 21%	
100135	MDS RARS	Low	Buccal	**SF3B1** K700E[2] 45%	**TET2** R550X[26] 84%	**CEBPA** P183del† 9%
100162	MDS RARS	Int-I	Skin	**SF3B1** K700E[2] 30%	**TET2** Q652X† 34%	

Table S7. Summary of mutations coexisting with mutant and wild-type spliceosome components in 117 patients. Mutations were grouped according to their functional relevance; splicing factor mutations, *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2*; epigenetic modifier mutations, *TET2*, *IDH1/2*, *ASXL1*, *EZH2*, and *DNMT3A*; Cell signaling/transcription regulator mutations, *FLT3*, *NRAS*, *C-KIT*, *RUNX1*, *CCBL*, *JAK2* and *MPL*; mutations in tumor suppressor *TP53*. The top half of the table compares mutations in the grouped genes. The bottom half of the table indicates individual gene mutations coexisting with splicing factor mutations, 37 patients were wild-type for all genes present in our panel screen. Table cells indicate number of patients within mutant (white cells) or non-mutant (shaded cells) spliceosome groups, followed by the percentage where appropriate.

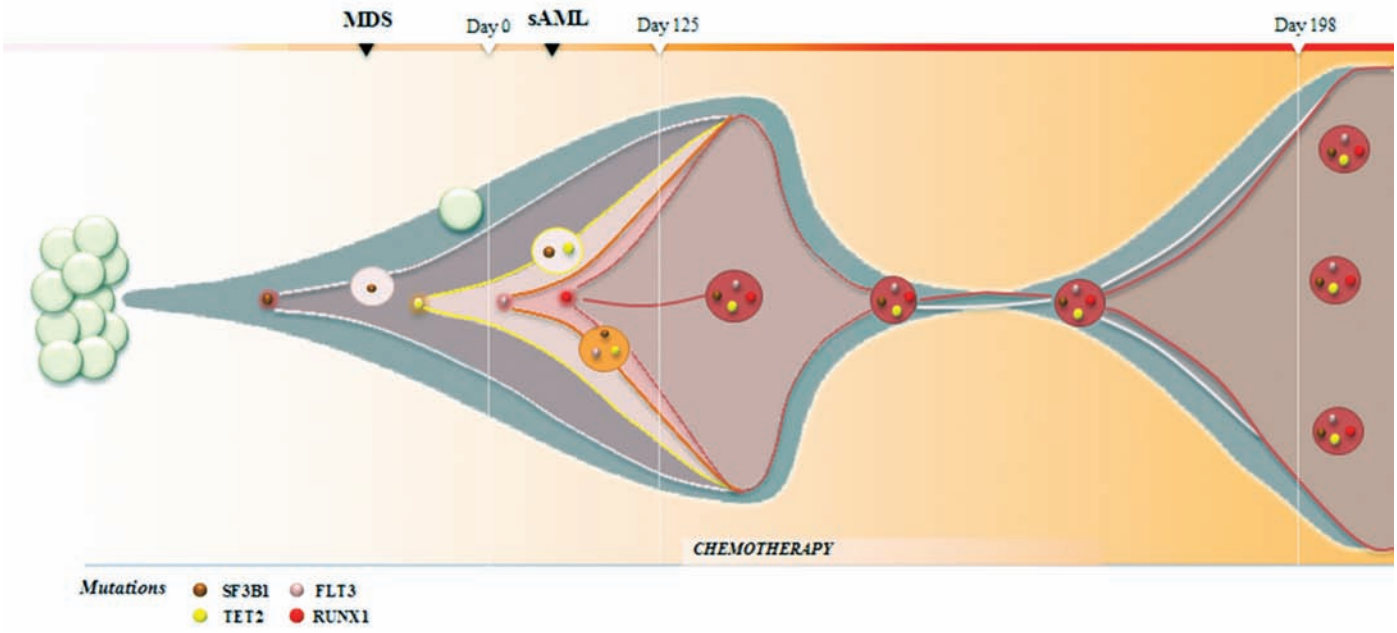
	Mutant Spliceosome	SF3B1	SRSF2	U2AF1	ZRSR2	Wild Type Spliceosome With Other Mutations
Total mutant cases	59	24	20	15	2	58
Cases with additional co-existing non-splice factor mutations	48 (81%)	19 (79%)	17 (85%)	12 (80%)	2 (100%)	27 (47%)
Cases with >1 additional co-existing non-splice factor mutations	15 (25%)	3 (13%)	7 (35%)	5 (33%)	0	5 (9%)
Cases with co-existing *TP53* mutations	3 (5%)	0	0	3 (20%)	0	16 (28%)
Cases with co-existing Cell Signalling /Transcription regulator mutations	16 (27%)	3 (13%)	10 (50%)	2 (13%)	1 (50%)	12 (21%)
Cases with co-existing Epigenetic modifier mutations	36 (61%)	14 (58%)	14 (70%)	9 (60%)	1 (50%)	44 (76%)
CO-EXISTING GENE MUTATIONS						
TP53	3 (5%)	0	0	3 (20%)	0	16 (28%)
FLT3	3 (5%)	0	3 (15%)	0	0	1 (2%)
NRAS	5 (8%)	0	4 (20%)	1 (7%)	0	4 (7%)
RUNX1	4 (7%)	1 (4%)	2 (10%)	1 (7%)	0	5 (9%)
CCBL	3 (5%)	1 (4%)	2 (10%)	0	0	3 (5%)
C-KIT	1 (2%)	0	1 (5%)	0	0	0
JAK2	2 (3%)	1 (4%)	1 (5%)	0	0	1 (2%)
MPL	1 (2%)	0	0	0	1 (50%)	0
ASXL1	9 (15%)	1 (4%)	5 (25%)	3 (20%)	0	17 (29%)
DNMT3A	6 (10%)	4 (17%)	0	2 (13%)	0	9 (16%)
EZH2	2 (3%)	0	1 (5%)	1 (7%)	0	9 (16%)
TET2	17 (29%)	6 (25%)	8 (40%)	3 (20%)	0	17 (30%)
IDH2	7 (12%)	3 (13%)	1 (5%)	4 (27%)	1 (50%)	6 (11%)
IDH1	1 (2%)	0	0	1 (7%)	0	1 (2%)
CEBPA	5 (8%)	5 (21%)	0	0	0	1 (2%)

Online Supplementary Table S8. Multivariable analysis of overall survival. The variables included are age, WHO category, bone marrow blasts, IPSS cytogenetic risk groups, transfusion dependency, *SF3B1* mutations, *NRAS* mutations and *TP53* mutations.

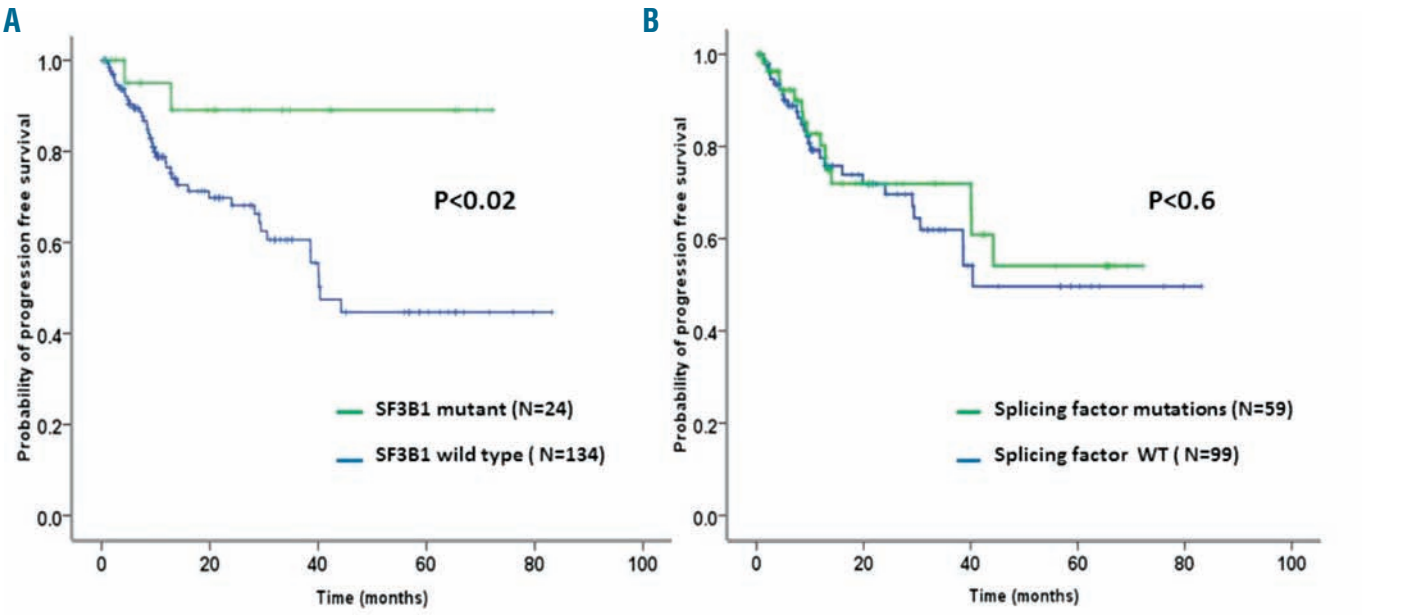
Variable	OS		
	Hazard Ratio	95% CI	p value
Age	1.0	1-1.1	0.05
WHO category	1.1	0.9-1.3	0.12
IPSS Cytogenetic	1.9	1.3-2.6	0.001
Medullary blast count	1	0.9-1	0.2
Transfusion dependency	1.7	0.9-3	0.08
SF3B1	0.2	0.1-0.8	0.03
TP53	2.1	1.1-4.4	0.04



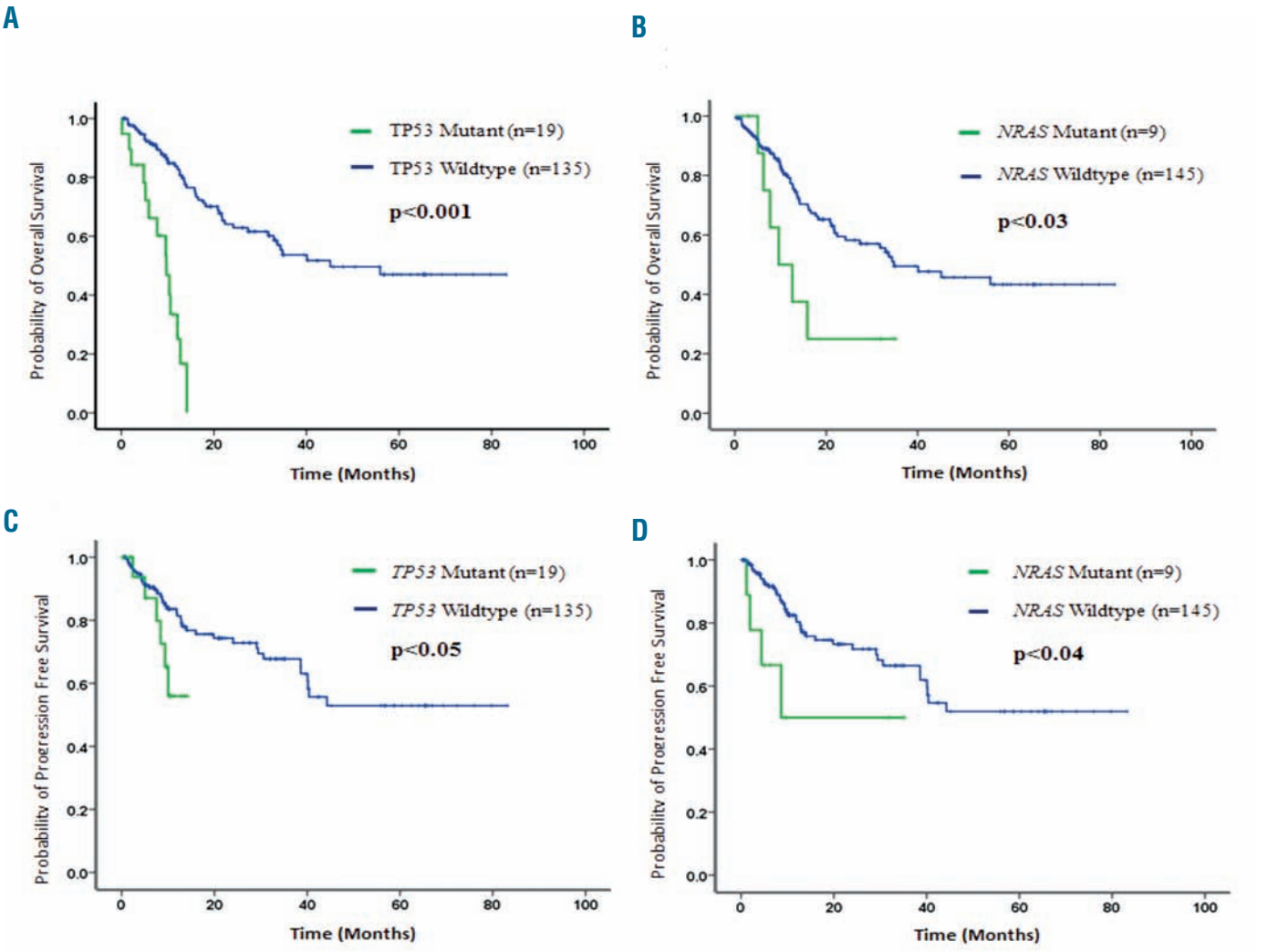
Online Supplementary Figure S1. Frequency and distribution of various gene mutations in MDS patients as depicted by circos diagrams.[9] The arc length corresponds to the frequency of mutations in the relevant gene and the ribbon width corresponds to the percentage of the patients who have other coexisting mutations. Diagram showing the coexistent mutations between 22 individual genes (A). Diagram showing the coexistent mutations between splicing factor genes (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*), epigenetic modifiers, cell signalling/transcription regulators, *TP53* and *CEBPA* (B).



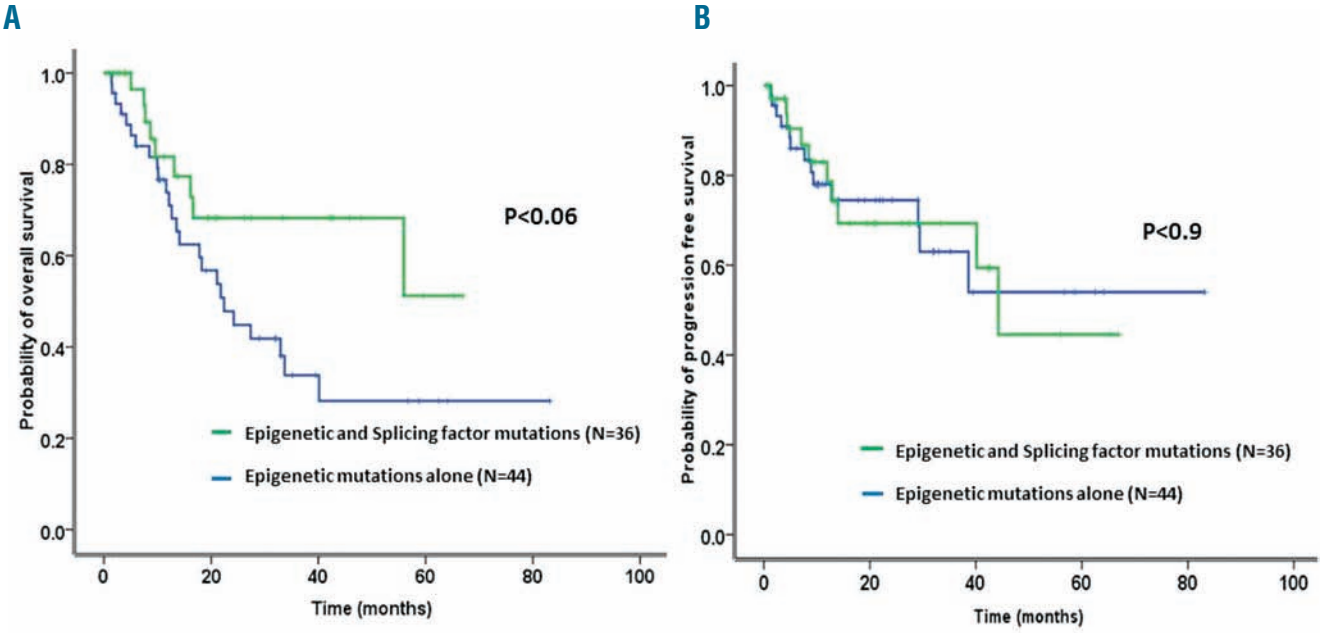
Online Supplementary Figure S2. Clonal evolution from MDS to AML with sequential acquisition of mutations. A normal hematopoietic stem cell (green cell) acquires mutations in *SF3B1* and *TET2* genes during the early phase of MDS. Subsequently, one sub-clone harboring *SF3B1* and *TET2* mutations acquires additional mutations in *FLT3* and *RUNX1* genes and evolves to become the dominant clone at transformation to AML. This is followed by further clonal expansion during subsequent progression with an accompanying rise in blast count. There was a period of a transient morphological remission between day 125 and 198 due to chemotherapy, although samples were not available at the time of the transient remission. The internal areas are representative of disease sub-clone size. Sample time points are indicated by vertical white lines representing time of diagnosis (day 0), transformation to AML (day 125) and further disease progression after a short duration of morphological remission (day 198). The order in which *SF3B1* and *TET2* mutations occur is unknown and only exampled here.



Online Supplementary Figure S3. Progression-free survival for patients with *SF3B1* mutations (n=24) compared with wild type *SF3B1* (n=130) (A) and progression-free survival for patients with spliceosome mutations (n=59) compared with patients without splicing factor mutations (n=99) (B).



Online Supplementary Figure S4. Overall survival (A) and progression-free survival (C) for patients with *TP53* mutations (n=19) compared with wild-type *TP53* (n=135). Overall survival (B) and progression-free survival (D) for patients with *NRAS* (n=9) compared with patients without *NRAS* mutations (n=145)



Online Supplementary Figure S5. Overall survival (A) and progression-free survival (B) for patients with epigenetic modifier mutations stratified according to coexistence of splicing factor mutations with epigenetic modifier mutations (n=36) or epigenetic modifier mutations alone (n=44).